

# Automated analysis of the American College of Radiology mammographic accreditation phantom images

Kenneth W. Brooks<sup>a)</sup>

*The Department of Radiation Oncology, The Emory Clinic, 1365 Clifton Road, Atlanta, Georgia 30322*

Jon H. Trueblood

*Medical College of Georgia, Department of Radiology, Medical Physics Section, 1459 Laney Walker Boulevard, Room AE-2018, Augusta, Georgia 30912-3950*

Kimberlee J. Kearfott

*Department of Nuclear Engineering, University of Michigan, 108 Cooley Lab, North Campus, Ann Arbor, Michigan 48109-2104*

Daryl T. Lawton

*Digital Imaging and Graphics Group, Microsoft Corporation, Redmond, Washington 98052*

(Received 27 January 1995; accepted for publication 11 February 1997)

A significant metric in federal mammography quality standards is the phantom image quality assessment. The present work seeks to demonstrate that automated image analyses for American College of Radiology (ACR) mammographic accreditation phantom (MAP) images may be performed by a computer with objectivity, once a human acceptance level has been established. Twelve MAP images were generated with different x-ray techniques and digitized. Nineteen medical physicists in diagnostic roles (five of which were specially trained in mammography) viewed the original film images under similar conditions and provided individual scores for each test object (fibrils, microcalcifications, and nodules). Fourier domain template matching, used for low-level processing, combined with derivative filters, for intermediate-level processing, provided translation and rotation-independent localization of the test objects in the MAP images. The visibility classification decision was modeled by a Bayesian classifier using threshold contrast. The 50% visibility contrast thresholds established by the trained observers' responses were: fibrils 1.010, microcalcifications 1.156, and nodules 1.016. Using these values as an estimate of human observer performance and given the automated localization of test objects, six images were graded with the computer algorithm. In all but one instance, the algorithm scored the images the same as the diagnostic physicists. In the case where it did not, the margin of disagreement was 10% due to the fact that the human scoring did not allow for half-visible fibrils (agreement occurred for the other test objects). The implication from this is that an operator-independent, machine-based scoring of MAP images is feasible and could be used as a tool to help eliminate the effect of observer variability within the current system, given proper, consistent digitization is performed. © 1997 American Association of Physicists in Medicine. [S0094-2405(97)00205-8]

Key words: mammography, phantom images, image processing, medical imaging

## I. INTRODUCTION

An integral portion of the American College of Radiology (ACR) mammography accreditation program (MAP) is the phantom image quality test, which alone accounts for 25% of the MAP's failures. An additional 10% fail due to both the phantom image quality criterion and a clinical mammogram criterion.<sup>1</sup> The MAP phantoms contain clinically relevant test objects designed to represent typical breast pathologies: fibrils (or fibers), microcalcification groups (or specks), and nodules (or masses). The objects decrease in size and contrast sufficiently to demonstrate a visibility threshold using a typical clinical image technique and viewing apparatus. For certification, a facility's phantom image is analyzed independently by three American Board of Radiology (ABR) certified medical physicists who are experienced in mammography quality control and trained in reading phantom images.

The criteria used by the readers has evolved to include a system of partial scores commensurate with the perceived object visibility as well as a system of deductions for image artifacts. The three independent object visibility net scores are averaged and must demonstrate visibility for at least four fibers, three microcalcification groups, and three masses.<sup>2</sup> A requirement for using phantoms to monitor image quality is nonvariable, consistent viewers.<sup>3</sup> However, there are two sources of variability in the MAP process which could lead to inconsistent accreditation failure. First, the perceptive variability among medical physicists which has been assessed in previous research.<sup>4-6</sup> Second, variabilities in the phantom manufacturing processes have also been documented in the literature<sup>5,7</sup> and may be compounded onto the viewer inconsistency. Ultimately the phantom image quality evaluation test for accreditation, as well as the phantom

manufacture itself, involve human variabilities, the ramifications of which may not be acceptable in a national mammography program.

The authors have previously proposed a computer-vision phantom image analysis approach to mitigate these effects.<sup>6</sup> Generally such approaches perform poorly, compared to humans, when interpreting complex scenes requiring higher-order human skills associated with memory and the visual cortex.<sup>8</sup> However, the phantom image test objects are standard geometric shapes on a uniform background which, in comparison to actual patient mammograms, are simple and suggest that perhaps an automated, computer-vision system for evaluating the phantom images can match human acceptance levels for test objects in mammography phantom images. Similarly, other researchers<sup>5</sup> have demonstrated the success of statistical methods for comparing each facility's accreditation image against a standard image. In their study, relative contrast of the two largest microcalcification groups and three largest nodules was measured on MAP images of varying quality and a linear, least-squares relationship to a high-quality reference image was assumed. Though their quantitative image measurements did not include all of the test objects in the MAP phantom image (three of the five nodules and two of the five microcalcification groups were utilized), the human observer responses for these targets were well predicted by the quantitative, linear relationship with much less overall variance than the human observers.

We chose another approach to quantitative phantom images analysis. The present work stemmed from earlier observer experiments which demonstrated variability in the MAP phantom evaluation process.<sup>4</sup> The complete problem of scoring image artifacts was not included in this study. However, all test object groups and all sizes were included in an effort to encounter the human visibility threshold. We hypothesize that if this human visibility contrast threshold is quantified, then a visibility decision based on it may be performed by an autonomous computer algorithm. Since mammography is predominantly performed via x-ray film, transparency film digitization devices, appropriate for the phantom evaluation process, have been analyzed and successful technology identified.<sup>9</sup> A description of the phantom and imaging systems used are provided in the next section. Also, the human observer MAP phantom image experimental design and threshold contrast results are summarized. Image processing algorithms which were developed to automatically locate all MAP phantom image test objects and predict their visibility relative to threshold contrast are described and compared to human observers for these tasks.

## II. METHODS

### A. Phantom

A commercially available breast phantom [Radiation Measurement Incorporated (RMI) Model-156 Breast Phantom, Middleton, WI 53562] which meets MAP standards was used. This standard mammographic phantom (SMP) is constructed of a 10 cm×10 cm×4.5 cm thick acrylic block with a removable, tissue-equivalent wax insert in one face

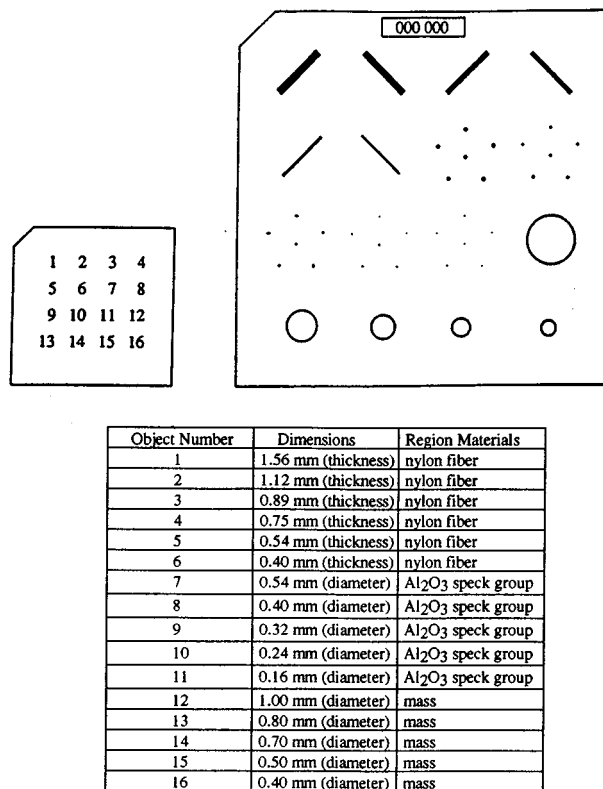


Fig. 1. Schematic diagram of the mammographic accreditation phantom (MAP) showing the locations and relative sizes of the features.

(serial #312 156 type 4 phantom). The insert is embedded with various sizes of nylon fibers to simulate soft-tissue edges, aluminum oxide particles to simulate microcalcifications, and water-density masses to simulate tumors.<sup>10</sup> These objects represent common breast pathologies and are present in sizes that range from being easily visible to invisible in the phantom film image. The wax insert is 0.4 cm thick and contains the fibers, microcalcification groups, and masses as depicted in Fig. 1.

### B. Mammographic systems

Two dedicated mammography units were used. The General Electric (Senograph 600T Series HF Mammography Machine, General Electric, 92137 Issy les Moulineaux, France) and the LoRad (M-II Mammography Machine, LoRad Medical Systems, Danbury, CT 06811) machines are both equipped with Molybdenum anodes and filtration. The General Electric machine has a 0.3 mm focal spot size and a 65 cm source-to-image distance was chosen. The LoRad has a 0.3 mm focal spot size and a fixed 50 cm source-to-image distance. All images were taken using standard 18×24 cm mammography film (Min R E, Kodak Company, Rochester, NY 14445) with a mammographic screen (Min R, Kodak Company, Rochester, NY 14650) and developed with the same undedicated darkroom film processor (Kodak RP X-omat automatic processor, Kodak Company, Rochester, NY 14650). A set of eleven representative phantom films were selected from 50 films which were generated using

TABLE I. System configurations and exposure techniques for the observer experiments.

