Computer-aided diagnosis: Detection and characterization of hyperparathyroidism in digital hand radiographs

Chair-Li Chang,[a] Heang-Ping Chan, Loren T. Niklason,[b] Mark Cobby,[c] Jeffery Crabbe, and Ronald S. Adler
Department of Radiology, University of Michigan Hospitals, Ann Arbor, Michigan 48109

(Received 7 February 1992; accepted for publication 13 May 1993)

An automated method is developed for the detection and staging of skeletal changes due to hyperparathyroidism on digitized hand radiographs. Subperiosteal bony resorption, particularly along the radial margins of the middle and proximal phalanges, is among the earliest manifestations of secondary hyperparathyroidism. In order to quantify the severity of bone resorption in these regions, the computer method analyzes the roughness of the phalangeal margins, as projected on the radiograph. The regions of interest, which contain the phalanges, are obtained from the digitized hand radiographs by an image preprocessor. The radial margin of each phalanx is detected by a model-guided boundary-tracking scheme. The roughness of these boundaries is then quantified by the mean-square variation and the first moment of the power spectrum. A receiver operating characteristic (ROC) study comparing the computer detection of hyperparathyroidism with the diagnosis by three experienced skeletal radiologists was performed by evaluating 84 hand images from 22 patients. Our present computer method can achieve a true-positive rate of 94% and a true-negative rate of 92%. Such a computer-aided diagnosis system may assist radiologists in their assessment of primary and secondary hyperparathyroidism, since it is both accurate and not subject to either intra- or interobserver variations.

Key words: computer-aided detection, hand radiographs, hyperparathyroidism, power spectrum, model-guided boundary tracking

I. INTRODUCTION

Hyperparathyroidism is an endocrine abnormality that manifests itself in the skeletal system in a predictable manner. Further, the activity of the disease correlates directly with the severity of skeletal changes. While this may be primary or secondary in etiology, the most common form of hyperparathyroidism in the United States is secondary and is related to chronic renal disease.

Hyperparathyroidism at its early stage causes subperiosteal resorption, which most often affects the radial margin of the middle and proximal phalanges of the index and middle fingers and all the terminal tufts; these features are the most sensitive and pathognomonic indicators of disease activity.[1–5] This resorption causes a fuzzy lace-like appearance of the phalangeal margin (Fig. 1), which progresses to a ragged spiculated appearance in the advanced stages of the disease. We therefore focused our investigation on the radial margin of the proximal and middle phalanges of the hand in this study to quantify the degree of subperiosteal resorption.

At present the diagnosis of hyperparathyroidism is made by detecting the appearance of the characteristic subperiosteal resorption on hand radiographs, by laboratory findings of increased level of immunoreactive parathyroid hormone (iPTH) and calcium in the blood.

Serum levels of iPTH concentration are generally more sensitive than the radiologic findings to diagnose secondary hyperparathyroidism. However, it has been observed[6] that iPTH levels may temporarily normalize due to fluctuating levels of serum calcium, while the radiologic examination still reveals evidence of subperiosteal resorption. Since changes in the bone would require substained parathyroid hormone stimulation or suppression for longer periods of time, the recognition of changes in the bone itself is obviously important. Although the radiologic findings may temporarily lag behind the biochemical changes, examination of hand radiographs is still the preferred method of evaluating hyperparathyroidism.

Because of the subtle nature of early subperiosteal resorption in the phalanges, there can be significant intra- and interobserver variations among radiologists in assessing the progression and regression of the disease. Computerized methods may minimize these variations by providing a reproducible, consistent, and objective evaluation of the extent of subperiosteal resorption of the phalanges. We have developed digital methods to analyze the bone resorption in hand radiographs; our approach to the determination of boundary roughness as a measure of subperiosteal resorption along the radial margins of the affected phalanges is presented here.

II. MATERIALS AND METHODS

A. Overall scheme of the computer detection algorithm

The overall scheme of our computer detection algorithm is shown in Fig. 2. There are five steps in the digital

 0094-2405/93/20(4)/983/10/$1.20

FIG. 1. Subperiosteal resorption on the phalangeal bone.

methods: digitization, preprocessing, boundary tracking, quantification of the roughness of the boundaries, and classification of normal and abnormal cases. These steps will be described in detail in the following sections.

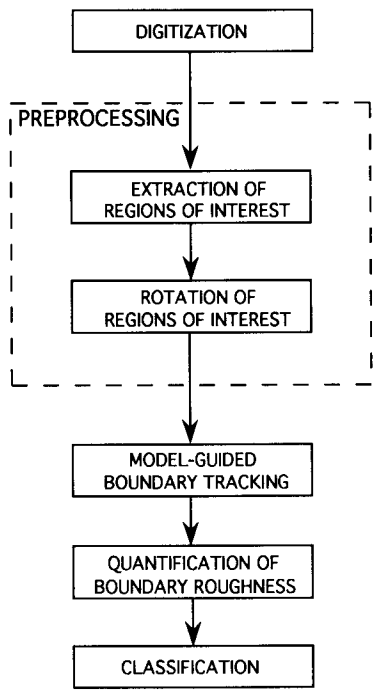

FIG. 2. Overall scheme of the computer detection algorithm.

B. Digitization of hand radiographs

Digital images were obtained by digitizing hand radiographs using a Lumisys (Sunnyvale, CA) DIS-1000 laser scanner. The hand radiographs were selected from files of patients with chronic renal failure by an experienced skeletal radiologist. Adult hand radiographs were taken with Dupont Quanta Detail/Cronex 10 screen-film systems and pediatric hand radiographs were taken with Kodak Lanex Fine screens with Kodak TMG films. The images were digitized at a pixel size of 0.1 mm$\times$0.1 mm and 4096 grey levels. The system is calibrated so that the optical density is linearly related to the pixel values in the range of 0.1–2.5 optical density units at 0.001 optical density units/pixel value. The gradient of the curve decreases gradually outside this optical density range.

C. Preprocessing—Extraction and rotation of regions of interest

To conserve image processing time, the regions of interest are extracted from the digitized hand images. An automated method[7] for the extraction of the ROI's and determination of the orientation of the phalanx is being developed. However, for the purpose of this study, these steps were performed manually to ensure their accuracy. Each ROI contains the proximal or middle phalanx of the index or the middle finger. The dimensions of the ROI's vary according to the size of the individual phalanx and range from 30$\times$50 pixels to 80$\times$110 pixels. If the phalanx is oriented at an angle relative to the vertical direction of the digitization matrix, the ROI is rotated so that the longitudinal axis of the phalanx is aligned with the vertical direction. This step is important because it standardizes the geometry of the spatial sampling for all phalanges, thereby reducing discrepancies in the measurement of boundary roughness introduced by different sampling geometries when the boundary points are tracked.

