SETI Institute

FINAL REPORT

NASA Cooperative Agreement # NCC 2-881

"Advanced Signal Processing Methods Applied to Digital Mammography"

August 1, 1994 to October 31, 1997

Principal Investigator: Richard P. Stauduhar

(Consists of Interim Report from August 1, 1994 to June 30, 1996 and Final report from July 1, 1996 to October 31, 1997)

NASA Cooperative Agreement NCC 2-881

Interim Report from July 1, 1996 to October 31, 1997

Principal Investigator: Richard P. Stauduhar

SETI Institute

The work reported here is on the extension of the earlier proposal of the same title, August 1994–June 1996. The report for that work is also being submitted. The work reported there forms the foundation for this work from January 1997 to September 1997.

After the earlier work was completed there were a few items that needed to be completed prior to submission of a new and more comprehensive proposal for further research. Those tasks have been completed and two new proposals have been submitted, one to NASA, and one to Health & Human Services (HHS).

The main purpose of this extension was to refine some of the techniques that lead to automatic large scale evaluation of full mammograms. Progress on each of the proposed tasks follows.

Task 1: A multiresolution segmentation of background from breast has been developed and tested. The method is based on the different noise characteristics of the two different fields. The breast field has more power in the lower octaves and the off-breast field behaves similar to a wideband process, where more power is in the high frequency octaves. After the two fields are separated by lowpass filtering, a region labeling routine is used to find the largest contiguous region, the breast.

Task 2: A wavelet expansion that can decompose the image without zero padding has been developed. The method preserves all properties of the power-of-two wavelet transform and does not add appreciably to computation time or storage. This work is essential for analysis of the full mammogram, as opposed to selecting sections from the full mammogram.

Task 3: A clustering method has been developed based on a simple counting mechanism. No ROC analysis has been performed (and was not proposed), so we cannot fully evaluate this work without further support.

Task 4: Further testing of the filter reveals that different wavelet bases do yield slightly different qualitative results. We cannot provide quantitative conclusions about this for all possible bases without further support.

Task 5: Better modeling does indeed make an improvement in the detection output. After the proposal ended, we came up with some new theoretical explanations that helps in understanding when the D4 filter should be better. This work is currently in the review process.

Task 6: N/A. This no longer applies in view of Tasks 4-5.

Task 7: Comprehensive plans for further work have been completed. These plans are the subject of two proposals, one to NASA and one to HHS. These proposals represent plans for a complete evaluation of the methods for identifying normal mammograms, augmented with significant further theoretical work.

Presentations made:

D. Kent Cullers, at the Aerospace Medical Association, 67th Annual Scientific Meeting, "Medical Applications of Space Research and Technology," Atlanta, GA, May 6–9, 1996.

John J. Heine, at the Site Visit by the Office on Womens's Health Review Pannel, under direction of Dr. F. Shtern, NASA Ames Research Center, February 25, 1997.

D. Kent Cullers, at the Technology Transfer Workshop on Breast Cancer Detection, Diagonosis, and Treatment, Washington, D.C., May 1–2, 1997.

J.J. Heine, S.R. Deans, D.K. Cullers, R. Stauduhar, and L.P. Clarke, "Multiresolution statistical analysis of high resolution digital mammograms," IEEE Trans. on Medical Imaging," Vol. 16, No. 5, pp. 503–515, October 1997.

Publication:

J.J. Heine, S.R. Deans, D.K. Cullers, R. Stauduhar, and L.P. Clarke, "Multiresolution probability analysis of gray scaled images," J. Opt. Soc. Am. A., (to be published).

NASA Cooperative Agreement NCC 2-881

Interim Report from August 1, 1994 to June 30, 1996

Principal Investigator: Richard P. Stauduhar

SETI Institute

1. Specific Aims

The long term goal of this project is to recognize negative mammograms. The short term goal is the development of a dependable multiresolution statistical model for normal tissue. The model will be used in a detection technique that should meet the criteria of high specificity and sensitivity.

2. Background

2.1 SETI Program: During the development of NASA's Search for Extraterrestrial Intelligence (SETI) Program, it was conjectured that one might generalize the very successful special techniques developed for detecting weak TV and radar signals. There was considerable interest in doing this. First, it was obvious that the number of signal types to which SETI was sensitive did not completely cover the class of intelligent signals. Thus there was interest in developing generalized feature extraction algorithms that could potentially differentiate systematic patterns from random background. Second, SETI efforts were also directed toward the complex problem of time varying features. It might be possible to develop a more general model for randomly occurring interference as nongaussian noise. These efforts provide the basis for a technology transfer from NASA/SETI to the private sector with application to breast cancer screening.

2.2 Breast Cancer Screening: Since the cause of breast cancer is still unknown, the most cost effective strategy for early diagnosis is screen film mammography. The NCI cancer control agency hopes to promote screening for 80% of eligible women in the United States by the year 2000 at a projected annual cost of two billion dollars. Since over 95% of screening mammograms are normal, the proposed method would significantly reduce the cost for breast cancer screening and would be timely with the conversion to direct digital mammography by the year 2000. The proposed software method could be used for computer assisted diagnosis (CAD) approaches, which aim at the detection and/or classification of a breast abnormality. The emphasis of the method is on achieving high specificity, the identification of negative mammograms at a very low false negative rate.

3. Technology Breakthrough

The identification of microcalcifications using single scale methods has proved of limited success. The USF research group has pioneered the use of multiresolution and multiorientation wavelet transforms for the segmentation and enhancement of microcalcification clusters and masses [1]-[4] with patents submitted as listed below. The SETI/USF research groups have developed a complementary but new approach that incorporates a rigorous theoretical analysis where the multiresolution wavelet decomposition is followed by a statistical analysis of selected expansion components. The method allows a statistical decision to be made for identifying a region of the mammogram as normal or abnormal and for reasonable estimates of the false positive rate prior to processing. A patent application is planned as listed.

4. Research Progress

4.1 Outline of Method: This study was initiated in August 1994 and completed in June 1996. Briefly, the technique is as follows: For each mammogram considered, a statistical model for the distribution of pixel intensities is computed at different resolutions obtained from wavelet decompositions. Statistics of subregions of the mammogram are compared with a global model for normal tissue. If these subregions have characteristics that deviate substantially from the global model for that mammogram, the deviating regions are marked as possibly abnormal.

4.2 Retrospective Case Study: This study has been confined to the analysis of conventional film mammograms scanned and digitized at a resolution of 35μm, 12 bits. This resolution was chosen because it is close to the anticipated resolution of direct digital x-ray mammography, a technology that is rapidly emerging. At the present time, 30 images from a database of over 100 have been fully analyzed. Of these 30 images, 17 are clinically abnormal and 13 are pathology free (two year follow-up). Thirty folders are on file that contain the full analysis for each of the images; these are available for inspection on request.

4.3 Clinical Evaluation: The results suggest that the proposed approach has real promise. An evaluation of the merit of this method was performed by a resident radiologist based on verified ground truth files. It showed that it is reasonable to expect about a 40-50% identification of normals while keeping the sensitivity of detecting calcification clusters close to 100%. Clearly, there is room for refinements and improvements to the entire procedure, before the method can be adapted to a clinical setting. The intent is to apply this method automatically to full mammographic images with evaluation from extended image databases that contain very subtle microcalcifications and other criteria to ensure difficult "normal" cases are considered, using an extension of methods recently published [5].

5. Documentation of Research Progress

A two part series of papers based on this work has been prepared for submission to *IEEE Transactions on Medical Imaging* [6,7]. (The first pages of these papers are appended.) These papers provide documentation of the theoretical basis for the methods and procedures used along with sample images to illustrate the processing.

6. Attachments

(1) First page of two papers submitted to IEEE Transactions on Medical Imaging.

(2) Copy of slides used by D. Kent Cullers in his presentation on May 8, 1996, at the conference:

Aerospace Medical Association
67th Annual Scientific Meeting
Medical Applications of Space Research and Technology
Atlanta, Georgia, May 6-9, 1996.

7. References

[1] W. Qian, L. P. Clarke, H. D. Li, R. A. Clark, and M. L. Silbiger, "Digital mammography: M-channel quadrature mirror filters for microcalcification extraction." *Comput. Med. Imag. Graphics* 18(5):301-314, 1994.

[2] W. Qian, M. Kallergi, L. P. Clarke, H. D. Li, P. Venugopal, D. Song and R. A. Clark, "Tree structured wavelet transform segmentation of microcalcifications in digital mammography," *Med. Phys.*, 22(8):1247-1254, 1995.

[3] B. Zheng, W. Qian, and L. P. Clarke, "Digital mammography: Mixed feature neural network with spectral entropy decision for detection of microcalcifications," *IEEE Trans. Med. Imag.*, (to be published).

[4] L. Li, W. Qian, L. P. Clarke, and R. A. Clark, "Digital mammography: Directional wavelet analysis for feature extraction and mass detection," *IEEE Trans. Med. Imag.*, (submitted).

[5] M. Kallergi, L. P. Clarke, W. Qian, M. Gavrielides, P. Venugopal, C. Berman, S. D. Holman-Ferris, M. Miller, and R. A. Clark, "Interpretation of calcifications in screen/film, digitized, and wavelet-enhanced monitor displayed mammograms: An ROC study" *Acad. Radiology* 3:285-293, 1996.

[6] J. J. Heine, S. R. Deans, D. K. Cullers, R. Stauduhar, and L. P. Clarke, "Multiresolution statistical analysis of high resolution digital mammograms, Part I: Theory." *IEEE Trans. Med. Imag.* (submitted August 1996)

[7] J. J. Heine, S. R. Deans, D. K. Cullers, R. Stauduhar, and L. P. Clarke, "Multiresolution statistical analysis of high resolution digital mammograms, Part II: Application." *IEEE Trans. Med. Imag.* (submitted August 1996)

8. Patents

1. Method and Apparatus for Analyzing Digital Images. P.I., L. P. Clarke. Patent pending. Filed with U.S. Patent and Trademark Office September 22, 1994. Serial No. 310,708.
2. Computer-Assisted Method and Apparatus for Analysis of X-ray Images. P.I., L. P. Clarke. Filed with U.S. Patent and Trademark Office, June 3, 1996. File No. 96-P-12194.13 HLA/JEH.
3. Method and Apparatus for Recognition of Normal Mammograms in Digital Mammography. P.I., J. J. Heine, S. R. Deans, D. K. Cullers, R. Stauduhar, and L. P. Clarke. In Progress. (To be submitted September 1996)

Multiresolution Statistical Analysis of High Resolution Digital Mammograms Part I: Theory

John J. Heine, Stanley R. Deans, *Senior Member, IEEE*, D. Kent Cullers, Richard Stauduhar, *Member IEEE*, and Laurence P. Clarke, *Member IEEE*

Abstract — **The multiresolution wavelet expansion of digitized mammograms can be analyzed using a parametric statistical model for each image of the expansion. The statistical analysis of the individual expansion components is relatively simple, whereas the analysis of the original image is complicated. An important application of this technique is the statistical modeling of normal tissue in digital mammograms. One possible application of this analysis is to the identification and separation of normal tissue from calcified tissue. The multiresolution probability modeling can be generalized and applied to other digitized medical images, or to any digital image where rigorous statistical evaluation is appropriate.**

This work was supported in part by the National Aeronautics and Space Administration under Grant NASA NCC 2-881, Ames Research Center, Mountain View, California.