Film No.	X-ray unit	kVp	mAs	Optical density at film center
1	LoRad	35	6	1.10
2	LoRad	30	13	1.21
3	LoRad	26	40	1.44
4	LoRad	28	12	1.62
5	SRM <sup>a</sup>	20	600	2.51
6	General Electric	30	115	0.99 <sup>b</sup>
7	General Electric	30	82	0.66 <sup>b</sup>
8	General Electric	22	400	1.11
9	General Electric	26	125	1.25
10	General Electric	35	12	0.61
11	General Electric	26	82	0.96
12	General Electric	26	74	0.69

<sup>a</sup>SRM=specimen radiography machine.

<sup>b</sup>Includes 1.3 cm scattering media placed on top of phantom.

these two machines and various techniques. The main selection criteria used was the background optical density in the center of the film, which ranged from 0.61 to 2.50. Since a wide range of film densities is acceptable in the MAP, presumably because radiologists preferences span a range of background densities, films were selected which spanned the range of qualities expected from facilities participating in the ACR MAP [private communication, ACR Mammography Accreditation Physicist reader (1992)]. The techniques are listed in Table I. The machines are subject to routine quality control including the ACR MAP certification for both of the dedicated units.

A specimen radiography machine (SRM) (Faxitron series, 43807N X-Ray system, Hewlett-Packard, Prunetridge, CA 95014-9826) was used to produce a reference phantom image. This machine is designed to operate for long exposures without tube damage. It has a source-to-image distance of 56 cm. The same type of mammography film was utilized and was also processed with an undedicated processor (Kodak RP X-omat automatic processor, Kodak Company, Rochester, NY 14650). The remaining film (Table I, No. 5) was obtained using a nonclinical technique with a 10 min exposure on a typical specimen radiography unit with the phantom's wax insert placed directly on the film, without an intensifying screen or cassette. The purpose for using a non-mammographic machine and technique for this film was to maximize object visibility and produce a film which represents the upper bound of image quality. The eleven clinical technique films all have less subject contrast and more blur than film No. 5.

### C. Film digitization

A previous study investigated the various digitization technologies which could be used for this imaging problem.<sup>9</sup> The smallest objects in the images are the last microcalcification group ( $\sim 160/\mu\text{m}$  diameter). From Nyquist sampling frequency considerations, this indicates a device capable of delivering an  $80 \mu\text{m}$  spot size or smaller. Similarly, the range of optical densities represented spans nearly three optical density units, which indicates a device capable of delivering into the thousands of unique gray values. The digital images were generated with a cooled charge-coupled device (CCD) (Photometrics, Model 2300, Phoenix, AZ) providing  $2033 \times 2045$  each at 12 bits. The camera was installed in a Black Containment Box (BCB) with a previously specified light source and detector arrangement.<sup>11</sup> Those details are briefly summarized here. The field of view (FOV) of the camera was adjusted to encompass the phantom image boundary. To minimize stray light, the light source was collimated by an opaque tray with a 10 cm square opening directly over the center of the light source. The light source consisted of a bank of fluorescent cool-white light bulbs arranged along the edges of a two-foot square to yield an approximately flat, symmetrical light source in the center. The light source, opaque tray, and CCD camera were all adjusted in their relative positions to yield maximum brightness values in the center of the brightest object in the phantom image while minimizing the exposure time and maintaining the FOV. The CCD chip and the square opening were registered during digitization. This arrangement assured the consistent alignment of images during digitization without regard to the degree of rotation of the phantom image relative to the edges of the film. It also allowed the full spatial resolution of the CCD to sample the 10 cm square which yielded approximately  $50 \mu\text{m}$  pixel in each direction. A Macintosh IIfx computer, which interfaced with the CCD camera, was used to acquire and store the images on an optical platter for processing with a standard UNIX workstation and C compiler.

### D. Model-based vision computation

The algorithm for analyzing SMP digitized film images may be characterized as a constrained, two-dimensional, model-based recognition technique. The problem domain for the algorithm involves processing two-dimensional digital images of specific test objects in the SMP images, localizing the objects, and estimating their visibility according to experimentally measured observer data. The objects and approximate locations are defined *a priori*. The shapes include: rectangular-shaped fibers slanted at  $\pm 45^\circ$ , circular-shaped microcalcifications, and larger, disklike simulated tumor masses. In order to accomplish the localization requirements, the approach taken must not drastically alter the spatial location of the shapes in the images. Constrained rotation of the fibers ( $45^\circ$  rectangles) as well as translation of the fibers and other shapes must be allowable. Given these initial constraints, a template matching scheme was utilized for object localization.

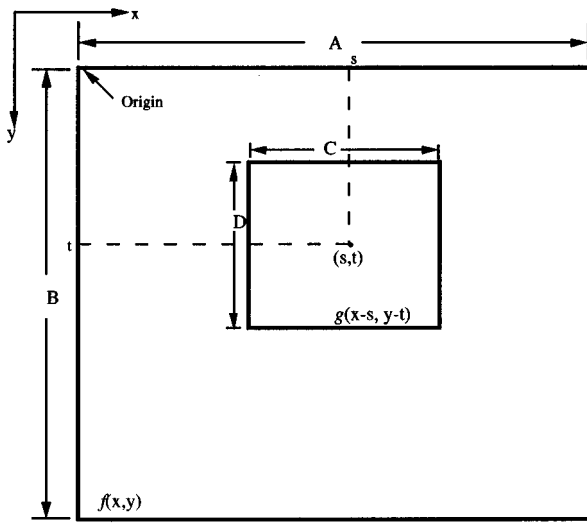


FIG. 2. Spatial domain schematic depicting the correlation of  $f(x,y)$  and template  $g(x,y)$  at point  $(s,t)$  (adapted from Ref. 13).

### 1. Frequency domain template matching

Figure 2 depicts the template matching process in the spatial domain. The digitized images are 2033 pixels by 2045 pixels (10 cm square FOV) by 12 bits/pixel due to the resolution requirements for sampling the 160  $\mu\text{m}$  microcalcification group. This dictates a prohibitively large object image and template image for a spatial domain approach. Therefore, the template matching approach utilized a fast Fourier transform (FFT), decimation in time, or Cooley–Tukey algorithm.<sup>12</sup> The only assumptions are that the array sizes are square and are evenly-divisible by 2. These conditions have been met by the proposed digitization scheme. Equation (1) provides the mathematical definition of the continuous, two-dimensional Fourier transform, where

$$F(u,v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \exp[-j2\pi(ux+vy)] dx dy. \quad (1)$$

The variables  $u$  and  $v$  are the associated frequency components for the  $x$  and  $y$  variables. The functions  $f(x,y)$  and  $F(u,v)$  are the Fourier transform pairs. The parameter  $j = \sqrt{-1}$  is the standard imaginary number. The convolution theorem provides the means for using frequency domain correlation as an alternative to a spatial domain approach. The theorem states that the spatial domain convolution, given by  $f(x,y)*g(x,y)$ , is equivalent to the corresponding frequency domain relation,  $F(u,v) \cdot G(u,v)$ , as shown in Eq. (2):

$$\begin{aligned} f(x,y)*g(x,y) &\Leftrightarrow F(u,v) \cdot G(u,v), \\ F(u,v) \cdot G(u,v) &\Leftrightarrow f(x,y)*g(x,y), \\ f(r, \theta + \theta_0) &\Leftrightarrow F(\omega, \varphi + \varphi_0). \end{aligned} \quad (2)$$

This is practically the process of centering the template image over the first pixel of the object image, multiplying the template's values by each value underneath it, replacing the original image pixels by the product, and moving the tem-

plate across the image in the spatial domain. In the frequency domain, this can be accomplished by simply multiplying the two Fourier-transformed functions and inverse transforming the product. The continuous version of the convolution process is given by Eq. (3):

$$f(x,y)*g(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\alpha,\beta) g(x-\alpha, y-\beta) d\alpha d\beta. \quad (3)$$

Since an image is formed of quantized gray values, Eq. (3) must be cast into discrete form. The two-dimensional, discrete Fourier transform is given by Eq. (5), which follows from application of Eq. (4) to Eq. (1):

$$\begin{aligned} f_e(x,y) &= \begin{cases} f(x,y), & 0 \leq x \leq A-1 \quad \text{and} \quad 0 \leq y \leq B-1 \\ 0, & A \leq x \leq M-1 \quad \text{and} \quad B \leq y \leq N-1, \end{cases} \\ g_e(x,y) &= \begin{cases} g(x,y), & 0 \leq x \leq C-1 \quad \text{and} \quad 0 \leq y \leq D-1. \\ 0, & C \leq x \leq M-1 \quad \text{and} \quad D \leq y \leq N-1, \end{cases} \end{aligned} \quad (4)$$

$$F(u,v) = \frac{1}{N} \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} f(x,y) \exp\left[\frac{-j2\pi(ux+vy)}{N}\right]. \quad (5)$$

Equation (4) allows  $f(x,y)$  and  $g(x,y)$  to become discrete arrays with finite bounds of size  $A$  by  $B$  and  $C$  by  $D$ , respectively. Equation (6) follows from application of Eq. (4) to Eq. (3). The values  $M$  and  $N$  are the assumed periodicity in the  $x$  and  $y$  directions, respectively:

$$\begin{aligned} f_e(x,y)*g_e(x,y) &= \frac{1}{N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [f_e(m,n) \\ &\quad \cdot g_e(x-m, y-n)], \\ M &\geq A + C - 1, \quad N \geq B + D - 1. \end{aligned} \quad (6)$$