A new pixel in the rotated image is generated by interpolation in a 5$\times$5 pixel region centered at the pixel of interest, with a bicubic spline method.[8,9] The bicubic spline is chosen because of its property of smoothness in image interpolation, and being an orthogonally separable two-dimensional interpolation function. Furthermore, two kinds of errors, resolution error and interpolation error, may be introduced by an imperfect interpolation function. An imperfect interpolation function may attenuate the high-frequency components of the image leading to a loss of image resolution, or introduce higher-order spectral modes leading to interpolation error. By using higher-order interpolation functions, such as the bicubic spline, interpolation error is reduced at the expense of resolution error.[9]

One of the difficulties in edge detection is caused by the presence of image noise. In the image preprocessing step, the ROI may be processed with a median filter for noise reduction. A median filter[10] may remove impulsive noise while preserving the edge gradient. A median filter thus appears to be useful for our detection task. The effect of

such noise-removal preprocessing is also investigated herein.

D. Model-guided boundary tracking

The target sites of boundary tracking are at the radial margins of the middle and proximal phalanges of the index and middle fingers. In order to accurately detect the level of disease activity, it is important to determine the degree of roughness of the phalangeal boundaries, both at the early stage, when subperiosteal resorption is subtle, and at the advanced stage, when the boundaries are fuzzy. The hand radiograph has a relatively high contrast compared with other types of x-ray images. However, the signal-to-noise ratio is low along the fuzzy boundaries. This requires that the boundary tracking algorithm be sensitive to detect fuzzy boundary points with small edge gradients and be insensitive to image noise. Most edge-detection algorithms work well when the grey-level transition is abrupt, similar to a step function. However, as the transition region of the edge becomes broader and shallower, their accuracy decreases, due to the inability in distinguishing the noise or false edge points from real edge points.

The edge detector used in this study is a one-dimensional spline function,[11] which acts as a linear filter, as shown in Appendix A. This filter is convolved with an input window of pixel elements to obtain the filter response. The extrema of the filter response are located. However, the extrema may not correspond to real edge points in the input window. To reduce these spurious responses, a pattern matching technique,[11] assuming that only white Gaussian noise is present in the hand images, is used to select the real edge points. In the presence of other kinds of image noise including structured noise, the edge detector may be suboptimal, as is the case here. The filter response with the selected extrema can then be treated as a multiple-scale pattern in which edge points are of different gradients. So a gradient threshold is set to eliminate "small-gradient" edge points that may be caused by other image noise.

Our model-guided boundary tracking algorithm consists of two stages: model building and boundary tracking. For an input phalanx image, it first generates a model for its radial margin; the model is then used as a guide in the boundary-tracking stage to track the detailed features along the boundary accurately. Figure 3 is a flow chart of the model-guided boundary-tracking algorithm. In the model-building stage, the aim of our algorithm is to make use of the edge detector to locate the dominant outermost edge points along the radial margin. These edge points approximately describe the outline of the radial margin and are referred to as the model points of the phalanx. The model points are obtained by applying the edge detector line by line along the horizontal direction. In this study, the approximate x coordinate of a model point on the starting line is chosen by a human operator. A model-building window of 41 pixels wide is centered horizontally at the x coordinate of the model point. The filter response within the window is computed by convolving the spline function with the image profile along the line. From the extrema of

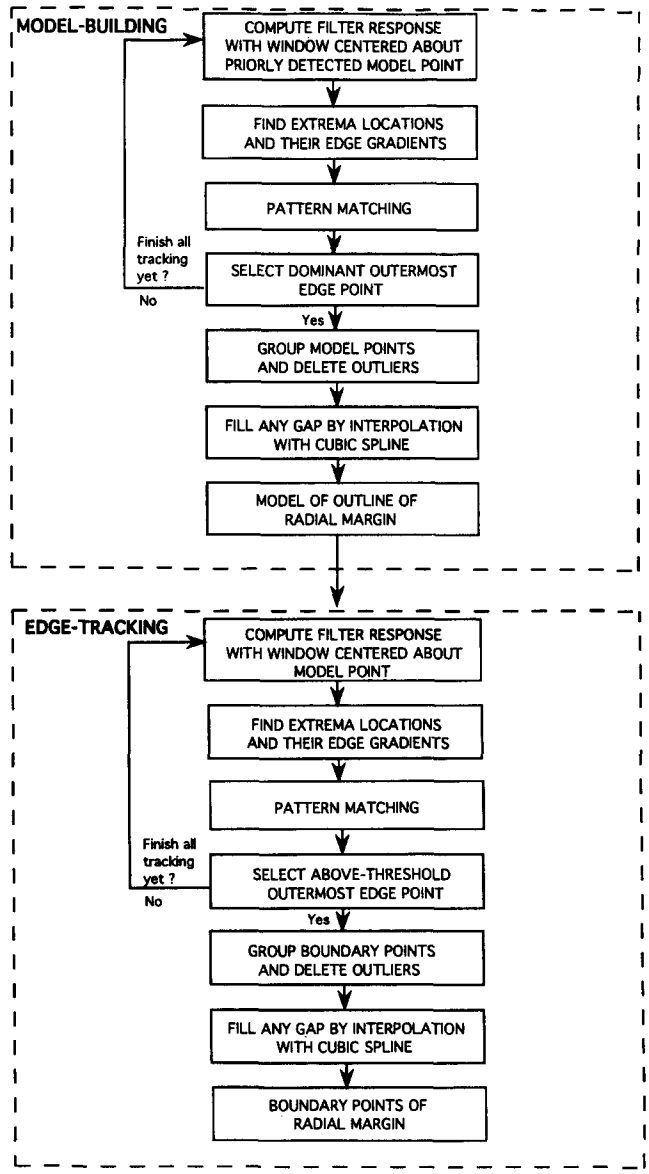

FIG. 3. Flow chart of the model-guided boundary-tracking algorithm.

the filter response, the algorithm chooses the first dominant edge point, which has a gradient greater than a threshold selected for model building, as the edge detector scans from the exterior to the interior across the boundary of the phalanx. The model-building gradient threshold is chosen to maximize correct detection based on the statistical properties of a population of typical hand images, as described below (Sec. III and Fig. 8). For subsequent lines, the 41-pixel model-building window is centered at the model point of the previous line. This procedure is repeated for 128 consecutive horizontal lines centered at the midpoint of the radial margin. The program then inspects the detected model points and groups them into one or more segments. Within a segment, the x coordinate of any model point must be one pixel within that of the model points in the adjacent horizontal lines. To further reduce spurious edge points due to noise, any segment that is composed of two or less model points is considered to be an outlier, and

is deleted. Thus, segments are separated by gaps due to deletion of outliers or the algorithm not being able to detect model points along one or more of the 128 consecutive horizontal lines. Based on the knowledge of continuity of the bone margin, all gaps will be filled by cubic spline interpolation using the remaining model points. The resulting 128 consecutive points form the model, which approximates the outline of the radial margin.