John J. Heine is with the Digital Medical Imaging Program, Department of Radiology, University of South Florida, Tampa, FL 33612-4799 (e-mail: heine@rad.usf.edu).

Stanley R. Deans is with the Department of Physics, and Member H. Lee Moffitt Research Center, University of South Florida, Tampa, FL 33620-5700, (e-mail: deans@rad.usf.edu).

D. Kent Cullers and Richard Stauduhar are with the SETI Institute, 2035 Landings Drive, Mountain View, CA 94043, (e-mail: k_cullers@seti-inst.edu, rick@seti-inst.edu).

Laurence P. Clarke is with the Department of Radiology, and H. Lee Moffitt Research Center, University of South Florida, Tampa, FL 33612-4799 (e-mail: clarke@rad.usf.edu).

Multiresolution Statistical Analysis of High Resolution Digital Mammograms Part II: Application

John J. Heine, Stanley R. Deans, *Senior Member, IEEE*, D. Kent Cullers, Richard Stauduhar, *Member IEEE*, and Laurence P. Clarke, *Member IEEE*

Abstract **— A multiresolution statistical method for identifying clinically normal tissue in digitized mammograms is used to construct an algorithm for separating normal regions from potentially abnormal regions; that is, small regions that may contain isolated calcifications. This is the initial phase of the development of a general method for the automatic recognition of normal mammograms. The first step is to decompose the image with a wavelet expansion that yields a sum of independent images, each containing different levels of image detail. When calcifications are present, there is strong empirical evidence that only some of the image components are necessary for the purpose of detecting a deviation from normal. The underlying statistic for each of the selected expansion components can be modeled with a simple parametric probability distribution function. This function serves as an instrument for the development of a statistical test that allows for the recognition of normal tissue regions. The distribution function depends on only one parameter, and this parameter itself has an underlying statistical distribution. The values of this parameter define a summary statistic that can be used to set detection error rates. Once the summary statistic is determined, spatial filters that are matched to resolution are applied independently to each selected expansion image. Regions of the image that correlate with the normal statistical model are discarded and regions in disagreement (suspicious areas) are flagged. These results are combined to produce a detection output image consisting only of suspicious areas. This type of detection output is amenable to further processing that may ultimately lead to a fully automated algorithm for the identification of normal mammograms. A ground truth evaluation of the merit of this method reveals that reasonable predictions of isolated false positives is possible prior to detection, and a specificity of 46% can be maintained while keeping the sensitivity at 100%.**

This work was supported in part by the National Aeronautics and Space Administration under Grant NASA NCC 2-881, Ames Research Center, Mountain View, California.

John J. Heine is with the Digital Medical Imaging Program, Department of Radiology, University of South Florida, Tampa, FL 33612-4799 (e-mail: heine@rad.usf.edu).

Stanley R. Deans is with the Department of Physics, and Member H. Lee Moffitt Research Center, University of South Florida, Tampa, FL 33620-5700, (e-mail: deans@rad.usf.edu).

D. Kent Cullers and Richard Stauduhar are with the SETI Institute, 2035 Landings Drive, Mountain View, CA 94043, (e-mail: k_cullers@seti-inst.edu, rick@seti-inst.edu).

Laurence P. Clarke is with the Department of Radiology, and H. Lee Moffitt Research Center, University of South Florida, Tampa, FL 33612-4799 (e-mail: clarke@rad.usf.edu).

SETI Institute and The University of South Florida at Moffitt Cancer Research Center

Digital Mammography: Multiresolution Statistical Methods for Normal Image Recognition

Richard P. Stauduhar
SETI Institute

D. Kent Cullers
SETI Institute

Stanley R. Deans
Department of Physics USF

Laurence P. Clarke
Department of Radiology USF

John J. Heine
Ph.D. Student USF

Copy of slides used by D. Kent Cullers in his presentation on May 8, 1996, at the conference:
Aerospace Medical Association
67th Annual Scientific Meeting
Medical Applications of Space Research and Technology
Atlanta, Georgia, May 6–9, 1996.

Research Aims

- Development of multiresolution statistical methods to recognize normal mammograms; typically greater than 95% of the screened cases are normal

- Automatic screening of normals or a "second opinion" strategy

- Clinical model: microcalcification detection with emphasis on low false negative (FN) detection rate, as opposed to sensitivity of detection