The issue of wrap-around error is of concern with the convolution process. This issue is treated in detail in many references.<sup>12,13</sup> To completely avoid this error, the images should be adjusted until the conditions for  $M$  and  $N$  in Eq. (6) are met. It is sufficient to summarize this effect by stating that both the image  $[f(x,y)]$  and template  $[g(x,y)]$  need to be zero-padded out to the maximum positive or negative duration of the objects of interest. For instance, if an object of interest occupied the middle-half of an image, and the template was sized similarly, the template must be expanded by adding zero values to perimeter locations until its array size is larger by one-half the dimension of the object. The image must also be increased in this fashion or, if it is already large enough to meet this requirement, the same corresponding locations must be either zero padded or ignored as they will be corrupted from wrap-around error. The process of zero-padding effectively selects a window of interest through which to view the image. This is unavoidable in practice since most images are of objects which are themselves finite in extent (i.e., nonzero mean). Unless the image is band-limited and periodic, all spatial frequencies cannot be completely recovered after forward and reverse Fourier transformations have taken place. The effect adds a small degree of blur to the inverse-transformed resultant image. In

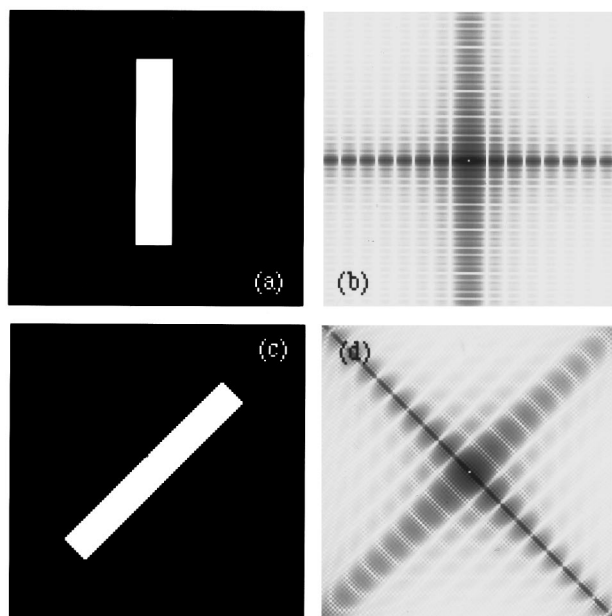


FIG. 3. (a) Simple object in spatial domain, (b) power spectrum of simple object, (c) simple object rotated 45°, and (d) power spectrum of rotated simple object.

the current context, this loss of very high-frequency information is negligible since the Fourier results are used for relative correlation coefficient estimation only and this error does not affect the original image values or the contrast calculation.

Since the FFT is a symmetric linear operator, arbitrary translation and rotation of the test images are maintained throughout Fourier space processing. This is demonstrated by Fig. 3 where the power spectrum rotates along with the object, assuring that the FFT will not alter the original translation and rotation of the image information. The power spectrum display [Eq. (7)] allows visualization of the Fourier frequency domain information. The terms  $R(u,v)$  and  $I(u,v)$  are

$$P(u,v) = R^2(u,v) + I^2(u,v), \quad (7)$$

The real and imaginary components of the transformed function,  $F(u,v)$ . This is also evident in the polar coordinate representation of the convolution theorem statement provided by the last line of Eq. (2).

## 2. Object localization algorithm

The implementation of the Fourier convolution technique as well as other model-based techniques are combined into a single algorithm referred to as the Mammography Quality Control Program (MQCP). Figures 4 and 5 depict the algorithmic data and control flow. The main assumption made in MQCP is that the digitized input images are cropped or optimally digitized at the apparent edge of the wax insert. This assumption is readily met in practice by the CCD device and the BCB arrangement. If the film is translated during digitization, the result will become evident in the displacement report but this will not add error to the localization process

unless the film is shifted more than about 1 cm or more. The rotation- and translation-independent nature of the Fourier transform localization methods provides this immunity. The 1 cm limit is due to the constraints of the phantom manufacturer whereby any displacement error of more than a few millimeters is sufficient for rejection [Private Communication, RMI-156 Breast Phantom Product Manager (1993)]. Thus, these constraints were utilized in MQCP to limit the search areas.

The Fourier-domain template matching approach used at the start of the MQCP provides only a partial measure of object localization. Following Fig. 6, the control flow: (1) reads the image and its dimensions and scale; (2) extracts the first subimage and performs its FFT; (3) either reads a pre-computed mask-FFT (shaded lines in Fig. 6) for the particular shape or generates a zero-centered, binary mask for the shape and performs the FFT; (4) the object FFT and the mask FFT are then multiplied, element-by-element, to perform the convolution; (5) the inverse-FFT is performed on the product image; (6) the resultant image is a matrix of correlation coefficient values indicating the degree of correlation the template exhibited for the value's location in the original image. This process is shown with corresponding images in Figs. 6, 7, and 8 for the three shapes. For the second portion of the MQCP control flow, shown in Fig. 5, the correlation surface resulting from part 1 is searched for maxima in the search ranges determined *a priori* and these locations are stored. Typical maxima are indicated in Figs. 6, 7, and 8. Since one objective for fibers is to measure the amount of rotation, the ideal angle for a particular fiber was used to create the mask and minor angular differences between the mask and the actual grayscale image do not appreciably affect measurement of the underlying fiber angular placement error. The same argument holds for displacement error of the other shapes. This effect should be apparent in Fig. 7, where only a portion of the speck group is used as an input image and the approach localizes the specks in a correctly registered, correlation surface image. This is true as long as displacement is no more than the manufacturer-specified 1 cm in any direction for all three shapes (including individual specks). The MQCP was not designed or tested for objects displaced beyond this limit.

The fiber correlation surface, shown in Fig. 6, is essentially a ridge of high values which are somewhat noisy, depending upon the noise and artifact levels in the original image. Multi-element, unidirectional derivative filters were tested to determine which would demonstrate the peak values of fiber correlation or maximum ridge. The results from the filter testing were used to select the optimal filter size for MQCP's localization of fiber peak values. The peak values of fiber correlation can deviate from a straight line depending on the image noise and the amount of dislocation of the fiber relative to the subimage lateral boundaries. The latter effect is due to the frequency-domain errors associated with the FFT convolution approach as discussed previously. To avoid these effects, the peak values are only considered around a 1 cm square vicinity of the ideal, *a priori* fiber location. The image noise may still cause a discontinuity in the ridge val-