To more accurately locate the radial margin of phalanges affected by either subtle or severe subperiosteal resorption, the edge detector has to be more sensitive and thus be able to detect edge points with smaller gradients. The sensitivity of the edge detector may be increased by selecting a low threshold for the edge gradient. However, such an edge detector will also be sensitive to image noise and the tracking will tend to stray away from the radial margin. Our model-guided approach allows a sensitive edge detector to be used in the boundary-tracking stage while reducing the likelihood of straying away from the radial margin.

The edge detector used for boundary tracking uses the same spline function as that described for model building (Appendix A). However, at this stage, the algorithm uses an 11-pixel wide boundary-tracking window centered about the x coordinate of the model point obtained during the model-building stage. When the edge detector scans from the exterior to the interior across the boundary of the phalanx within the boundary-tracking window, the first edge point that exceeds a boundary-tracking gradient threshold is chosen as a boundary point. The width of the boundary-tracking window and the boundary-tracking gradient threshold are again optimized for correct detection based on the sample population of hand images (Sec. III and Figs. 9 and 10). The choice of the boundary-tracking gradient threshold represents a compromise such that the edge detector will be insensitive to noise, without sacrificing the ability of the edge detector to search out fuzzy edge points in the vicinity of the model points. This procedure is repeated for the same 128 consecutive horizontal lines along the radial margin, as in the model-building stage. The subsequent procedures of grouping boundary points, deleting outliers, and filling gaps are similar to those described above for the model points, except that in grouping boundary points, the x coordinate of any boundary point must be two, instead of one, pixels within that of the boundary points in the adjacent lines.

Figure 4 shows an example in which the model and boundary points are superimposed on the phalanx, the boundary points are judged visually to be more accurate than the model points in following the fuzzy details of the radial margin. This is also verified by the results from our ROC analysis (Sec. III and Figs. 8 and 9).

E. Quantification of boundary roughness

To characterize the severity of hyperparathyroidism in terms of subperiosteal resorption, the boundary roughness is quantified by the following two feature measures: mean-square variation, which quantifies spatially the deviation of the boundary points from an estimated natural smooth contour of the radial margin, and the first moment or the

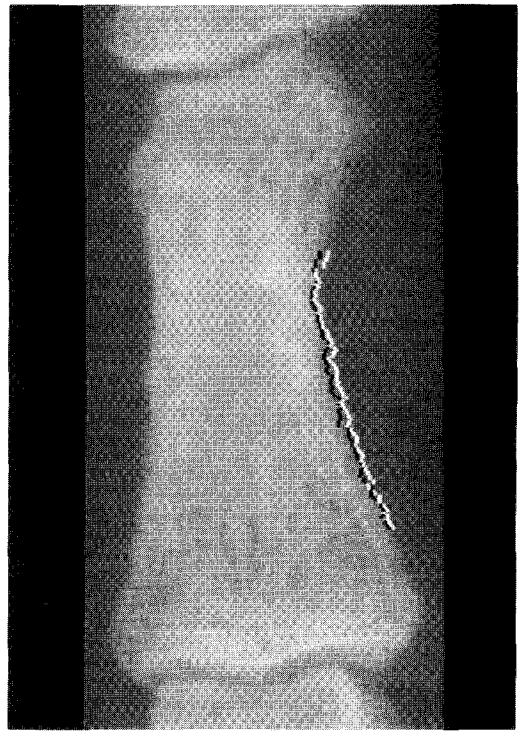

FIG. 4. Model and boundary points superimposed on the phalanx. Black dots represent the model points and white dots represent the boundary points. Note that some of the model and boundary points overlap each other.

weighted spatial frequency of the one-dimensional power spectrum. A high value of first moment indicates increased boundary roughness. These boundary-roughness measures are defined as follows:

$$\text{Mean-square variation} = (1/N)\Sigma_{n=1}^{N} |X(n) - \hat{X}(n)|^2, \tag{1}$$

where $\hat{X}(n)$ = estimated smooth contour of the radial margin, $X(n)$ = boundary points, and N = number of boundary points.

First moment of the power spectrum

$$= \sum_{k=0}^{N-2} \{[P(f_k) + P(f_{k+1})]/2\}(f_k + \Delta f/2), \tag{2}$$

where f_k = discrete spatial frequency, $P(f_k)$ = power spectral value at f_k, and Δf = frequency interval. The power spectrum is calculated from a finite number of samples with the maximum entropy method,[8,12] assuming that the boundary data is at least wide-sense stationary.

The natural shape of the bone contour varies from one phalanx to another. The variation in the location of the boundary points is caused largely by the bone contour. It is therefore important to correct for this variation in the boundary location before the true variation due to subtle erosion in the bony surface can be compared. We estimate the smooth bone contour by applying least-squares polynomial curve fitting to the detected boundary points (Sec. III and Fig. 11). The estimated contour is subtracted from the boundary points. The mean-square variation of the