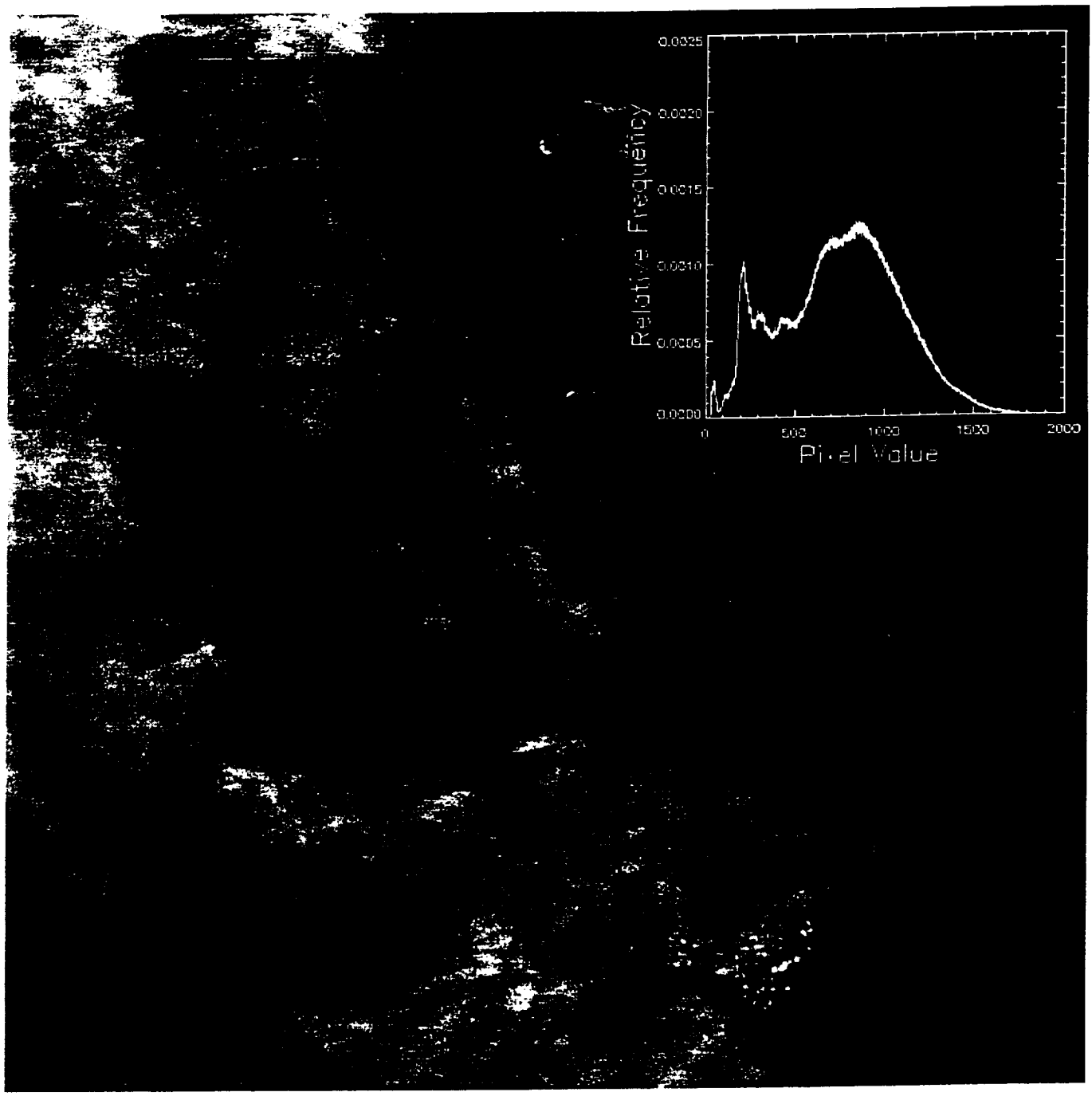
Technology Exchange

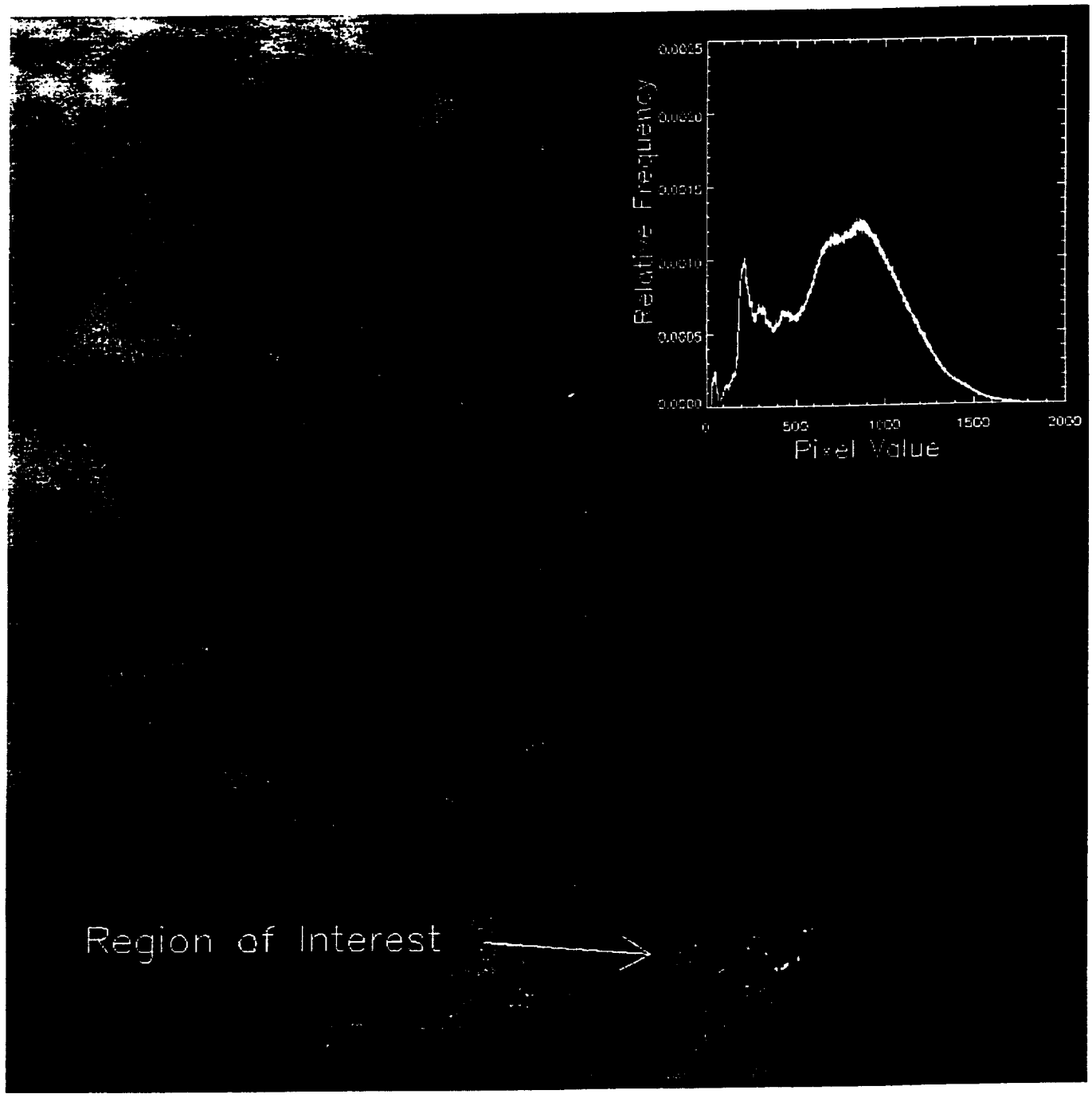
- **SETI:** Advanced Statistical methods currently used in Project Phoenix

- **USF:** Advanced wavelet methodology applied to calcification and tumor detection

- **Moffitt:** Generation of image data base Physician based analysis

START WITH THE RAW IMAGE


0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
Relative Frequency
0
500
1000
1500
2000
Pixel Value


Relative Frequency
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0
500
1000
1500
2000
Pixel Value
Region of Interest

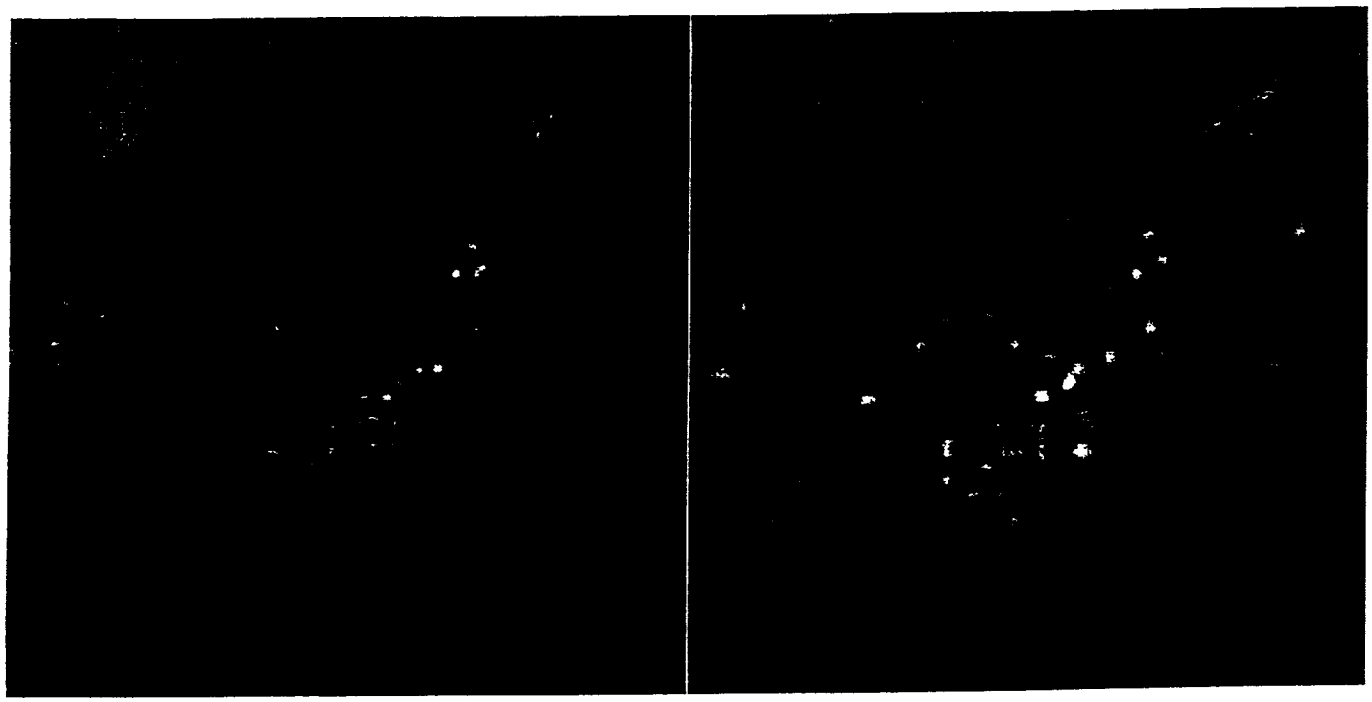
Methods

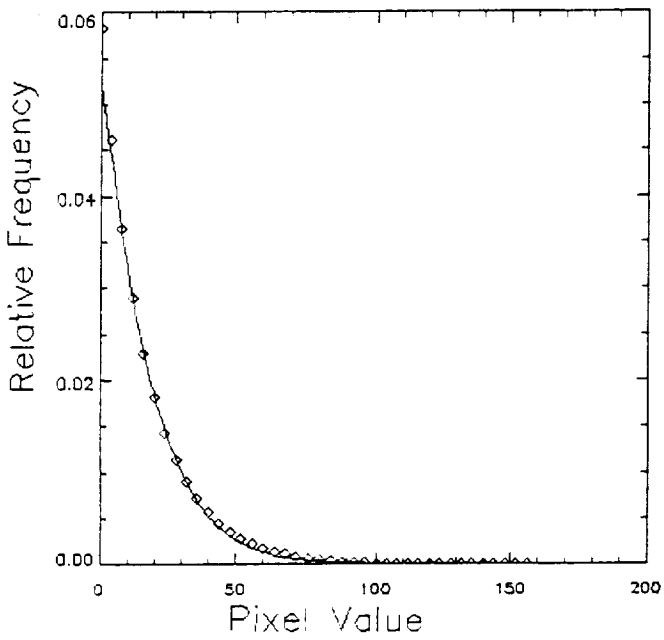
➜ Multiresolution wavelet image decomposition into full size independent frequency subimages

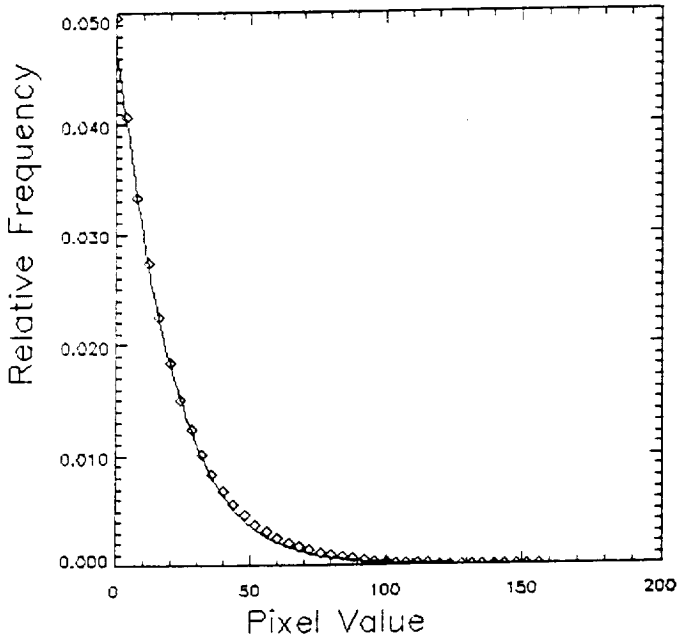
➜ Subimage selection and probability modeling

- Independent detection in select subimages and combination output

- Statistical modeling selectively applied to subimages, as opposed to the raw data, allows a low FN detection rate for a moderate level of sensitivity (projected estimate 50%) for recognition of normal mammograms




Relative Frequency
0.06
0.04
0.02
0.00
0
50
100
150
200
Pixel Value


Relative Frequency
0.050
0.040
0.030
0.020
0.010
0.000
0
50
100
150
200
Pixel Value