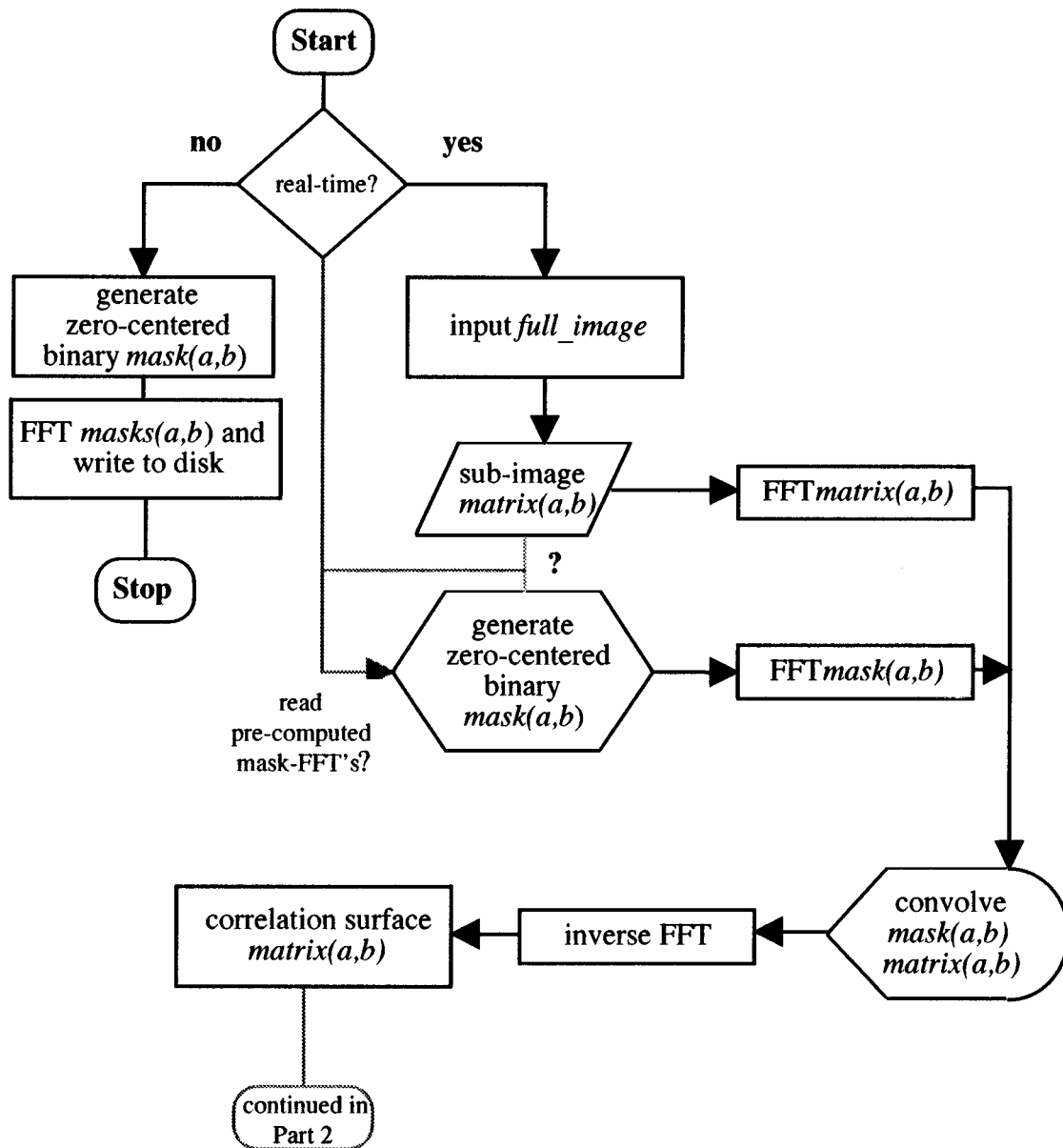


FIG. 4. Images corresponding to Mammography Quality Control Program (MQCP) control flow for fibers. The original fiber image and an appropriate binary mask are Fourier transformed (the power spectra of each are shown below each original image) and multiplied in the frequency domain. The product is then inverse Fourier transformed to yield the convolution image. The convolution image is depicted by scaling the correlation coefficients to 8 bits. The area of maximum correlation is evident by the darkest area in the vicinity of the original fiber.

ues which fall inside this spatial constraint. This effect is handled by iteratively selecting portions of the ridge until a section, as long as the particular fiber is wide, gives angles from a least-squares fit to within  $10^\circ$  of the *a priori*  $45^\circ$ . If the constraint is not met after ten iterations, MQCP will stop searching and default to the *a priori* location as the centroid coordinates. Equation (8) mathematically describes the least-squares fit used by MQCP. The angle was taken as the inverse tangent of the slope from the least-squares fit where  $x_i$ =peak location  $i$ ,  $x$  coordinate;  $y_i$ =peak location  $i$ ,  $y$  coordinate;  $n$ =total number of peak values:<sup>14</sup>

$$\text{angle} = \tan^{-1} \left[ \frac{\left( \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right) \right)}{\sum_{i=1}^n x_i - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n}} \right]. \quad (8)$$

The peak values should include the numerically largest value of correlation in the image. However, a minor discrepancy could arise from application of the derivative filter used to find the peak values. If this occurs, the final centroid coordinates for a fiber are taken to be the  $x$  coordinate from the

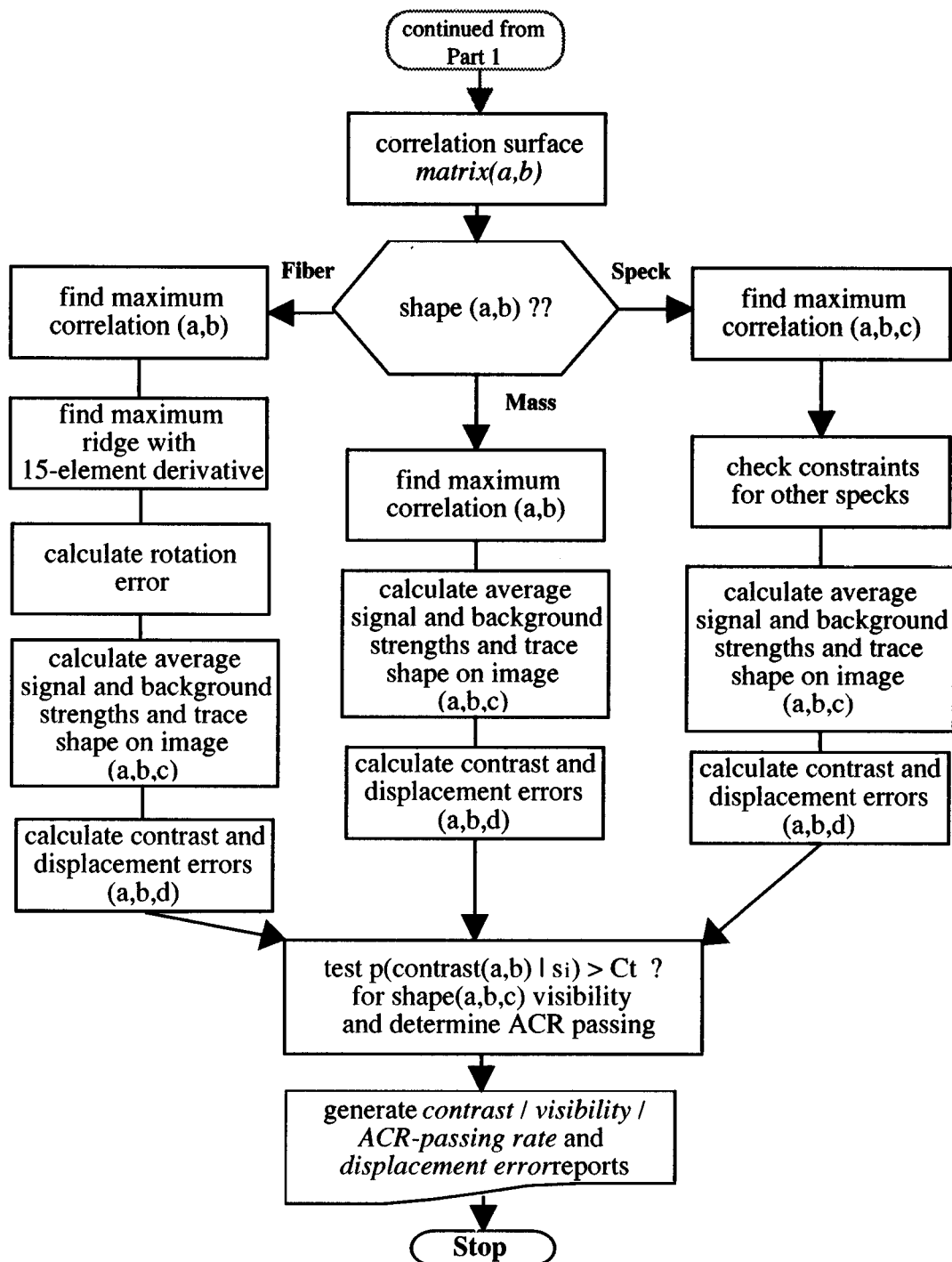


FIG. 5. Images corresponding to Mammography Quality Control Program (MQCP) control flow for speck groups. The original speck group image and an appropriate binary mask are Fourier transformed (the power spectra of each are shown below each original image) and multiplied in the frequency domain. The product is then inverse Fourier transformed to yield the convolution image. The convolution is depicted by scaling the correlation coefficients to 8 bits. The area of maximum correlation is evident by the darkest area in the vicinity of the original fiber.

maximum correlation location and the  $y$  coordinate is taken as the peak value corresponding to the  $x$  coordinate. Since the specks and masses are rotationally symmetric, there is no angular error component. The ideal locations for all 16 objects are schematically depicted in Fig. 9. These locations are measured from the edges of the image, thus it is assumed

that the input images are cropped at the wax edges as previously discussed. The centroid coordinates were selected which were within 1 cm of the ideal locations shown in Fig. 9 and the maximum of the correlation surface in the same search area. The speck groups are arranged in the corners and center of a regular pentagon. The ideal location of the

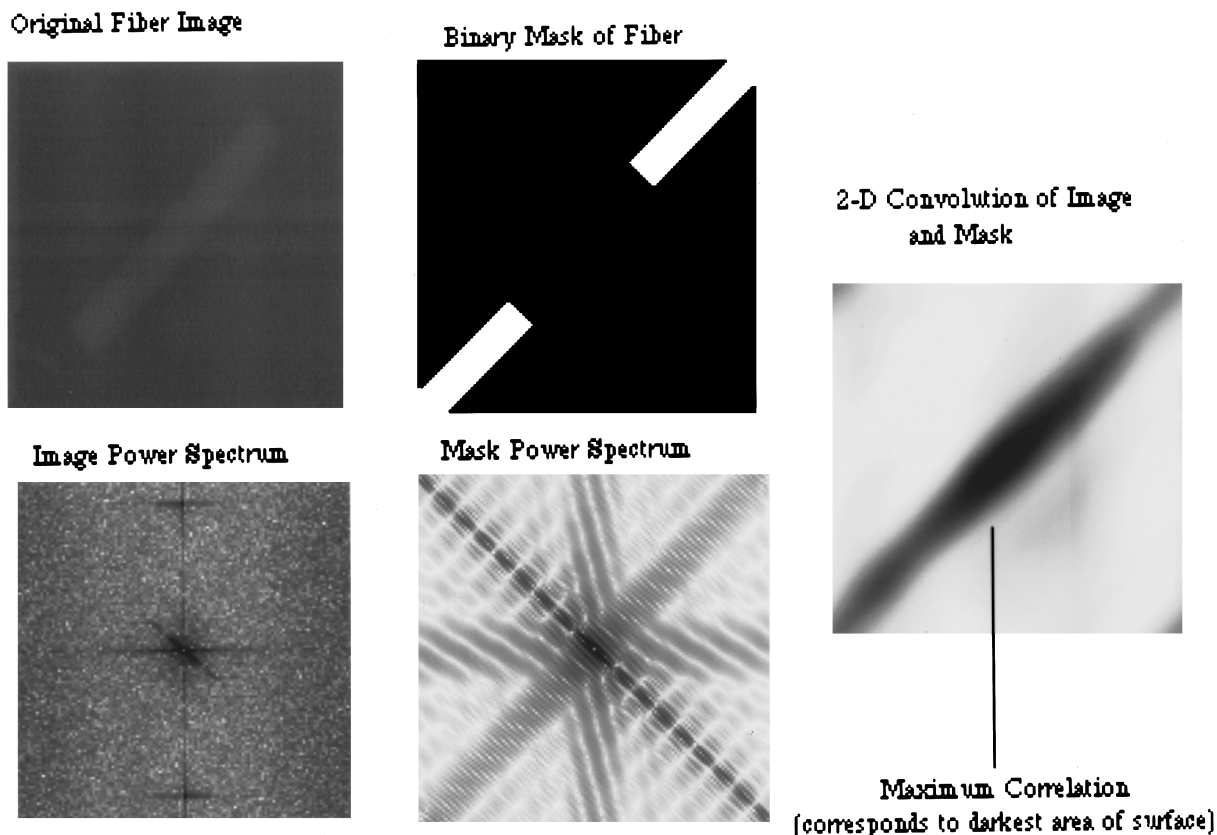


FIG. 6. Part 1—Mammography Quality Control Program (MQCP) algorithmic data and control flow.