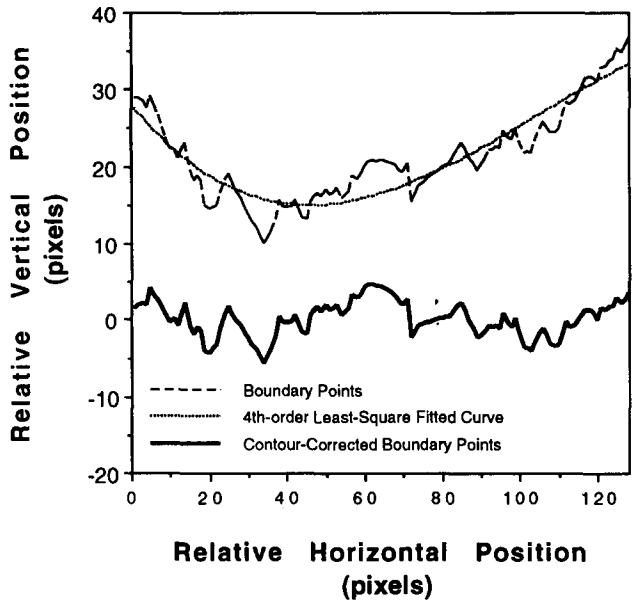

FIG. 5. Plot of the detected boundary points overlaid with its fourth-order least-squares fitted curve and the contour-corrected boundary.

contour-corrected boundary points will then reflect the actual roughness due to subperiosteal resorption. Figure 5 shows a plot of the detected boundary points overlayed with its fourth-order least-squares fit as well as the contour-corrected boundary points.

Figure 6 shows the power spectra of contour-corrected boundary points for an abnormal case with hyperparathyroidism and a normal case. The total area under the power spectrum curve is larger in the abnormal case than that of the normal case, indicating that the radial margin is rougher in the former. The shape of the curve also indicates that the average spatial frequency of the severe case is higher than that of the normal case resulting from the contribution of higher spatial frequency components along the severely affected radial margin.

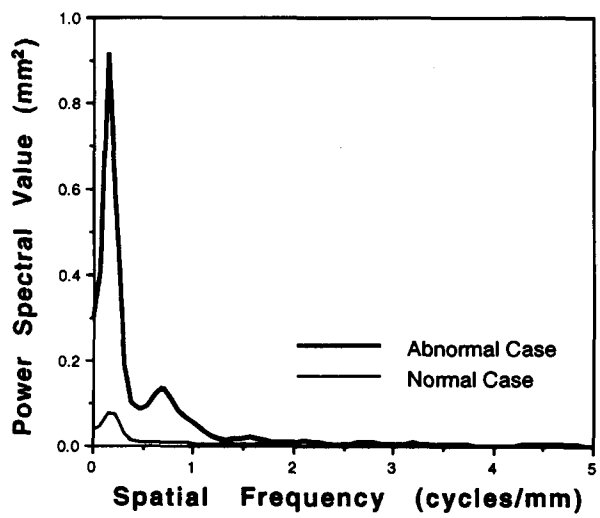

FIG. 6. Power spectra of contour-corrected boundary points for a normal case and an abnormal case of hyperparathyroidism.

F. Classification

The computer detection program was applied to each hand image to obtain the boundary points of the radial margin of each of the four phalanges. The mean-square variation and the first moment of the power spectrum for the four phalanges in each hand were calculated and the averages of these quantities were determined. The resulting averaged mean-square variation and the averaged first moment of the power spectrum were used as the components of the feature vector in the construction of scatter plots in a two-dimensional feature space. For each case, three experienced skeletal radiologists had ranked the severity of subperiosteal resorption from 0 to 3, with rank 0 as normal and rank 3 as florid, as described in Sec. II G. We used the radiologists' ranking as the "truth" for comparison with the computer's performance. For this purpose, all cases of rank 1 to rank 3 were grouped as abnormal cases and rank 0 as normal cases. Then a classifier is used to classify the cases as positive or negative using the computer extracted features. Based on the limited number of cases used in this study, our analysis shows that the features used for the classification are not normally distributed. Stepwise logistic regression,[13] which does not make the assumption of multivariate normality, is thus utilized to classify the features instead of using discriminant analysis,[13] which assumes multivariate normality. Stepwise logistic regression calculates the predicted probability of calling a case negative (or normal), as determined by the features. By selecting different decision thresholds on the predicted probabilities of calling a case negative, a ROC curve,[14–16] which expresses the relationship between the true-positive rate and the false-positive rate, can be constructed to evaluate the detection accuracy of the algorithm. The dependence of the computer detection accuracy on the image-processing parameters can then be determined by comparison of the ROC curves.

G. Observer study

An observer study was conducted using 84 hand images from 22 patients. The selected films demonstrated none or varying degrees of subperiosteal resorption. The left and right hands were treated as independent cases.

The observer study was conducted with three experienced radiologists who independently ranked each of the middle and proximal phalanges of the index and middle fingers based on the degree of subperiosteal resorption (i.e., boundary roughness). The four rankings are florid, definite, subtle and normal, or equivalently, rank 3, rank 2, rank 1, and rank 0, respectively. If there was substantial disagreement among the radiologists for a particular case, that is, if the difference in ranking was two or more, they were asked to rank that case again independently. If there was still substantial disagreement for that case, the three radiologists ranked the case together to reach a consensus. On the other hand, if there was no substantial disagree-

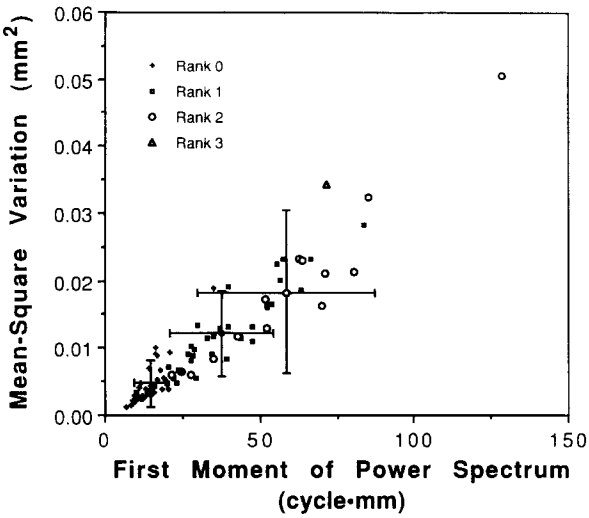

FIG. 7. Scatter plot of mean-square variation versus first moment of power spectrum for cases ranked 0, 1, 2, and 3. The mean and standard deviation of cases of the same rank are also shown. Only one case was ranked as 3 (florid).