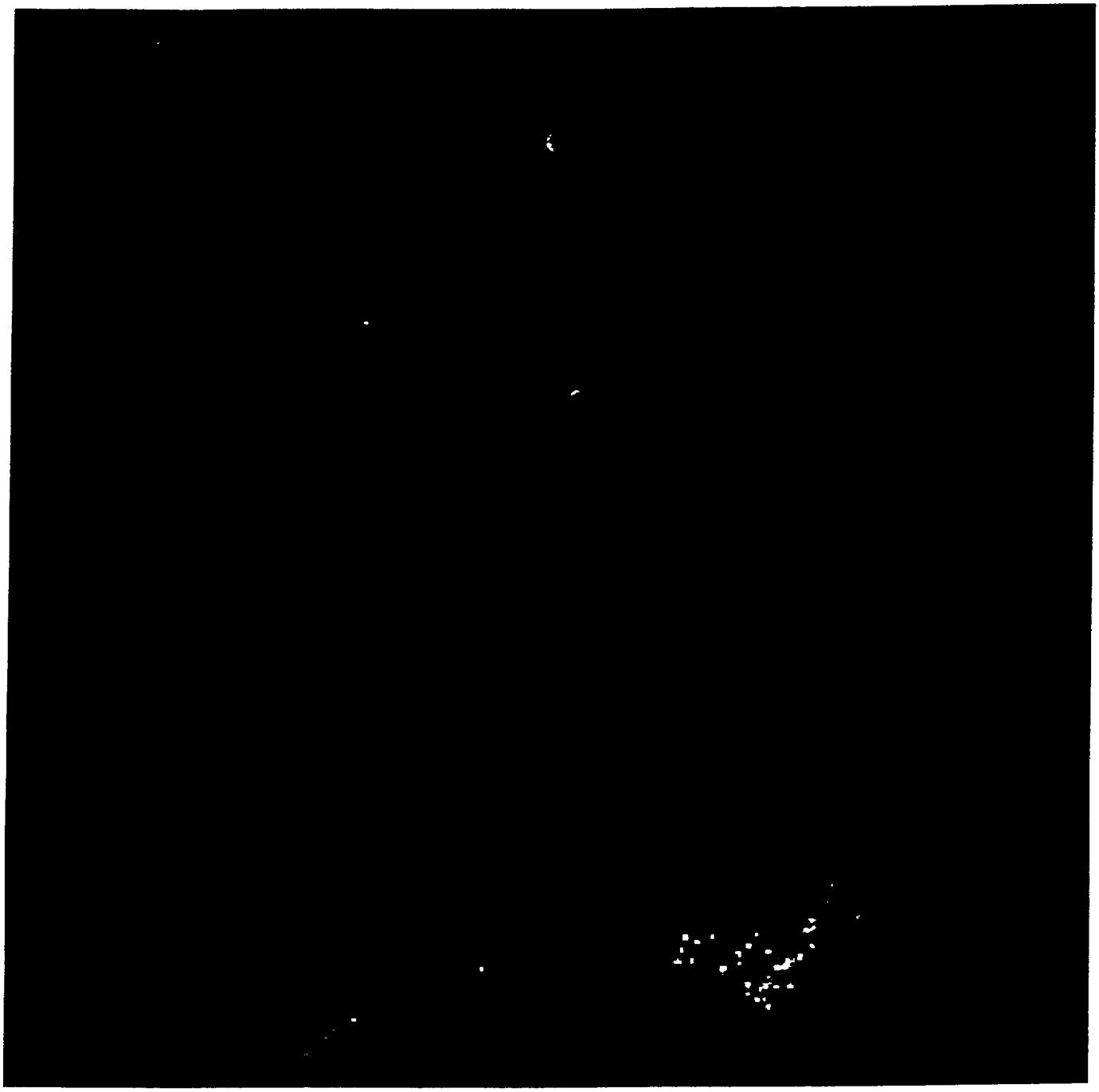
Methods

- Multiresolution wavelet image decomposition into full size independent frequency subimages

- Subimage selection and probability modeling

→ Independent detection in select subimages and combination output

- Statistical modeling selectively applied to subimages, as opposed to the raw data, allows a low FN detection rate for a moderate level of sensitivity (projected estimate 50%) for recognition of normal mammograms



Methods

- Multiresolution wavelet image decomposition into full size independent frequency subimages

- Subimage selection and probability modeling

- Independent detection in select subimages and combination output

→ Statistical modeling selectively applied to subimages, as opposed to the raw data, allows a low FN detection rate for a moderate level of sensitivity (projected estimate 50%) for recognition of normal mammograms

Concluding Remarks

- Multiresolution statistical methods can be readily expanded for other clinical features such as suspicious masses

- The methods compliment a decade long transition to direct digital mammography where more image detail will be present

- Methods compliment parallel research efforts at USF / Moffitt in the development of computer assisted diagnosis (CAD) techniques using wavelet approaches for detection / classification of clinical features as a "second opinion" strategy

Submitted to *IEEE Trans. Med. Imag.* August (1996).

Multiresolution Statistical Analysis of High Resolution Digital Mammograms Part II: Application

John J. Heine, Stanley R. Deans, *Senior Member, IEEE*, D. Kent Cullers, Richard Stauduhar, *Member IEEE*, and Laurence P. Clarke, *Member IEEE*

Abstract — **A multiresolution statistical method for identifying clinically normal tissue in digitized mammograms is used to construct an algorithm for separating normal regions from potentially abnormal regions; that is, small regions that may contain isolated calcifications. This is the initial phase of the development of a general method for the automatic recognition of normal mammograms. The first step is to decompose the image with a wavelet expansion that yields a sum of independent images, each containing different levels of image detail. When calcifications are present, there is strong empirical evidence that only some of the image components are necessary for the purpose of detecting a deviation from normal. The underlying statistic for each of the selected expansion components can be modeled with a simple parametric probability distribution function. This function serves as an instrument for the development of a statistical test that allows for the recognition of normal tissue regions. The distribution function depends on only one parameter, and this parameter itself has an underlying statistical distribution. The values of this parameter define a summary statistic that can be used to set detection error rates. Once the summary statistic is determined, spatial filters that are matched to resolution are applied independently to each selected expansion image. Regions of the image that correlate with the normal statistical model are discarded and regions in disagreement (suspicious areas) are flagged. These results are combined to produce a detection output image consisting only of suspicious areas. This type of detection output is amenable to further processing that may ultimately lead to a fully automated algorithm for the identification of normal mammograms. A ground truth evaluation of the merit of this method reveals that reasonable predictions of isolated false positives is possible prior to detection, and a specificity of 46% can be maintained while keeping the sensitivity at 100%.**

This work was supported in part by the National Aeronautics and Space Administration under Grant NASA NCC 2-881, Ames Research Center, Mountain View, California.

John J. Heine is with the Digital Medical Imaging Program, Department of Radiology, University of South Florida, Tampa, FL 33612-4799 (e-mail: heine@rad.usf.edu).

Stanley R. Deans is with the Department of Physics, and Member H. Lee Moffitt Research Center, University of South Florida, Tampa, FL 33620-5700. (e-mail: deans@rad.usf.edu).

D. Kent Cullers and Richard Stauduhar are with the SETI Institute, 2035 Landings Drive, Mountain View, CA 94043, (e-mail: k_cullers@seti-inst.edu, rick@seti-inst.edu).

Laurence P. Clarke is with the Department of Radiology, and H. Lee Moffitt Research Center, University of South Florida, Tampa, FL 33612-4799 (e-mail: clarke@rad.usf.edu).

I. INTRODUCTION

In Part I of this work [1], we developed a multiresolution statistical analysis method that permits parametric modeling of normal tissue in high resolution digitized mammograms. Here, we illustrate that the previous study naturally leads to a powerful technique that enables the separation of normal regions from potentially clinically suspect regions. Briefly, the previous study shows that the development of a multiresolution statistical model mammograms is possible. The normal tissue model can be used to make comparisons with local image regions: If a small region deviates significantly from the global model it can be flagged as potentially suspicious, and if a region is in agreement it can be discarded. The systematic identification of abnormal regions can be regarded as a detection algorithm, an algorithm that can be tested and evaluated using a standard data base. If no suspicious regions are located, a mammogram lacking any pathology can be identified by the detection process. As before, the term "normal" is used to define tissue regions that do not contain microcalcifications (benign or malignant), calcified veins, or image aberrations, such as small film defects. This specification is used to define the detection task and is not to be confused with the clinical meaning of normal or abnormal tissue. Also, in this work masses are not considered. The guiding premise is that the statistical interpretation of the raw image is rather difficult, but is relatively simple when applied separately to various components of the image following a wavelet expansion.

The ultimate goal is to detect normal mammograms. Since the radiologist spends an enormous amount of time investigating images lacking any malignancy, and the vast majority of mammograms are clinically normal, this approach has the potential for saving valuable time. Also, this method may be viewed as a "second opinion" strategy. In essence, an image that is declared normal by the detection scheme, and then reviewed by a mammographer has been analyzed twice. The desired performance is to detect roughly 40% to 50% of the normal images with a low probability of classifying abnormal images as normal. It should be emphasized that this study deals with calcifications and does not include images with tumors or masses. Clearly, in order to completely solve the problem of identifying normal images this will have to be addressed.

As a result of increased mammographic screening for early cancer detection, considerable effort has been devoted to computer aided diagnosis (CAD) schemes. The work most closely related to our approach utilizes various multiresolution methods for investigating mammograms [2]-[5],[7],[8]. Dengler *et al.* [2] use a difference of two Gaussians for the detection filter, and the final detection is based on a global threshold. Valatx *et al.* [3] generate a smooth approximation of the image with a B–spline expansion and apply a mixed distribution based local thresholding technique to both the raw and approximated image; the output image is formed by subtracting the two thresholded images. A calcification segmentation method is developed by Qian *et al.* [4] using two channel and multichannel wavelet transforms [5], based on subband selection and a rescaling (thresholding) technique for feature detection [6]. Strickland and Hann [7] apply the wavelet transform at full resolution (no downsampling) and detect independently in two sets (HH and LH + HL) of three full resolution subband images. The detection results are combined, further processed, and the inverse wavelet transform is implemented. De Vore *et al.* [8] implement the standard wavelet transform, select the important subbands, and invert the transform after wavelet coefficient suppression. The resulting image is empirically thresholded in order to remove the remaining background information.

Certain aspects the work presented here has similarities to the work referenced above; also, there are marked differences. The main conceptual difference is that the focus of this analysis is on modeling and identifying normal tissue, coupled with flagging regions that deviate from the model as suspicious. Our detection criterion is different too, since we can approximate error rates from knowledge of the distribution function.

II. IMAGE INFORMATION

The images under investigation are film mammograms, digitized at 35 μm per pixel resolution with 12 bit precision, using a Du Pont NDT Scan II Film Digitizer. The data base of over 100 mammograms contains combinations of normals (no pathologies) and abnormals (images with biopsy proven calcification clusters), all with varying parenchymal densities, as described by Kallergi *et al.* [9]. For this study 30 images have been selected from the same patient data base used by Zheng *et al.* [10]: 28 selected at random, and 2 specifically. Two images were deliberately picked because they contain very subtle clusters, and it is important to evaluate the detection performance with limiting cases. For all the selected images large sections consisting of (2048 x 2048) or (1024 x 2048) pixels are used (largest power of 2 region that does not include background). The term "image" refers to these large sections. For reliable statistical analysis it is essential to exclude all regions exterior to the breast. Of the 30 images studied 17 are clinically abnormal, and 13 are pathology free. This means that 2 of 17 ($\approx 12\%$) of the abnormal images in this study are difficult detection cases; this is well above the anticipated number of such cases likely to arise from a large data base.