center of the speck group is assumed to coincide with the location shown in Fig. 9 where specks occur. The ideal locations of the remaining specks in a given group are measured relative to the coordinates of a pentagon centered about the coordinates shown in Fig. 9. The displacement errors were measured by the standard distance formula applied to the MQCP-located centroid coordinates and the ideal coordinates for a particular shape.

### 3. Object visibility

Once the location of a particular shape is estimated, the contrast of the shape is determined. The contrast for MQCP is calculated by Eq. (9), where  $S$  = signal, average gray value of located object;  $B$  = background, average value of area surrounding located object; dark = dark current of camera system (if applicable):<sup>11</sup>

$$C = \frac{S - \text{dark}}{S + B - \text{dark}} \quad (9)$$

The background area dimensions were selected which encompassed enough background area to provide a stable average gray-value for the region. The MQCP utilizes a comparable area of surround, or greater, on each dimension of the object as suggested by Chesters.<sup>15</sup> The MQCP indicates the areas used as signal and background by marking the closest outside pixel black at the border of each region of interest. This computer graphics feature allows the MQCP local-

ization results to be quickly analyzed for accuracy. Samples of this process are provided in the results section.

Final classification of the localized objects is performed by a two-hypothesis (binary) Bayesian classifier. The classification variable is threshold contrast,  $C_t$ , and the decision for visibility is assumed to be at 50% visibility. Implementation of this classifier is accomplished by comparison of a particular shape's contrast with the threshold value for that shape and establishing visibility if it is greater than the threshold and nonvisibility if it is less than the threshold. The observer data were taken from previously presented measurements.<sup>4</sup> Section III contains the determination of  $C_t$  for the three shapes used in MQCP.

Testing of MQCP against humans for ACR passing rates was performed for six images which had a moderate observer response rate. That is, about half as many passed the images as failed them. Human observers in the present work did not include scoring of artifacts or partial object visibility. The moderate response rate for these films assured that a human threshold was encountered in each shape category, because a portion of the human observer population had to fail the images for the response rate to be moderate. The images used were digitized from film numbers 1, 2, 3, 5, 6, and 10 of those listed in Table I. The same images were also used to train the binary classifier by establishing the threshold contrasts for each shape. While using the same images to both establish a threshold and test the threshold could lead to bias in the classifier, the objective of this work is to demon-



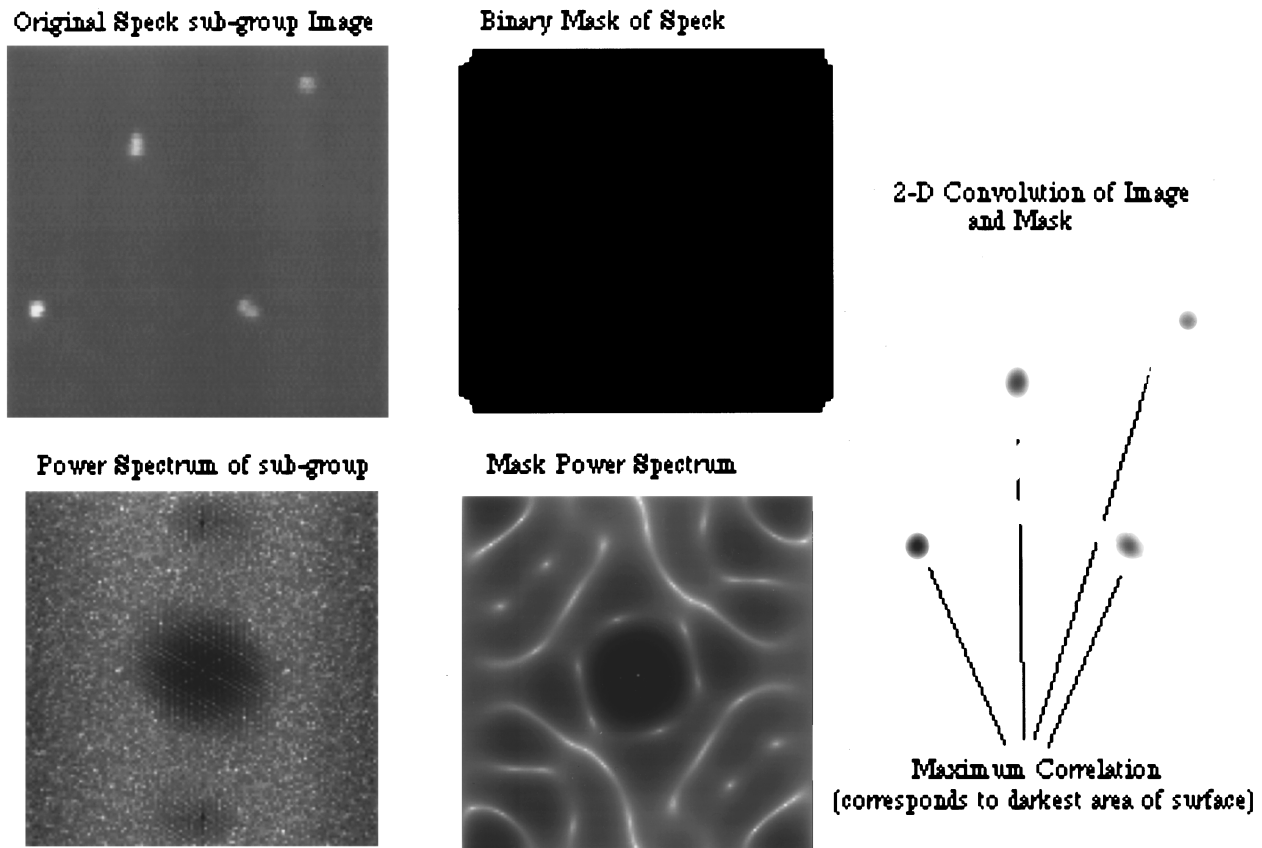


Fig. 7. Images corresponding to Mammography Quality Control Program (MQCP) control flow for masses. The original mass image and an appropriate binary mask are Fourier transformed (the power spectra of each are shown below each original image) and multiplied in the frequency domain. The product is then inverse Fourier transformed to yield the convolution image. The convolution is depicted by scaling the correlation coefficients to 8 bits. The area of maximum correlation is evident by the darkest area in the vicinity of the original fiber.

strate that automated processing is feasible. Establishment of a universally applicable threshold or other decision metric derived from actual ACR cases is necessary for future advancement of such an approach, but it is not addressed in this work.

For displacement error testing, three SMP wax inserts were used from a selection of rejected SMPs from the phantom vendor. These inserts were specifically rejected because of their displacement errors. Each image represents a rejection based upon one shape. The known errors consist of: film 1, the fifth fiber is significantly rotated past  $45^\circ$ ; film 2, the third speck group has a severely displaced speck; and film 3, the last mass is severely displaced. The errors were quantified by physically estimating displacement or rotational deviation from estimated ideal conditions. The SMP film images were digitized and the pixels manually counted. The ideal conditions were estimates used to demonstrate that MQCP could quantify errors. Comparison of MQCP and the measured errors are provided in the results in Table II.

After MQCP has estimated location, errors, contrast, and visibility for all shapes, final output reports are generated summarizing this information. Sample output results from MQCP for a test image are provided in the results in Tables III and IV. They include displacement errors and rotation errors for fibers, displacement errors for speck groups and

masses, and contrast information for the above as well as a final ACR passing decision based on the object visibility scores.