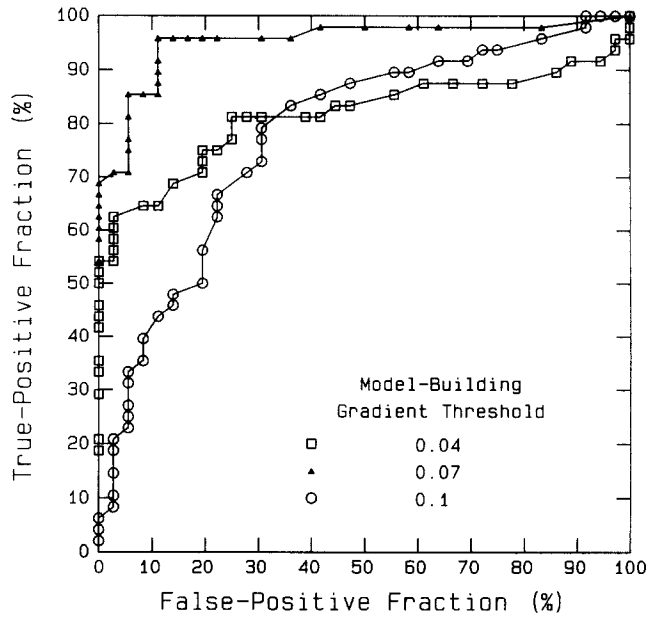

FIG. 8. Dependence of detection accuracy on the model-building gradient threshold. The least-squares fit was set at fourth order.

ment among the radiologists in ranking a particular case, the average ranking obtained from the individual rankings of the three radiologists is regarded as the consensus ranking. The intra- and interobserver variations are estimated by the intraobserver and observable interobserver standard errors, as described by Swets and Pickett.[15]

III. RESULTS

The distribution of the rankings among the 84 hands is as follows: 35 normal or rank 0, 34 subtle or rank 1, 14 definite or rank 2, and 1 florid or rank 3. Figure 7 shows the scatter plot of the two features for cases ranked 0, 1, 2, and 3. There is a positive correlation between the computer determined quantities and the radiologists' rankings. There are overlapping regions on the distributions of the features of these four groups. To visualize the extent of spread and overlapping, Figure 7 also shows the means and standard deviations of the feature measures for all the cases in each rank. At present, our "gold standard" for evaluation of the computer's performance is based on visual inspection of the hand radiographs by experienced skeletal radiologists. We do not know what portion of the overlap between rankings is due to difficulty in visually distinguishing the subtle difference in boundary roughness between cases of any two consecutive ranks. Thus we grouped all cases of rank 1 to rank 3 as abnormal cases and rank 0 as normal cases in our database for the following analysis.

To determine the optimal model-building gradient threshold, boundary-tracking gradient threshold, boundary-tracking window width and the least-squares fit order, we evaluated the effect of each parameter on detection accuracy using ROC analysis.

To investigate the effect of model-building gradient threshold on accuracy, only the model-building stage of the model-guided boundary tracking algorithm is executed. The model points obtained for the four phalanges in each of the 84 hand images are fitted with fourth-order least-squares fit to estimate the phalangeal contour. The two boundary-roughness measures are then calculated from the contour-corrected model points. The result is evaluated by plotting the ROC curves for a range of model-building gradient thresholds. A few representative ROC curves with model-building gradient thresholds of 0.04, 0.07, and 0.1 are shown in Fig. 8. By comparing the areas under the ROC curves for the different model-building gradient thresholds, we have found that the ROC curve of model-building gradient threshold of 0.07 provides the maximum area. Model-building gradient thresholds above 0.1 were also investigated, but the thresholds were often too high for the edge detector to detect sufficient edge points to give a good approximation to the outline of the radial margin in most of the phalanx images. On the other hand, model-building gradient thresholds below 0.04 were often too sensitive to image noise and the tracking strays away from the radial margin. It is also observed visually that a model-building gradient threshold of 0.07 does give a reliable approximation to the outline of the radial margin.

The next step was to determine the optimal value of the boundary-tracking gradient threshold with the model-building gradient threshold set at 0.04, 0.07, or 0.1, boundary-tracking window width at 11 pixels, and a least-squares fit of fourth order. From the ROC curves shown in Fig. 9, we found that there is no substantial difference in terms of the area under the ROC curves obtained with boundary-tracking gradient thresholds of 0.01, 0.04, and 0.07 at a model-building gradient threshold of 0.07. The ROC curves obtained with a boundary-tracking gradient threshold of 0.01, 0.04, or 0.07 and with model-building gradient thresholds of 0.04 and 0.1 (not shown) are all lower than the curves shown in Fig. 9. It is also observed that these ROC curves are higher than those with only the

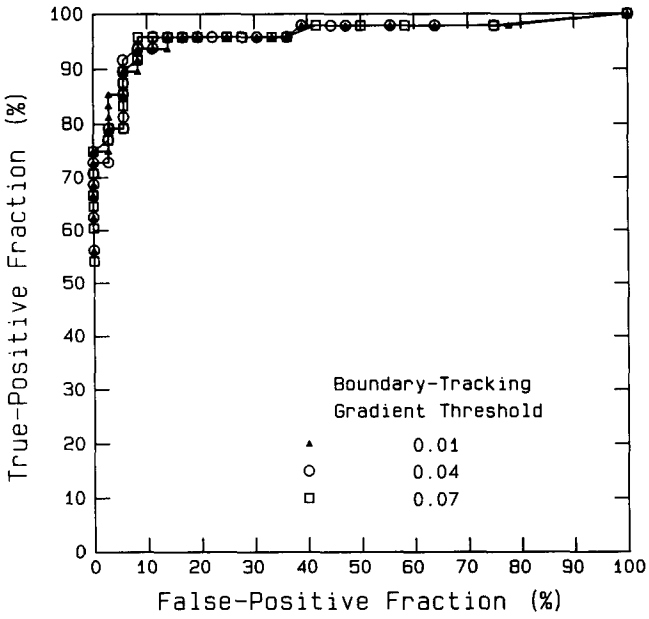

FIG. 9. Dependence of detection accuracy on the boundary-tracking gradient threshold. The model-building gradient threshold was set at 0.07, boundary-tracking window width of 11 pixels, and a fourth-order least-squares fit.