For demonstration purposes one image section (2048 x 2048) is used to illustrate the various stages of analysis, and this will be referred to as the raw image, see Fig. 1. Our use of these large sections rather than the whole image is for statistical reasons. Certainly, before this method of analysis will be useful in a clinical setting it will be necessary to develop a very accurate method to excise the interior breast region from the exterior background area. This project is currently under development at our laboratory.

III. WAVELET EXPANSION AND PRIMARY STATISTIC

The image domain wavelet expansion is the same as in Part I,

$$f_0 = d_1 + d_2 + \cdots + d_j + f_j \,,$$

where all images in the sum are independent and contain no redundancies. We find empirically that the d_3 and d_4 images are most pertinent for calcification detection at this digital resolution, see Fig. 2 and Fig. 3.

The wavelet transform and subimage selection are similar in some respects to other tried approaches [4],[5],[7],[8] in that the subband images (in the wavelet domain) are selected *a priori*. For example, the transformation application is not like [7] but the selection is similar. Our method is based on using two independent images after wavelet inversion rather than combining the d_3 and d_4 components. Each d_j image is constructed from three wavelet subband images: high-low (HL), low-high (LH), and high-high (HH), at the appropriate level j.

In Part I [1] we showed that a good approximation for the absolute value d_j image is given by

$$p(x;c) = \frac{1}{c}\exp\left(-\frac{x}{c}\right), \quad x \geq 0,$$

where x is an arbitrary pixel value and c is a constant. It is implied that both x and c depend on j.

Following maximum likelihood arguments [11] for independent samples of the variable x, the parameter c can be estimated by the average value of x,

$$c = \langle x \rangle .$$

We have assumed that the samples of x are independent; this is certainly not the case but, does not seem do be a serious detraction. The empirical and theoretical probability densities for the d_3 and d_4 images are shown in Fig. 4.

By considering the size of the image compared to the number pixels contained in a cluster it follows that the cluster has a minimal effect on the global statistic. (There are roughly 5×10^6 pixels in the image and about 2000 pixels for an average calcification cluster.) Therefore, the primary statistic can be considered as the model for normal tissue. This statistic is useful for developing robust statistical tests.

IV. HYPOTHESIS TESTING

Application of the Neyman Pearson lemma [11] leads to a robust statistical test based on knowledge of the primary statistic. [In the following discussion N is the size, ($8 \times 8 = 64$) or ($16 \times 16 = 256$), of a small region of the d_j subimage.] For N samples of the random variable X with pdf given by $p(x;c)$, the likelihood function is defined as

$$L(x;c) = \prod_{i=1}^{N} p(x_i;c) .$$

This is the joint pdf for N independent samples of the variable X; again, the correlation between successive samples of X is ignored. A test can be derived from evaluating the possibility that $c = c_o$ (for normal region) against the alternative that $c = c_a$ (for abnormal). The parameter c_o is associated with the global or normal statistic of the d_j subimage. By implementation of a procedure, known as the null hypothesis, a hypothesis is set up to see if it can be rejected. The test, commonly referred as the likelihood ratio, is given by

$$\ell = \frac{L(x;c_o)}{L(x;c_a)} < \kappa ,$$

where $c_a > c_o$ and κ is a constant to be determined. This is the ratio of the joint pdfs, or likelihoods, and results in two alternate choices: (1) accept the region as normal (accept the null hypothesis) if the ratio is not too small; or (2) reject the null hypothesis and assume the region is suspicious. If the ratio is small the probability is greater of rejecting the null hypothesis. The rejection criterion must be determined and is addressed in the next section. Applying this test specifically to the d_j image primary statistic results in

$$\left(\frac{c_a}{c_o}\right)^N \exp\left[-\left(\frac{1}{c_o} - \frac{1}{c_a}\right)\langle x \rangle\right] < \kappa .$$

The natural logarithm of this expression, followed by some rearranging, gives

$$\langle x \rangle > \left(\frac{c_a c_o}{c_a - c_o} \right) [- \log \kappa - N \log(c_o / c_a)] .$$

The quantity on the right side of this equation is a positive constant, designated by γ. Thus the discriminating test is

$$\langle x \rangle > \gamma .$$

The hypothesis $c = c_o$ is rejected (the region is not normal) if $\langle x \rangle$ is too large. Clearly, there are two types of errors involved with this decision: (1) decide to reject the null hypothesis and assume that the region is suspicious when it is not. In classical detection theory this is known as a false alarm and is analogous to the standard medical imaging false positive (FP) error; or (2) decide to accept the null hypothesis and consider the region as normal when it is not. In medical imaging terminology this is a false negative (FN). The value of γ determines the FP rate. In order to select this value, and thus set the desired FP rate, the sample distribution for the parameter c must be found.

It is important to emphasize that the likelihood ratio test gives an analytical method for comparing image regions against some global criterion. However, the approach does not reveal the spatial extent of the comparison; this must come from empirical evidence.

V. SUMMARY STATISTIC AND ERROR RATES

The summary statistic is established as described in Part I [1], where the gamma distribution designated by $g(c; \alpha, \beta)$ is computed for the d_3 and d_4 images. This procedure results in a new image reduced in both spatial dimensions by a factor of 8 or 16 for d_3 or d_4, respectively. The normalized histogram of the reduced image is the empirical pdf for c.

The theoretical and empirical pdfs for the reduced images are illustrated in Fig. 5. An error of the first kind or FP rate can be estimated from this pdf prior to detection processing. Again, it is assumed that the calcified regions have a minimal effect on this distribution and can be considered as outliers located in the far right tail region. The FP rate (the fractional number false calcifications per image) can be obtained by

$$P_f = \int_{\tau}^{\infty} g(c; \alpha, \beta)\, dc ,$$

where τ denotes the threshold. This equation deserves special consideration. The test criterion given by $\langle x \rangle > \gamma$ is obtained as follows: (1) select $\tau = \gamma$, (2) pick a value for P_f, (3) solve this equation for τ. The total expected number of false positives in the entire d_j image can be approximated by

FP(total) = $P_f \times$ (number of pixels in reduced image).

It should be emphasized that this is an estimation that may be obtained as an average after processing many images. The intriguing aspect is that the false positive rate P_f can be set prior to detection. However, this does not completely specify the error prediction, since no preprocessing estimation can be made concerning the error of the second kind or FN rate. The FN rate follows from the calcification distribution which is unknown.

VI. LOCALIZED NORMAL TISSUE RECOGNITION

The detection technique is implement by shifting a 8×8 or 16×16 pixel search window through the d_3 and d_4 images, respectively. A detection flow diagram that illustrates the various stages of processing is given in Fig. 6. The intent is to match the search window size to the average spatial extent of the calcifications that may exist in each subimage. When the spatial extent of the wavelet function and calcifications are similar the response (in the d_j image) is maximized, and the area is flagged as suspicious.

For early cancer detection calcifications with spatial extent less than ≈ 0.5 mm are most important for clinical diagnosis. This corresponds to calcifications ranging roughly from 16 pixels to 3 or 4 pixels in diameter (≈ 0.1 mm), and the search window is matched to this scale. We assume that calcifications smaller than this are not discernable. The window is shifted with a 50% overlap in both spatial dimensions during the search; this is to reduce the risk of missing a feature (calcification). When a region is assumed normal (accept the null hypothesis) it is set to zero. If the null hypothesis is rejected (accept as suspicious) the region is left intact. This is how potentially small calcified regions are detected by default. Following the independent detection the images are combined, and the total detected image results as the output. In effect the dual output combination can be viewed as a mask. From this any d_j image combination or even the raw image can be returned as the output image. This can be accomplished by making the total output image into a binary image (ones or zeros) and simply multiplying by the desired type of output. This is important if further processing is desired because the calcified regions can be returned with full resolution and detail. For this demonstration the detection output given from the sum of the first 5 d_j images. The detection results are illustrated in Fig. 7.

The detection scheme takes into account that calcifications have a spatial extent or connectivity quality: a calcification appears as a clump of large pixel values and normal regions have a diffuse distribution of pixel values. The regions considered are matched to spatial scale by adjusting the search window size accordingly. The window size and shift increment are a compromise. If the window is greater than half the size of the feature and the shift greater than half the window size there is a possibility that the feature will be missed. Assume that the smallest object to be detected has a spatial extent M, a $2M \times 2M$ window must be used and shifted with an increment of M to insure that the object will not be missed. The window is most sensitive when the feature fills it entirely, which is not generally expected here. An alternative method would be to scan the image with the limiting window size with single pixel increments. But, this may add many false positives to the outcome.

Although not apparent there is redundancy built into this detection scheme. This can be assessed by looking at the d_3 and d_4 detected images prior to recombination, see Fig. 8 and Fig. 9. There are flagged regions in the d_3 image that are not flagged in the d_4 image and vice versa. This indicates that the wavelet response to the feature was stronger in the respective image. However, in some regions the test is triggered in both images at roughly the same spatial location. This indicates that the wavelet functions response is similar in both images (relative to the background and window size). Thus some calcified regions have the possibility of being detected in both images and represents a redundancy. This can be viewed as a safety measure.

The focus of this detection scheme is very localized. However, a possible sign of early cancer is the presence of a microcalcification cluster, and this is of more clinical concern than

isolated events. A single cluster is defined roughly as 3 to 5 microcalcifications assembled within a square centimeter (cm). This definition implies that one or two calcifications within a square cm are not clinically important. In order to reach the goal of recognizing images that are normal from the clinical point of view clearly requires another stage of processing.

The additional stage of processing is needed to eliminate FP normal diffuse regions on the order of a square cm. The FP normal diffuse region may result from isolated events within a square cm proximity: two PF calcifications (flagged regions that are normal) and one true calcification (a correctly flagged region); or vice versa; or three FP regions.