### III. RESULTS

#### A. Automated localization performance

Table II contains the comparison between MQCP and measured distances for displacement errors as well as angle error for the three films which failed the phantom vendor's quality control requirements. The distances given are centroid displacement distances, measured relative to Fig. 9. Both the angle errors and the distance errors indicate that MQCP is capable of tracking displacements and angular error relative to any arbitrary reference frame. This is also evident from the results shown in Fig. 10 for film No. 5. The observer passing response for this film was unanimous at 100%.<sup>4</sup> These results are indicative of the localization performance expected from MQCP when a very good quality image is provided and high-resolution digitization used. For film No. 2, MQCP's localization performance is shown in Fig. 11. These results are indicative of a film which is only marginally meeting the ACR passing rate criterion. This film averaged less than 50% passing rate from all of three groups

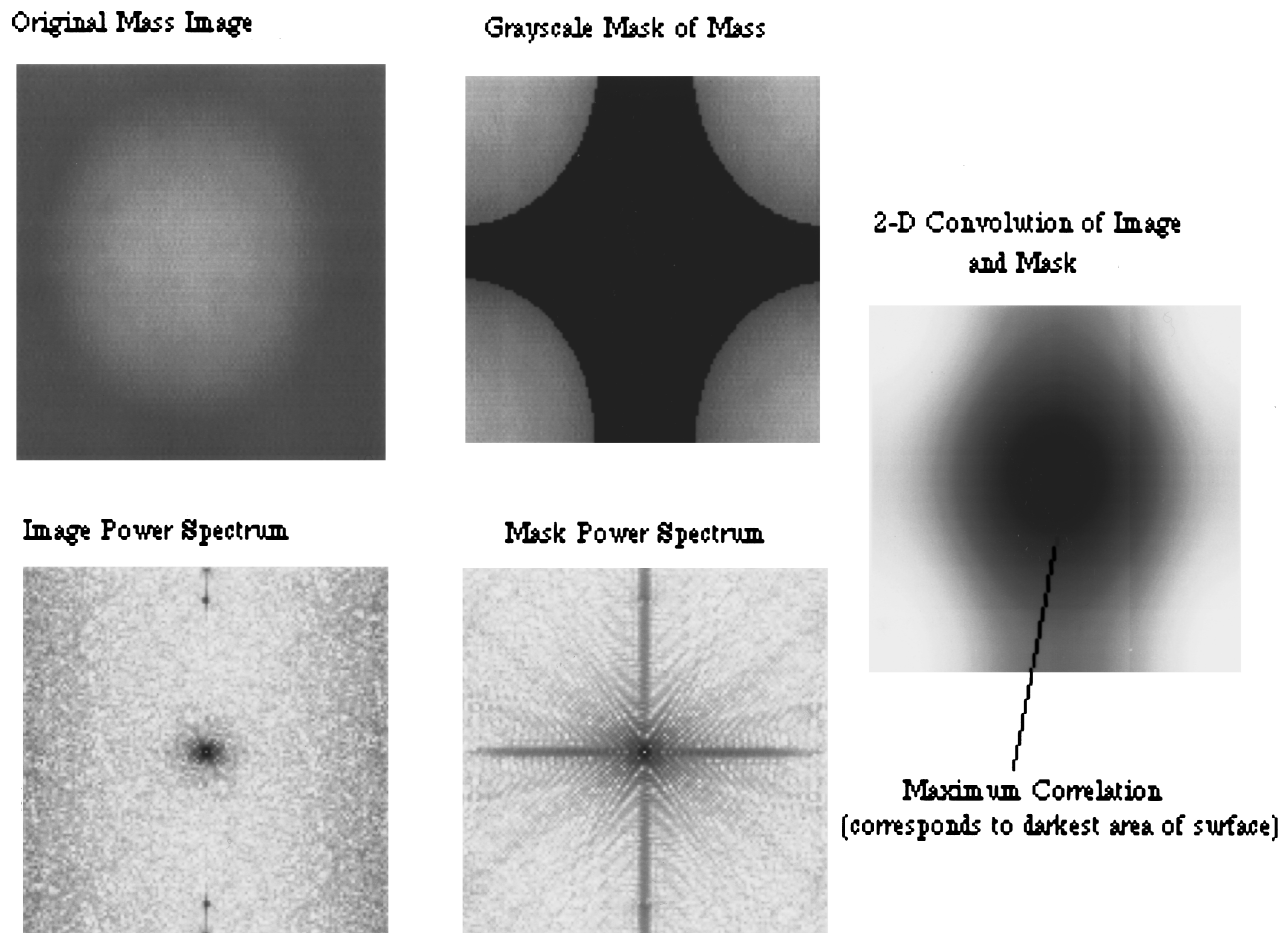


FIG. 8. Part 2—Mammography Quality Control Program (MQCP) algorithmic data and control flow.

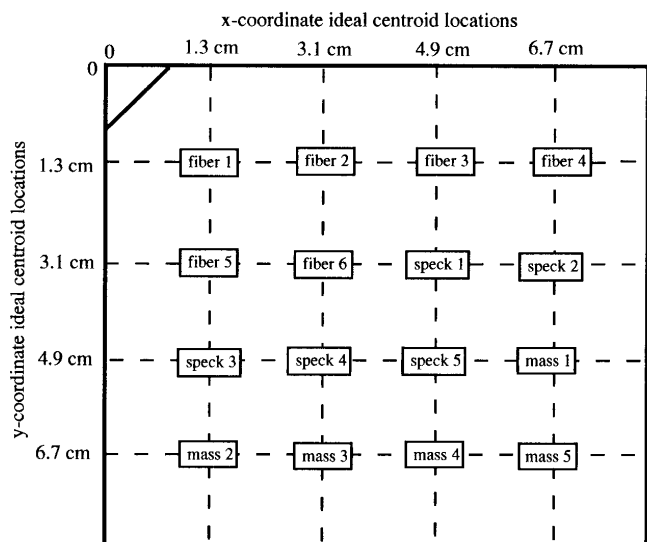


FIG. 9. Ideal centroid coordinates for Mammography Quality Control Program (MQCP) displacement error calculations. The lateral edges of this ideal reference frame corresponds to the edges of a square mask used during the digitization of each phantom image. This provided a convenient reference frame and allowed the spatial resolving capabilities of the CCD camera to operate only on the image contents without wasted space around the dark edges.

of 30 observers each.<sup>4</sup> The image is noisy and all of the objects are not necessarily visible to the human eye.

**B. Automated vision performance**

The results from training MQCP's binary decision classifier are shown in Fig. 10. The low spatial frequency objects (fibers and masses) exhibit very similar contrast visibility and are distinct from the response for high spatial frequency objects (specks) for the observers tested. These results are in agreement with similar results presented previously.<sup>15</sup> The threshold contrast values corresponding to the decision prob-

TABLE II. Mammography Quality Control Program (MQCP) localization performance compared to phantom vendor quality control. The angle and distance measurements were made manually on the film image. The computed (MQCP) estimates for angle and distance were made relative to the ideal reference frame shown in Fig. 9.

Film no.	Error	Angle (°)	MQCP angle (°)	Distance (cm)	MQCP distance (cm)
1	fiber angle	72	69.7	n/a	n/a
2	speck	n/a	n/a	1.64	1.78
3	mass	n/a	n/a	1.93	2.04

TABLE III. Mammography Quality Control Program (MQCP) summary output report. Contrast, visibility, and displacement error for film 5. The first half of this table gives the computer measured contrast and indicates the pass/fail decision. The last half provides an example summary displacement report. The displacement is relative to the ideal reference frame outlined in Fig. 9. This shows how a computerized approach could be used for documenting manufactured locations for quality assessment.

Fibers		Contrast and visibility report					Visible
		Contrast(min. 1.010)					
	1	1.093					Yes
	2	1.079					Yes
	3	1.069					Yes
	4	1.060					Yes
	5	1.034					Yes
	6	1.031					Yes
Number of fibers visible: 6		(ACR requirement: 4)					
Speck Groups	Contrast (min. 1.156)					Visible	
	1	2	3	4	5	6	
1	1.276	1.329	1.342	1.327	1.318	1.309	Yes
2	1.570	1.578	1.544	1.534	1.323	1.602	Yes
3	1.502	1.546	1.578	1.571	1.610	1.611	Yes
4	1.338	1.332	1.441	1.360	1.453	1.375	Yes
5	1.130	1.132	1.101	1.107	1.067	1.089	No
Number of speck groups visible: 4		(ACR Requirement: 3)					
Masses		Contrast (min. 1.016)					Visible
	1	1.229					Yes
	2	1.121					Yes
	3	1.098					Yes
	4	1.057					Yes
	5	1.023					Yes
Number of masses visible: 5		(ACR requirement: 3)					
Displacement report (mm from ideal)							
Fiber No.	Right(+) Left(-)	Up(+) Down(-)	Centroid distance	Rotation angle (degrees)			
1	-1.498	-0.005	1.498	45.324			
2	6.840	0.277	6.846	-46.685			
3	11.094	-0.756	11.120	45.987			
4	20.277	-1.366	20.323	-43.953			
5	-1.545	-10.643	10.755	44.678			
6	4.305	-5.432	6.93	-45.000			
Speck group No.	Centroid distance						
	1	2	3	4	5	6	
1	13.507	13.493	14.962	15.441	13.515	11.622	
2	19.902	19.325	21.255	21.453	19.265	17.888	
3	12.599	10.662	12.337	15.005	14.330	12.297	
4	13.405	12.598	14.697	15.892	14.956	12.979	
5	18.485	16.799	18.746	15.410	15.756	15.765	
Mass No.	Right(+) Left(-)	Up(+) Down(-)	Centroid distance				
1	18.587	-12.174	22.219				
2	-0.136	-19.103	19.104				
3	0.418	-19.291	20.331				
4	12.784	-19.291	23.143				
5	18.164	-17.930	25.523				

ability of 50% visibility were: fibers, 1.010; specks, 1.156; and masses, 1.016. These values were utilized in MQCP to test for visibility of located objects.