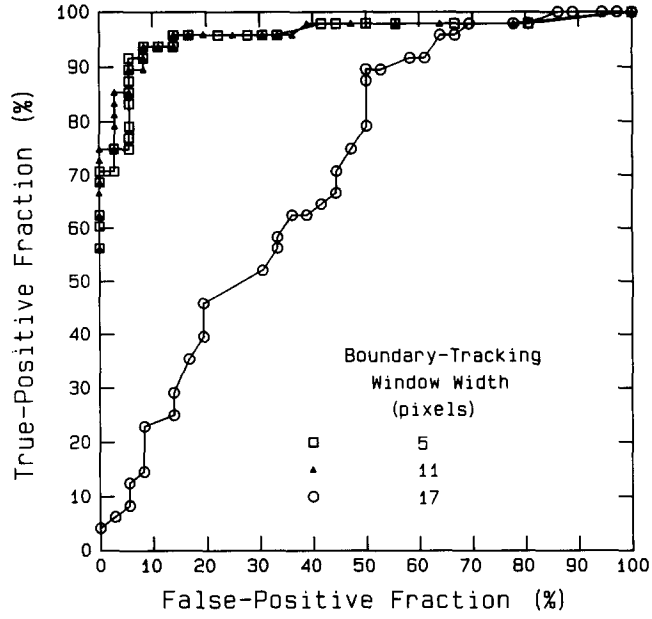

FIG. 10. Dependence of detection accuracy on boundary-tracking window width. The model-building gradient threshold was set at 0.07, boundary-tracking gradient threshold at 0.01 and a fourth-order least-squares fit.

model-building stage executed (Fig. 8). Therefore, although there is no substantial difference in detection accuracy among the different boundary-tracking gradient thresholds in the range 0.01 to 0.07, there is significant improvement in detection accuracy when the model-guided boundary-tracking method is used.

With the model-building gradient threshold at 0.07, the boundary-tracking gradient threshold at 0.01, and a least-squares fit of fourth order, we proceeded to choose the optimal boundary-tracking window width. Figure 10 shows ROC curves with boundary-tracking window widths of 5, 11, and 17 pixels. Boundary-tracking window width of 17 pixels is undesirable; this wide window combined with a sensitive edge detector results in a tendency of tracking to stray far from the model points. A boundary-tracking window width of 11 pixels was, therefore, chosen.

We have assessed the effect of the degree of least-squares fit on the accuracy of the computer detection with the model-building gradient threshold set at 0.07, the boundary-tracking gradient threshold at 0.01, and the boundary-tracking window width at 11 pixels. The resulting ROC curves are shown in Fig. 11. The curves for fourth and fifth orders are similar, whereas the curve for third order is slightly lower. Since high-order least-squares fits require more computation time, we have chosen a fourth-order least-squares fit.

The effect of noise removal with median filtering on the accuracy of computer detection was also investigated. With the model-building gradient threshold at 0.07, the boundary-tracking gradient threshold at 0.01, a boundary-tracking window width of 11 pixels and fourth-order least-squares fit, the detection accuracy obtained with and without median filtering is compared in Fig. 12. For a given false-positive detection rate, the true-positive detection rate was lower with median filtering than without median filtering. This result indicates that preprocessing with a noise smoothing filter may distort the subtle features along the boundary, thereby reducing the sensitivity of classification based on boundary roughness. It also indicates the degree

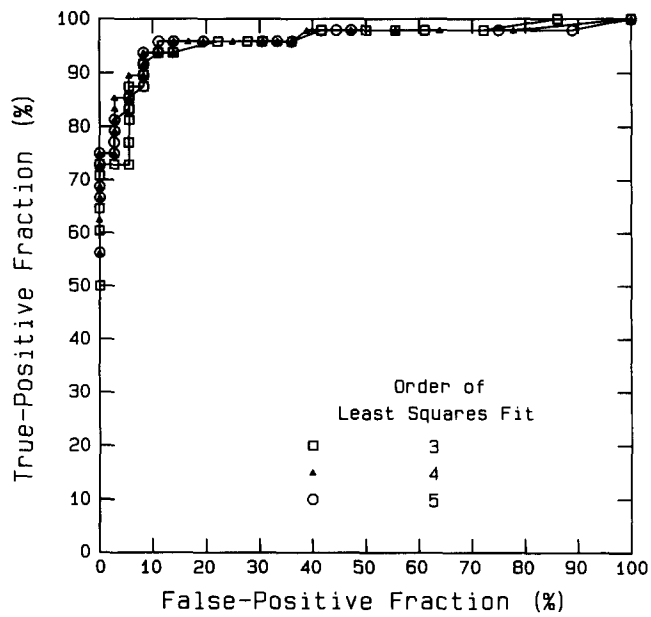

FIG. 11. Dependence of detection accuracy on order of least-squares fit. The model-building gradient threshold was set at 0.07, boundary-tracking gradient threshold at 0.01 and boundary-tracking window width of 11 pixels.

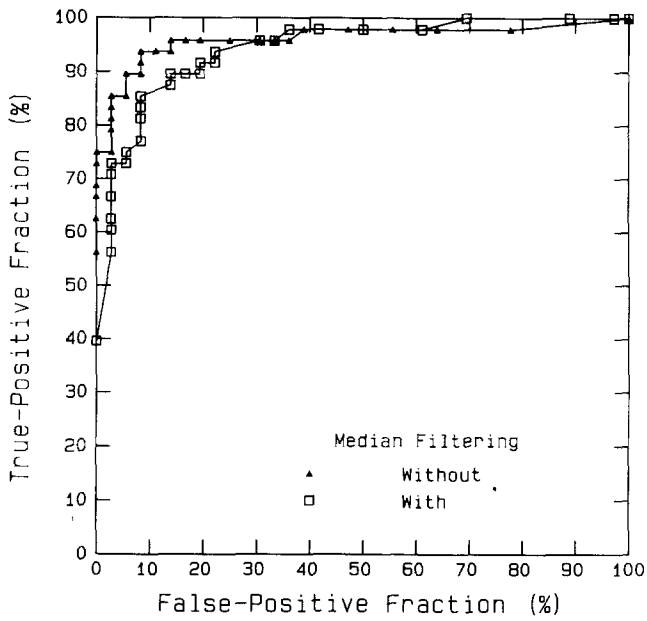

FIG. 12. Comparison of detection accuracy of the computer method with and without noise smoothing with median filtering. The model-building gradient threshold was set at 0.07, boundary-tracking gradient threshold at 0.01, boundary-tracking window width of 11 pixels and fourth-order least-squares fitted contour-corrected boundary points. The curve without median filtering represents the overall detection accuracy of our current computerized method. The computer achieves a true-positive rate of 94% at a true-negative rate of 92% (i.e., a false-positive rate of 8%). The area under the ROC curve is 0.96.

of difficulty and the sensitivity required for analysis of subtle subperiosteal resorption.