VII. EXPERIMENT, ANALYSIS, AND EVALUATION

A. Experiment

The localized normal region detection must be conducted such that the FN and FP rates are optimized. These rates are in opposition in that decreasing one causes an increase in the other. In terms of the threshold, if τ is set low enough the FN rate can be reduced to zero, but then the FP rate is high. So, the problem is to adjust the threshold. We want the threshold as high as possible while keeping the FN rate essentially zero. This optimum value can be found by probing the detection operating characteristics. We do this by processing the images 5 times, each time with a slightly higher threshold, or equivalently, a lower value for the FP rate, P_f.

B. Analysis Method

The evaluation of the local area detection method for each of the trials was performed by a resident radiologist using three figures of merit. [First, for clarification, a true positive (TP) with respect to an isolated calcification is defined as: a calcified region, benign or malignant, that has not been set to zero. The TP cluster follows from this definition also.] The figures of merit are: (1) the isolated FP calcifications per image; (2) the number of TP clusters; and (3) the number of FN clusters. The cluster analysis is based on the biopsy verified ground truth files, and the results are presented as averages. There are many methods used for counting clusters; consequently, the technique used here requires a brief explanation.

Following from the definition of a cluster (as defined previously), if 3 events are located (this includes FPs or TPs) within a square cm the region is classified as a cluster. If the nearest neighbor calcifications of two different clusters are within a cm in either the horizontal or vertical direction the total cluster is counted as one; this is sometimes defined as a diffuse cluster situation, and it admits the possibility of chaining clusters together.

C. Tabulated Results

The 5 sets of detection results are shown in Table I. Each trial corresponds to a different threshold τ or $P_f(\tau)$. The thresholds corresponding to the 5 trials are arranged so that $\tau_1 < \tau_2 < \cdots < \tau_5$, and the corresponding values of $P_f \times 10^4$ are in the last column. The goal is to identify the τ where the experimental value of the sensitivity begins to drop below 100%. In this table the following definitions are used:

Specificity = (Number of normals correctly classified)/(Total number of normals),

and

Sensitivity = (Number of clusters found)/(Total number of abnormals).

Table I
Evaluation of each of five trials

Trial	Specificity %	Sensitivity %	FP clusters/image	$P_f(\tau) \times 10^4$
τ_1	15	100	1.20	6.00
τ_2	15	100	1.36	3.00
τ_3	46	100	0.93	2.50
τ_4	46	94	0.67	1.00
τ_5	92	89	0.13	0.05

In going from trial 1 to trial 2 there is no measurable change in the evaluation. This means τ was not changed enough. The parameters associated with trial 3 are the best. since it is possible to keep the sensitivity at 100%, and still identify 46% of the normals.

An estimate for the theoretical maximum number of isolated (individual) calcifications per image can be found by the formula

$$\text{Max} = 2 \times P_f \times (18 \times 256 \times 256 + 12 \times 256 \times 128)/30$$

and the minimum is given by: Min = Max/2. This formula comes from considering that there are 2 reduced images for each raw image and there are two possible sizes of reduced images (256×256) or (256×128). There are 18 large images and 12 small images. The Max is two times the Min because both d_3 and d_4 can contribute to detection, and it is possible to have no overlapping error in each image. These results are summarized in Table II.

Table II
Theoretical limits for each trial.

Trial	Max	Min	Actual Counts	$P_f(\tau) \times 10^4$
τ_1	63.92	31.46	44	6.00
τ_2	31.46	15.73	31	3.00
τ_3	26.20	13.10	18	2.50
τ_4	10.48	5.24	13	1.00
τ_5	0.52	0.26	3	0.05

D. Observations

The specificity rates (Table I) are very encouraging, since a feasible operating P_f can be found. In this case it is trial 3, and only one very subtle detection case is missed in trial 4. This indicates we can hope to operate at 100% sensitivity while identifying 46% of the normal images. The theoretical isolated FP rates (Table II) are in general agreement with the counted data. As the FP rate is reduced the agreement diverges somewhat because the integral required to find the FP rate is only an approximation. The final estimates are good order of magnitude results. These evaluation results indicate that the detection method behaves as predicted, and gives credence to the statistical modeling. If the model was merely a crude approximation it is quite likely the detection results would not be in such close agreement.

VIII. CONCLUSIONS

A cursory examination of mammograms may indicate that the underlying statistics are irregular, and parametric modeling is most likely not a tractable approach. However, this study in conjunction with the previous work [1] clearly indicates otherwise. The multiresolution statistical analysis allows for the parametric modeling of the information needed for the recognition of normal tissue and detection of abnormal regions. In the vast majority of images studied, the primary and summary statistic appear to be dependable estimators for the detection scheme. The intriguing aspect of this analysis is that the detection procedure follows from theoretical calculations derived from the primary statistic. Also, estimates of the FP rate can be set ahead of time. In essence, the technique merges two powerful analysis techniques: classical signal detection theory, and multiresolution decomposition.

The detection process was illustrated with the symmlet basis [1]. Other wavelet bases can be used for comparison purposes to optimize the choice of bases. Thresholds can be set the same and the experiment repeated. The 30 images are a fair representation of a clinical mammography data base. Thus we have the potential to quantify the "best basis" for usage in mammography.

The evaluation results provide a strong impetus for further pursuit and analysis of the multiresolution statistical technique. An automated method must be developed to segregate the isolated calcifications that do not belong to a cluster, and it still remains to find a reliable technique to segment the breast from the background. The breast background boarder region must be delineated, and the statistical analysis constrained to the interior region.

Detection of diffuse normal areas is an important point to consider when attempting to reach the goal of recognizing images that are clinically normal. However, the local region detection part of the algorithm is the foundation of the technique: if this fails (large FN rates) any ensuing stage of proceeding will naturally fail. Another stage of detection based on recognizing normal properties of larger normal areas is under development.

ACKNOWLEDGMENT

It is a pleasure to thank M. Kallergi, and R. P. Velthuizen of the Department of Radiology and Moffitt Research Center, at the University of South Florida for helpful discussions regarding this manuscript and digital mammography in general. Special thanks go to Priya Venugopal, M.D., for reading the mammograms and providing the information needed for evaluation.

References

[1] J. J. Heine, S. R. Deans. D. K. Cullers, R. Stauduhar, and L. P. Clarke, "Multiresolution statistical analysis of high resolution digital mammograms, Part I: Theory," *IEEE Trans. Med. Imag.*, vol. , no. . pp. , 199 .

[2] J. Dengler, S. Behrens, and J. F. Desaga, "Segmentation of microcalcifications in mammograms," *IEEE Trans. Med. Imag.*, vol. 12, no. 4, pp. 634-642, 1993.

[3] L. Valatx, I.E. Magnin, and A. Brémond, "Automatic microcalcification and opacities detection in digitized mammograms using a multiscale approach," in *Digital Mammography*, A. G. Gale, S. M. Astley. D. R. Dance. and A. Y. Cairns, Eds. Amsterdam: Elsevier Science B. V., pp. 51-57, 1994.

[4] W. Qian, L. P. Clarke, M. Kallergi, and R. A. Clark. "Tree structured nonlinear filters in digital mammography." *IEEE Trans. Med. Imag.*, vol. 13, no. 1, pp. 25-36, 1994.

[5] W. Qian, M. Kallergi, L. P. Clarke, H.-D. Li, P. Venugopal, D. Song, and R. A. Clark, "Tree structured wavelet transform segmentation of microcalcifications in digital mammography," *Med. Phys.*, vol. 22, no. 8, pp. 1247-1254, 1995.

[6] W. Qian and L. P. Clarke, "Wavelet-based neural network with fuzzy-logic adaptivity for nuclear image restoration." *Proc. IEEE*, vol. 84, no. 10, pp. , 1996.

[7] R. N. Strickland and H. I. Hann, "Wavelet transforms for detecting microcalcifications in mammograms," *IEEE Trans. Med. Imag.*, vol. 15, no. 2, pp. 218-229, 1996.

[8] R. A. DeVore, B. Lucier, and Z. Yang, "Feature extraction in digital mammography," in *Wavelets in Medicine and Biology*, A. Aldroubi and M. Unser, Eds. Boca Raton, FL: CRC Press, pp. 145-161. 1996.

[9] M. Kallergi, K. Woods, L. P. Clarke, W. Qian, and R. A. Clark, "Image segmentation in digital mammography: Comparison of local thresholding and region growing algorithms," *Comput. Med. Imag. Graphics* vol. 16, no. 5, pp. 323-331, 1992.

[10] B. Zheng, W. Qian, and L. P. Clarke, "Digital mammography: Mixed feature neural network with spectral entropy decision for detection of microcalcifications," *IEEE Trans. Med. Imag.*, (to be published).

[11] W. Mendenhall and R. L. Scheaffer, *Mathematical Statistics with Applications*, North Scituate, MA: Duxbury. 1973.

Figure Captions

Fig. 1. The raw image 2048 x 2048 pixels scaled by a factor of 2/5 for viewing purposes. The arrow points to the region containing the biopsy proven cluster.

Fig. 2. The d_3 image. The negative of the image is illustrated for better viewing purposes. Normally the calcifications are bright (positive biased and large intensity values).

Fig. 3. The d_4 image.

Fig. 4. The empirical (solid) and theoretical (diamond) pdfs for the d_3 (left) and d_4 (right) images. The plot is displayed in this fashion for clarification due to the close theoretical and empirical agreement.