The results from testing MQCP with six test images and comparing them to human judgments are shown in Table V. The contrast, visibility, and displacement error results are

summarized in Tables III and IV for the two images shown in Figs. 11 and 12, respectively. Localization graphics for these images indicate the areas used for signal and background in the contrast calculation. The MQCP gave the same object scores as humans for film 5 (Fig. 10.) This was evidence of the ability of the system to correctly operate on

TABLE IV. Mammography Quality Control Program (MQCP) summary output report. Contrast, visibility, and displacement error for film 5. The first half of this table gives the computer measured contrast and indicates the pass/fail decision. The last half provides an example summary displacement report. The displacement is relative to the ideal reference frame outlined in Fig. 9. This shows how a computerized approach could be used for documenting manufactured locations for quality assessment.

Fibers		Contrast and visibility report Contrast (Min. 1.010)					Visible
1					1.019	Yes	
2					1.014	Yes	
3					1.009	No	
4					1.002	No	
5					1.003	No	
6					1.005	No	
Number fibers visible: 2		(ACR requirement: 4)					
Speck Groups	Contrast (Min. 1.156)		3	4	5	6	Visible
1	1.123	1.201	1.111	1.076	1.167	1.105	No
2	1.195	1.193	1.122	1.162	1.083	1.165	Yes
3	1.166	1.234	1.236	1.229	1.245	1.178	Yes
4	1.101	1.123	1.131	1.169	1.138	1.145	No
5	1.115	1.085	1.142	1.099	1.115	1.074	No
Number speck groups visible: 2		(ACR requirement: 3)					
Masses		Contrast (Min. 1.016)					Visible
1					1.061	Yes	
2					1.023	Yes	
3					1.021	Yes	
4					1.003	No	
5					1.007	No	
Number of masses visible: 3		(ACR requirement: 3)					
Displacement report (mm from ideal)							
Fiber No.	Right(+)		Up(+)		Centroid distance	Rotation angle (degrees)	
	Left(-)		Down(-)				
1	2.493		6.991		7.422	45.567	
2	6.418		0.230		6.422	-45.749	
3	8.700		0.324		8.706	45.134	
4	11.451		6.991		13.416	-46.987	
5	-3.469		0.812		3.563	46.005	
6	8.296		0.390		8.305	-46.003	
Centroid distance							
Speck group no.	1	2	3	4	5	6	
1	10.957	11.0691	12.404	12.802	10.924	9.131	
2	17.200	16.713	18.468	18.612	16.860	15.646	
3	10.876	9.066	10.413	13.049	12.710	10.865	
4	12.480	10.183	12.161	13.514	12.717	10.675	
5	14.778	13.563	20.137	15.850	15.165	13.306	
Mass No.	Right(+)		Up(+)		Centroid distance		
	Left(-)		Down(-)				
1	17.742		-9.732		20.236		
2	-2.155		-16.991		17.127		
3	4.681		-17.460		18.077		
4	9.920		-11.310		15.044		
5	21.028		-12.577		24.503		

extremely good images. Film 2 gave conflicting results between MQCP and the trained diagnostic physicists. The reason film 2 failed was due to the fact that only half of the fourth fiber is actually visible. The observers were instructed to make a judgment between visible or nonvisible, though

40% of the trained observers commented that the fiber was half-visible while viewing this film. Thus the contrast is lowered by an appreciable amount, causing it to drop below the threshold. In this instance, the binary classifier still comes very close to reaching the same decision as humans (off by

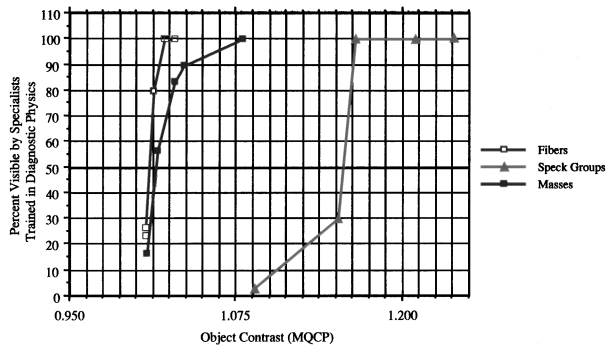


FIG. 10. Mammography Quality Control Program (MQCP)-measured contrast versus percent visibility for fibers, specks, and masses. The observers were specialists trained in diagnostic physics.

10%). The specks and masses for film 2 registered passing scores with MQCP as they did with humans. A more complex classifier, such as a multi-hypothesis decision rule, may provide closer results in these cases. The MQCP passed or failed the remaining films the same as the human observers.

#### IV. DISCUSSION

Since current SMP designs do not provide an absolute reference frame such as fiducial markers, the methods used had to locate the test objects autonomously. To help with this, a digitization mask was created which cropped the film image at the apparent edge of the wax insert. At this point, the images were positioned so that the shapes were in a known order. Having properly digitized the images, the image processing was performed in two stages: object localization and object visibility. First, object localization is crucial since no attempt to model visual responses with computers can work without first finding gray values which have a high probability of being related to the correct objects. Thus, low-level processing utilizing Fourier domain template matching was employed to provide a registered map of correlation coefficients. Intermediate-level processing utilized derivative filters operating on the correlation coefficient map to find local maxima. This terminology is consistent with that of Maar.<sup>16</sup> These results look promising for salient localization of the test objects in SMP images. The algorithms performed

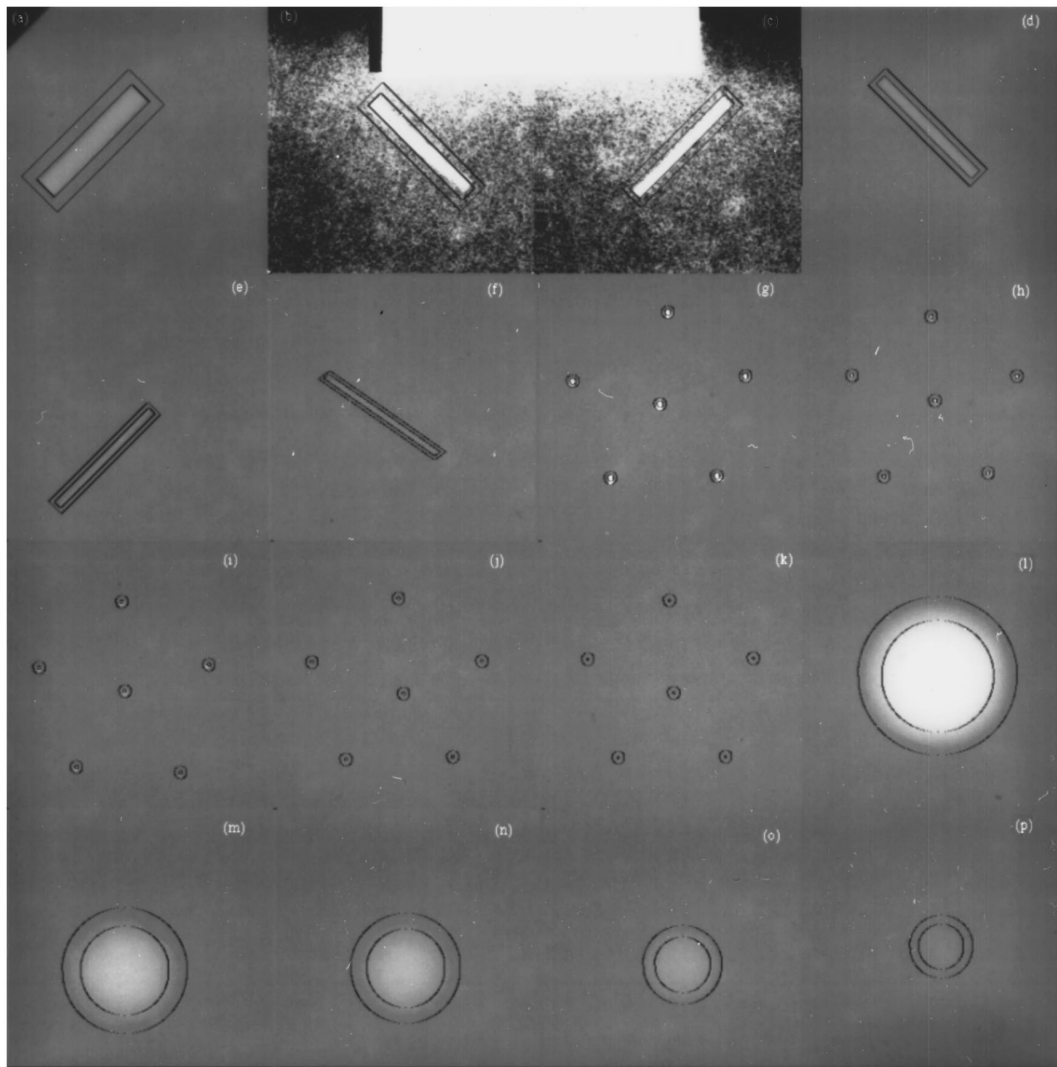


FIG. 11. Mammography Quality Control Program (MQCP) localization, film 5.