Using the set of parameters optimized for detection, we applied the program to the 84 hands. Figure 13 shows the histograms of predicted probabilities of calling a case negative for the normal and abnormal cases, as determined by logistic regression. The small overlap between the two

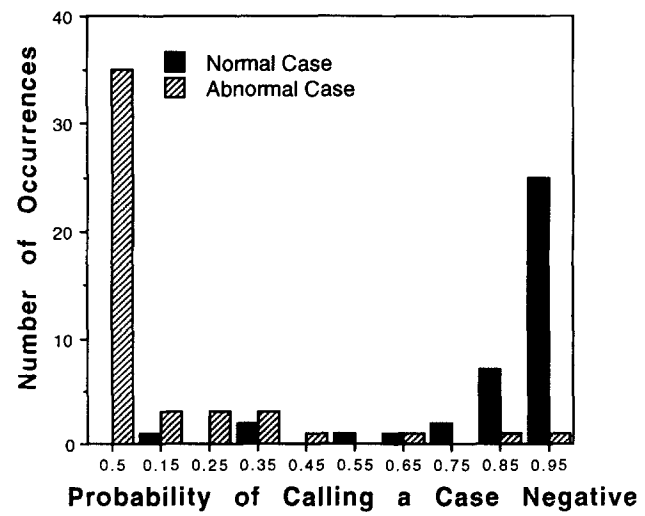

FIG. 13. Histograms of predicted probabilities of calling a case negative for the normal and abnormal cases in our database as determined by logistic regression. Each probability category has a range of ±0.5.

groups suggests that our feature vectors are good indicators of subperiosteal resorption caused by hyperparathyroidism. This assessment is further confirmed by the ROC curve without median filtering shown in Fig. 12. At a true-positive rate of 94%, the false-positive rate is 8%, which corresponds to a true-negative rate of 92%.

IV. DISCUSSION

We have developed a computer algorithm for the detection of boundary roughness of phalanges due to subperiosteal resorption, a skeletal manifestation of primary and secondary hyperparathyroidism. Our preliminary results suggest that computer-aided analysis is valuable in the detection of these skeletal changes.

One of the most important objectives of the detection algorithm is to accurately determine the boundary points of the radial margins of the phalanges. A pixel size of 0.1 mm was used. We have not attempted to investigate the relationship between this spatial resolution and the Nyquist criterion, neither have we compared the detection accuracy of this algorithm at higher resolution. Murphey[17] has shown by ROC studies that a pixel size of 0.08×0.08 mm is required for digitizing magnification (2×) hand radiographs in order to obtain observer performance in detection of mild subperiosteal resorption comparable to that achieved with the original magnification films. However, the digitization requirement for hand radiographs without magnification has not been studied. At present, the only digital systems available for general radiography are storage phosphor plate systems. These systems provide a high resolution mode of 0.1 mm×0.1 mm. In a practical clinical situation, both data archiving and transmission requirements limit the digitization resolution used.

The correspondence between edge points and extrema in the filter response using any type of edge detector is important. To accomplish a one-to-one correspondence, the edge detector is required to have finite dimension and be smooth everywhere.[11] The edge detector used in our algorithm satisfies such criteria, therefore theoretically, it gives exact correspondence between an edge point and an extremum in the filter response. If, instead, other kinds of edge detectors are used, which do not have finite dimension, such correspondence may not be justified; this causes inaccuracy in determining the edge location.

We have investigated the effect of the model-building gradient threshold on the tracking algorithm. The accuracy of the boundary-tracking algorithm depends strongly on the model-building stage. If the model-building stage gives a very poor approximation to the outline of the radial margin, the boundary points may not be included within the borders of the boundary-tracking window. We cannot simply increase the boundary-tracking window width because tracking may stray away from the model points due to various kinds of image noise, including structured noise such as calcified blood vessels. An adaptive method that estimates the boundary-tracking window width at each model point may be more desirable in order to accurately track the radial margin. Tradeoff between being insensitive to image noise and being sensitive to fuzzy boundary points

TABLE I. Intraobserver standard error for each of the three radiologists.

Consensus rank	Number of cases selected for repeated reading	Observer I	Observer II	Observer III
0	5	0.447	0.316	0.316
1	18	0.264	0.312	0.289
2	2	0.354	0.000	0.354
3	1	0.000	0.500	0.000

always exists. However, the gradient threshold may be adjusted adaptively at each model point depending on the signal-to-noise ratio within the window.

This study has primarily dealt with quantifying the boundary roughness of the radial margins of the middle and proximal phalanges of the index and middle fingers. To further improve the accuracy of discriminating abnormal from normal cases, additional features may be included: one good candidate that may characterize the radial margin is a feature that quantifies the edge strength of the boundary. Another possibility would be characterization of the trabecular pattern.[7] It has been suggested that tuftal erosion, in which the sharp cortical outline of the terminal tuft is lost, is the earliest manifestation of hyperparathyroidism. This could also be characterized by a feature and used, in addition to those features already mentioned, to discriminate the presence and absence of disease.

Since visual assessment of hand radiographs by experienced radiologists is the preferred method of evaluating hyperparathyroidism, the consensus of the three experienced radiologists was used as the "gold standard." This standard is limited by intra- and interobserver variations, which reduce accuracy in the visual evaluation of progressive subperiosteal resorption. Thus, the evaluation of the accuracy of the classifier trained, based on such a "gold standard," may be biased. We have measured the accuracy of the observers in terms of the intra- and interobserver variations in ranking. Table I and Table II, respectively, show the standard errors[15] estimated due to these variations. If accurately trained, the computer algorithm may be a more consistent, reproducible, and objective method in evaluating the phalangeal surface roughness due to subperiosteal resorption in hyperparathyroidism.

Processing time is a crucial consideration in the clinical settings. In this study, we have not attempted to minimize the computational time. The most important objective in this study was to obtain a high detection accuracy. In the future, we will evaluate the tradeoffs between detection accuracy and computational cost for various parameters used in boundary-tracking and boundary-roughness quantification.

TABLE II. Observable interobserver standard error for cases within each consensus rank.