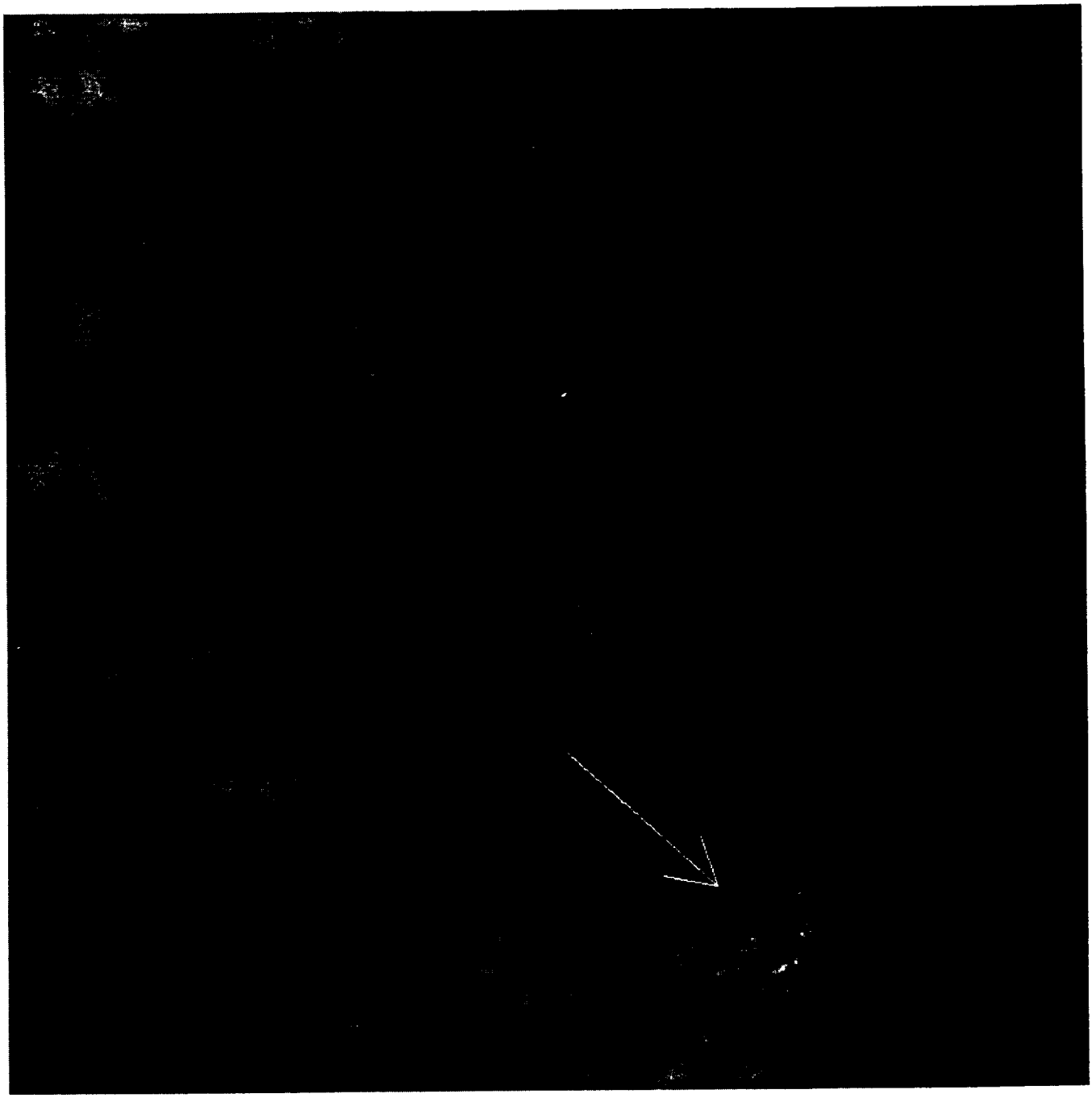
Fig. 5. The empirical (solid) and theoretical (dash) summary pdfs for the d_3 (left) and d_4 (right) images.

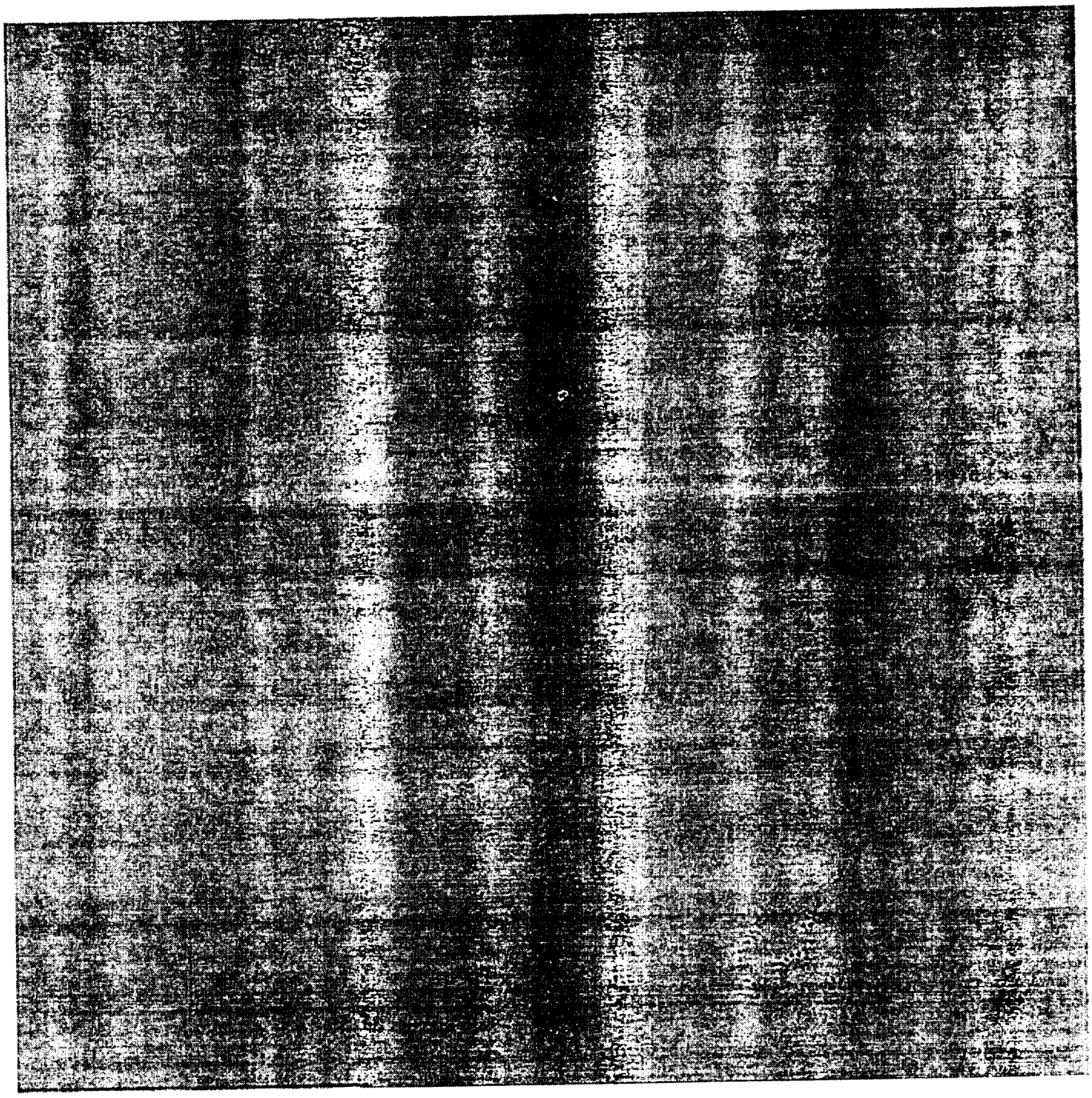
Fig. 6. Detection flow chart.

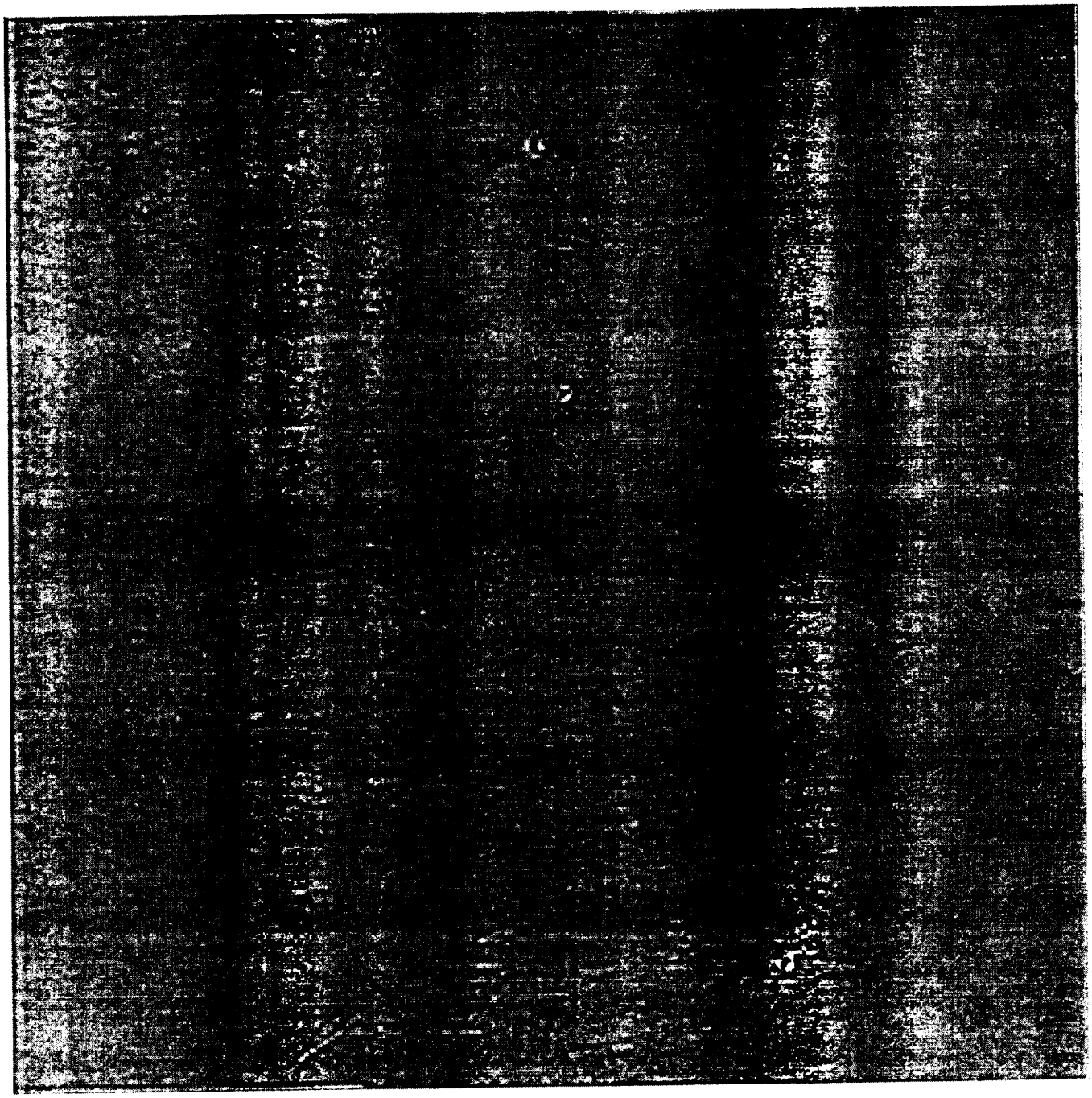
Fig. 7. The total combined detection.