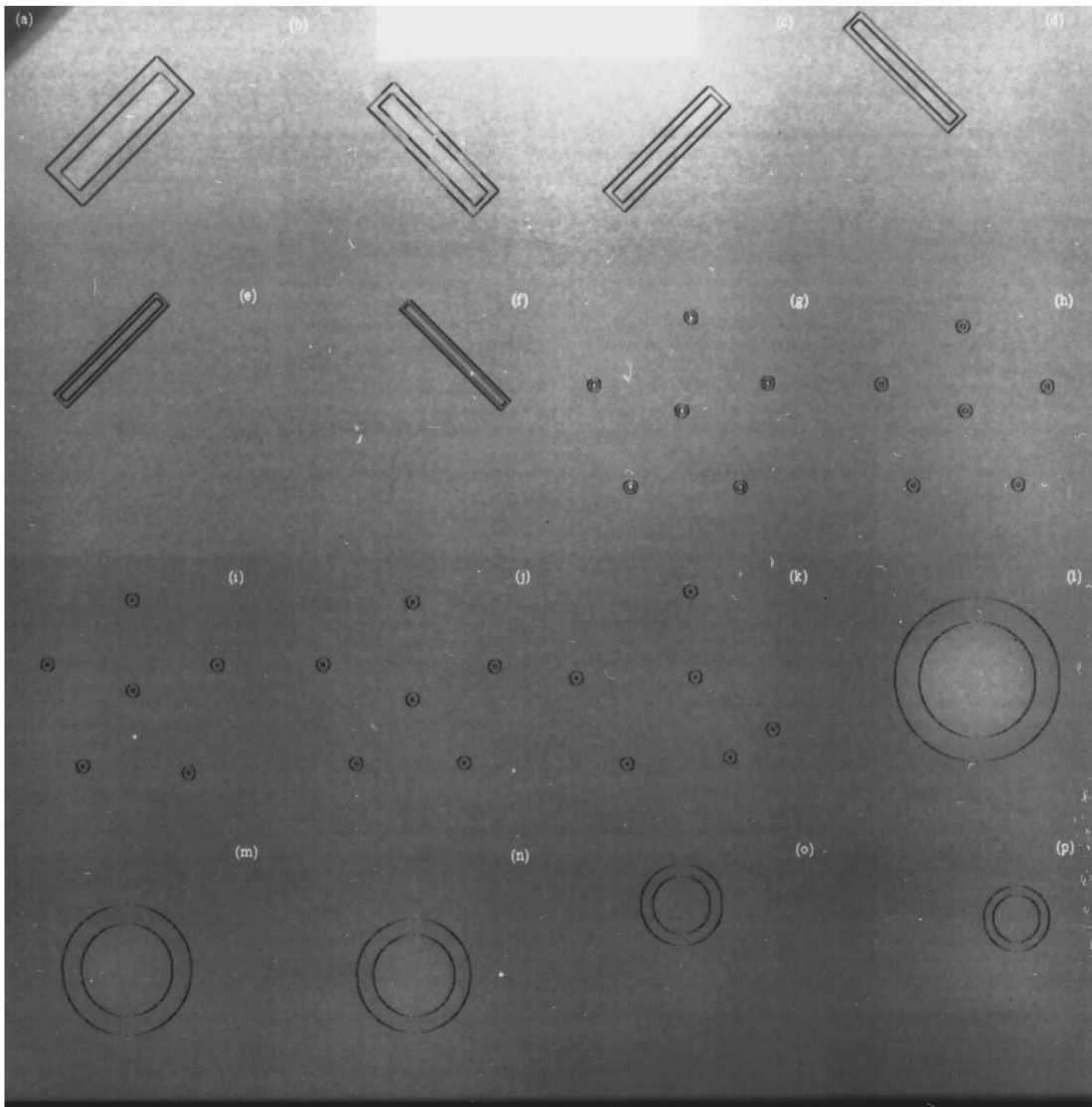


FIG. 12. Mammography Quality Control Program (MQCP) localization, film 2.

at least as well as humans and without human variability. This was to be expected since the computer has access to much more information than the human eye can process and is in agreement with other published results.<sup>5</sup> The final stage of processing was the high-level classification which was modeled by a Bayesian classifier using threshold contrast as

TABLE V. Mammography Quality Control Program (MQCP) passing rate results compared to human observers.

Film No.	MQCP judgment	Trained-physicist ACR passing rate
1	Fail (<50% probability)	40% passed
2	Fail (<50% probability)	60% passed
3	Pass (>50% probability)	100% passed
5	Pass (>50% probability)	100% passed
6	Fail (<50% probability)	20% passed
10	Fail (<50% probability)	40% passed

measured from the target observer group. Threshold contrast has previously been identified as a useful predictor variable for estimating human visibility.<sup>5,17</sup> The performance by MQCP, coupled with a cooled CCD 2033 by 2045 by 12 bit camera digitizer, is in good agreement overall with specially trained human observers. However, the same image test set was used to establish the threshold and test the classifier. While using the same images to do this could lead to bias in the classifier, the objective of this work was to demonstrate that automated processing is feasible, not to establish a generic classifier threshold which would apply universally. Now that this has been shown, any model of human perception that is effective may be used for the classification. We are currently considering other models of human visual classification for both the test objects and artifacts.

In any system useful for a national program, there is the additional requirement of artifact analysis. The current system of deductions and half-visible object scores was not in

place at the time our original observer measurements were taken and we have no set of human responses for the almost unlimited random shapes presented by artifacts. We feel that the role for MQCP is to preprocess the images to find the test objects. After the objects are located, then subtract them from the image and process the remaining image area focusing on artifacts. We are currently working to establish a database of known artifacts from which to devise a method for at least developing a computer-assisted tool for documenting artifacts and deducting points. In addition, we are extending this work by investigating phantom design enhancements to provide an all-encompassing quality assessment including spectral information, entrance exposure information, direct measurement of noise power, and MTF to provide more quantitative, objective measures of image quality.

In summary, machine interpretation is plausible and could be used in conjunction with the current MAP phantom image evaluation process to improve consistency and objectivity. The techniques used for object recognition and scoring could be extended to image artifacts augmented by decision inference rules. Digitization devices which meet the requisite technical performance criterion are commercially available and must be used consistently. Such an approach could be used to assist with elimination of manufacturing errors and provide an autonomous screening tool for future mammographic facility accreditation. The ultimate significance will be apparent when the implementation of such an approach provides more consistent, high-quality mammograms.

<sup>a)</sup>Electronic-mail: ken@radonc.emory.org

<sup>1</sup>M. C. Martin and J. B. Smathers, "Requirements of the American College of Radiology Mammography Accreditation Program," in *Current Regulatory Issues in Medical Physics*, proceedings of an American College of Medical Physics Symposium (Medical Physics Publishing, Madison, WI, 1992).

<sup>2</sup>P. A. Wilcox, "American College of Radiology Mammography Accreditation Program," in *24th Annual Conference on Radiation Control*, Proceedings of a CRCPD Conference, Orlando, FL, CRCPD-Pub. 92-5 (1992), pp. 145-152.

<sup>3</sup>NCRP Report #85, *Mammography* (National Council on Radiation Protection and Measurements, 1989).

<sup>4</sup>K. W. Brooks, J. H. Trueblood, and K. J. Kearfott, "Subjective evaluations of mammographic accreditation phantom images by three observer groups," *Inv. Radiol.* **29**, 42-47 (1994).

<sup>5</sup>D. P. Chakraborty and M. P. Eckert, "Quantitative versus subjective evaluation of mammography accreditation phantom images," *Med. Phys.* **22**, 133-143 (1995).

<sup>6</sup>K. W. Brooks, J. H. Trueblood, K. J. Kearfott, and D. T. Lawton, "Automated analysis of mammographic quality control images," *Med. Phys.* **20**, 881 (1993).

<sup>7</sup>R. K. Cacak, "Inconsistencies in mammographic phantoms used for ACR accreditation," *Radiology* **181**, 288 (1991).

<sup>8</sup>A. Rose, *Vision: Human and Electronic* (Plenum, New York, 1973).

<sup>9</sup>J. H. Trueblood, S. E. Burch, K. J. Kearfott, and K. W. Brooks, "Radiographic film digitization," in *Digital Imaging Proceedings of 1993 Summer School*, edited by W. R. Hendee and J. H. Trueblood, Medical Physics Monograph 22 (Medical Physics Publishing Co, Madison, WI, 1993).

<sup>10</sup>T. G. Langer, E. S. de Parades, S. Agarwal, and D. Smith, "QA in Mammography, Part 1: QA Physics," *Applied Radiology* (1992).

<sup>11</sup>J. P. Peng, K. J. Kearfott, J. H. Trueblood, and T. B. Kingsbury, "Variable illumination light box for film digitization," *Med. Phys.* **18**, 832 (1991).

<sup>12</sup>W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*, 2nd ed. (Cambridge U.P., Cambridge, 1992).

<sup>13</sup>R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (Addison-Wesley, New York 1992).

<sup>14</sup>R. L. Scheaffer and J. T. McClave, *Probability and Statistics for Engineers*, 2nd ed. (Duxbury, Boston, 1986).

<sup>15</sup>M. S. Chesters, "Human visual perception and ROC methodology in medical imaging," *Phys. Med. Biol.* **37**(7) (1992).

<sup>16</sup>D. Maar, *Vision* (W. Freeman, New York, 1982).

<sup>17</sup>A. Burgess, "Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation," *J. Opt. Soc. Am.* **2**, 1498-1507 (1985).