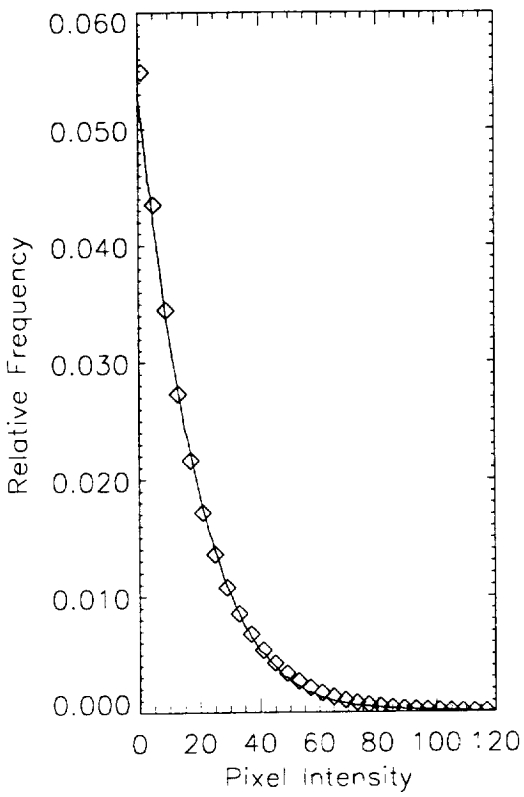
Fig. 8. The d_3 detected image.

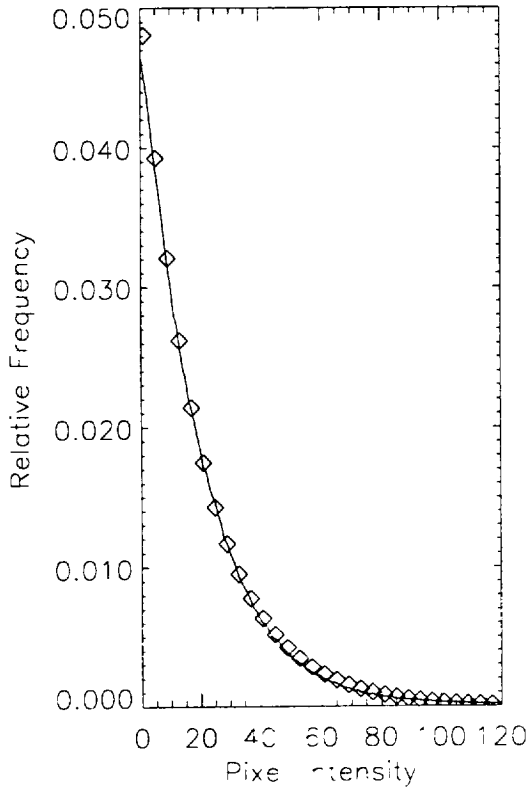
Fig. 9. The d_4 detected image.

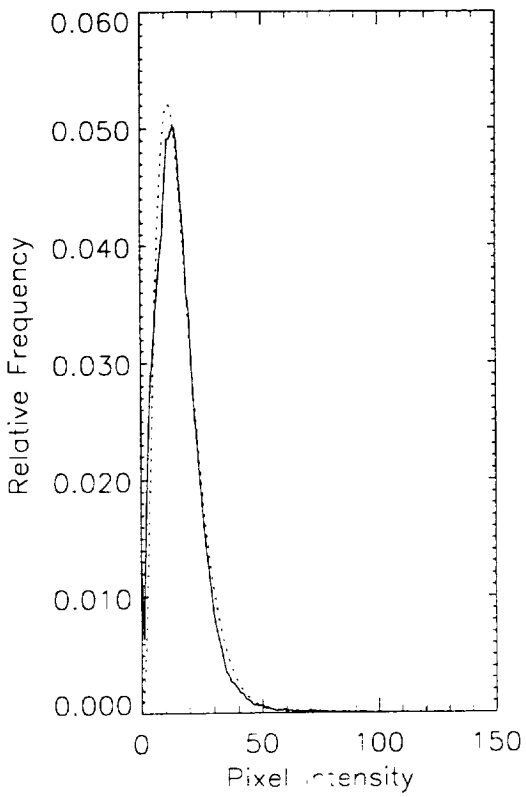


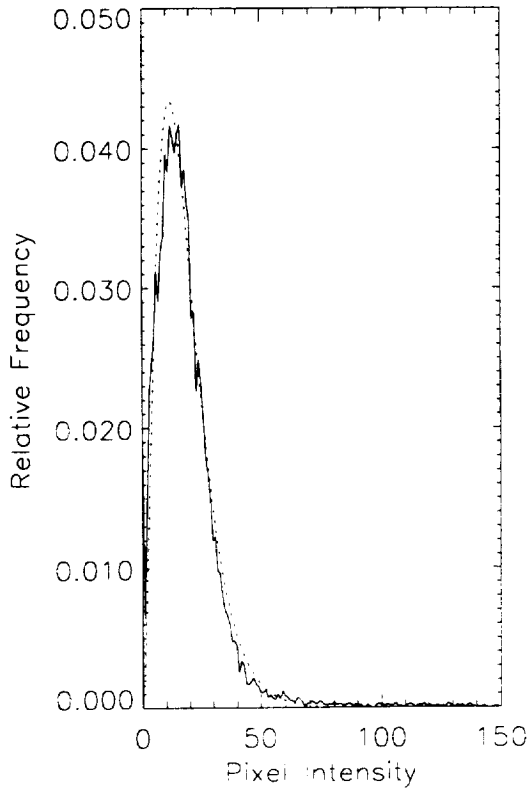


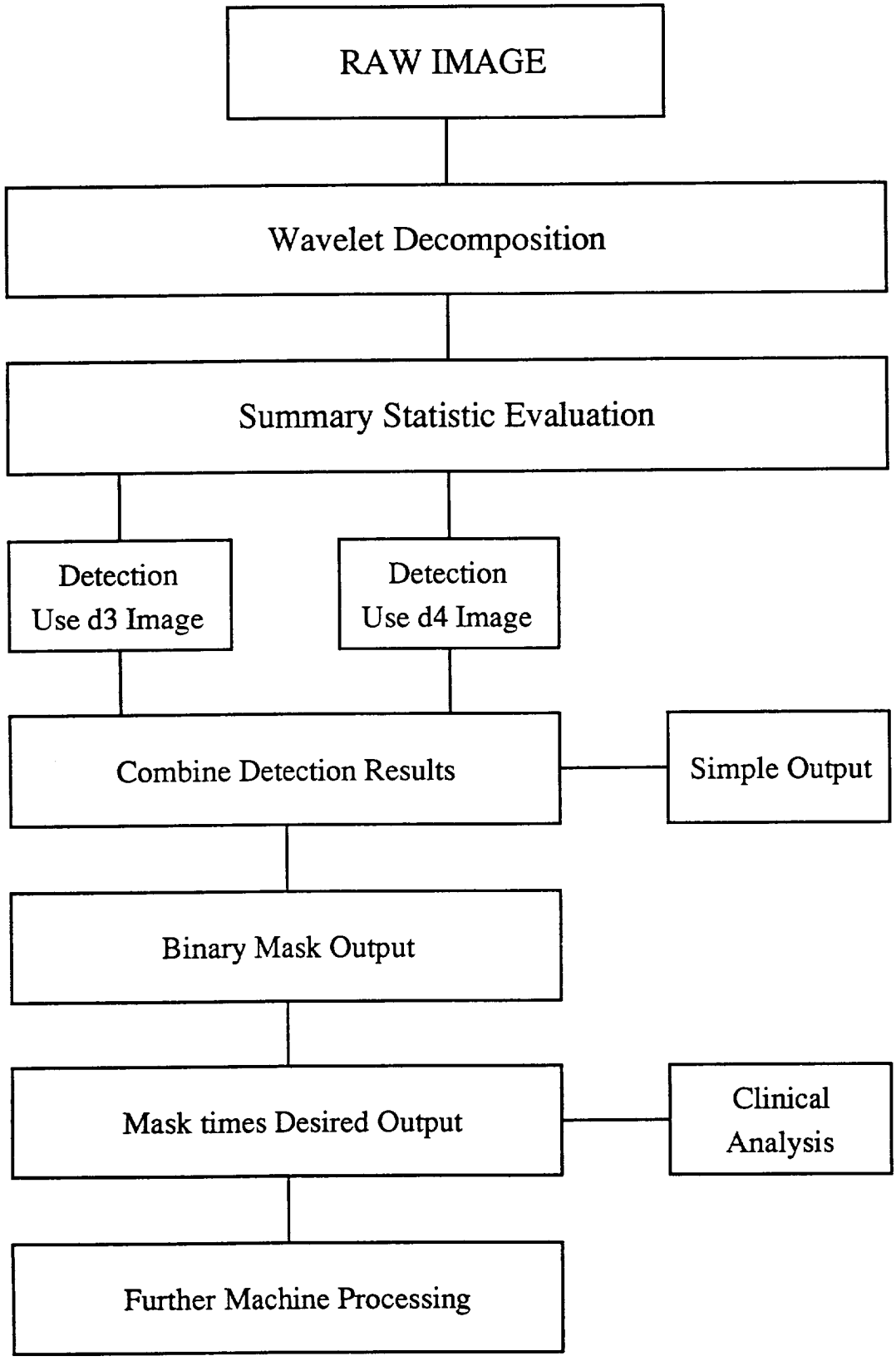



0.060
0.050
0.040
0.030
0.020
0.010
0.000
Relative Frequency
0 20 40 60 80 100 120
Pixel Intensity


0.050
0.040
0.030
0.020
0.010
0.000
Relative Frequency
0 20 40 60 80 100 120


0.060
0.050
0.040
0.030
0.020
0.010
0.000
Relative Frequency
0
50
100
150
Pixel Intensity


0.050
0.040
0.030
0.020
0.010
0.000
Relative Frequency
0
50
100
150
Pixel Intensity


RAW IMAGE
Wavelet Decomposition
Summary Statistic Evaluation
Detection
Use d3 Image
Detection
Use d4 Image
Combine Detection Results
Simple Output
Binary Mask Output
Mask times Desired Output
Clinical
Analysis
Further Machine Processing