Consensus rank	Observable interobserver standard error
0	0.402
1	0.495
2	0.535
3	0.577

A fully automated detection system has yet to be developed to be practical in a clinical setting. The selection of the ROI's has to be automated and be accurate. Various input parameters for the computer algorithm need to be determined in an adaptive manner. More features should be used to improve the accuracy in quantification of the boundary roughness. The tufts and trabecular pattern may also be included for the evaluation of hyperparathyroidism. More importantly, a large database that includes case samples that sufficiently represent the population will be required for training of a clinically practical system. Further, a reliable "truth" assessment for the case samples is also critically important. Because of the difficulty of correlating a biochemical test result with the radiographic manifestation, as discussed above, an approach to this problem may be the use of a larger number of experienced skeletal radiologists in the interpretation of the case samples and their consensus used as the "truth." This is equivalent to incorporating the knowledge of a collection of experienced radiologists into the computer-aided diagnosis system. The system thus trained may serve as an experienced consultant, which is both accurate and reproducible, to reduce variation in the performance of human observers. This is partly inferred from our results that, using the same set of "optimal" parameters trained with the consensus "truth" data from three radiologists, the ROC curve of the classifier with the consensus "truth" data (Fig. 12) is either comparable to or higher than the three ROC curves (not shown) with the individual "truth" data. The areas under the ROC curves are 0.96, 0.96, 0.95, and 0.94, respectively; or alternatively, at a false-positive rate of 8%, the true-positive rates are 94%, 93%, 87%, and 76%, respectively. The classifier therefore has been trained to provide a diagnosis in close agreement with the panel of radiologists and, as such, it may provide a second opinion to the individual radiologists if it is consulted during their individual readings.

V. CONCLUSION

We have developed a computer algorithm to detect subperiosteal resorption in patients with hyperparathyroidism from digitized hand radiographs. Our preliminary results indicate that we can achieve a sensitivity of 94% at a specificity of 92%, based on the consensus rankings of three radiologists. Our observer study indicates that there are substantial intra- and interobserver variations in the evaluation of the subtle skeletal changes, even among experienced radiologists. The computer-aided diagnosis system, if accurately trained, may improve the reproducibility and consistency in detection and staging of hyperparathyroidism. Further improvement in the boundary-tracking technique is needed to ensure accurate tracking of even the severely affected phalanges. Finally, a larger database must

be collected in order to establish the reliability and consistency of the computerized method when applied to cases encountered in the general population.

APPENDIX A: EDGE DETECTION

The edge detector used in our study has the following form:[11]

$$\phi'(t)=\lambda\begin{cases}\dfrac{-6t(t+A)}{A^3}, & -A\leqslant t\leqslant 0,\\ \dfrac{6t(t-A)}{A^3}, & 0<t\leqslant A,\\ 0, \quad \text{otherwise,}\end{cases} \tag{A1}$$

where $\phi'(t)$ is the derivative of $\phi(t)$ and λ is a normalizing constant, such that $\int_{-\infty}^{\infty}\phi(t)G(t)dt=1$. $G(t)$ is the point spread function of the digitizer and is approximated by the Gaussian function:

$$G(t)=\frac{1}{\sigma\sqrt{2\pi}}e^{(-t^2/2\sigma^2)}, \tag{A2}$$

where $\sigma>0$ is determined by the system. In this study, σ is assumed to be 0.8 pixels, as suggested in Ref. 11, and

$$\phi(t)=\lambda\begin{cases}\dfrac{-t^2}{A^3}(2t+3A)+1, & -A\leqslant t\leqslant 0,\\ \dfrac{t^2}{A^3}(2t-3A)+1, & 0<t\leqslant A,\\ 0, \quad \text{otherwise,}\end{cases} \tag{A3}$$

$\phi(t)*G(t)$, where $*$ denotes the convolution operator, is the scaled pattern used in pattern matching[11] to find the real edge points from the extrema $T(t_j)$ of the filter response $T(t)$, for $j=1,\ldots$, which are superimposed with noise response. To have a match, the mean and variance of $T(t)-T(t_j)[\phi(t)*G(t)]$ have to be approximately equal to $E\{n\}\int_{-\infty}^{\infty}\phi'(t)dt$ and $\int_{-\infty}^{\infty}|F[\phi'](s)|^2P_n(s)ds$, respectively, where $E\{n\}$ is the ensemble average of an ergodic noise process n, which is assumed to be white Gaussian noise; $F[\phi'](s)$ is the Fourier transform of $\phi'(t)$; and $P_n(s)$ is the power spectrum of n.

[a]Current address: Bioengineering Program, University of Michigan, Ann Arbor, Michigan 48109.
[b]Current address: Radiological Sciences and Technology, Massachusetts General Hospital, Boston, Massachusetts 02114.
[c]Current address: Department of Radiology, Bristol Royal Infirmatory, Marlborough Street, Bristol, BS2 8HW, England.

[1]D. G. Pugh, "Subperiosteal resorption of bone: A roentgenologic manifestation of primary hyperparathyroidism and renal osteodystrophy," Am. J. Roentgenol. **66**, 577–586 (1951).
[2]M. Sundaram, P. F. Joyce, J. B. Shields, M. A. Riaz, and S. Sagar, "Terminal phalangeal tufts: Earliest site of renal osteodystrophy findings in hemodialysis patients," Am. J. Roentgenol. **133**, 25–29 (1979).
[3]D. Resnick and G. Niwayama, *Diagnosis of Bone and Joint Disorders* (Saunders, Philadelphia, 1988).
[4]P. S. Jensen and A. S. Kliger, "Early radiographic manifestations of secondary hyperparathyroidism associated with chronic renal disease," Radiology **125**, 645–652 (1977).
[5]D. Resnick, L. J. Deftos, and J. G. Parthemore, "Renal osteodystrophy: Magnification radiography of target sites of absorption," Am. J. Roentgenol. **136**, 711–714 (1981).
[6]H. E. Meema, D. G. Oreopoulos, and T. M. Murray, "Periosteal resorption and periosteal neostosis: Comparison of normal subjects and renal failure patients on chronic ambulatory peritoneal dialysis using MOP-3 image analysis system and a grading method," Skeletal Radiol. **15**, 14–20 (1986).
[7]S. N. C. Cheng, H. P. Chan, R. S. Adler, L. T. Niklason, and C. L. Chang, "Development of a neural network for early detection of renal osteodystrophy," *Proceedings of the SPIE*, Biomedical Image Processing II, 1991, Vol. 1450, pp. 90–98.
[8]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Verrerling, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University, Cambridge, 1988).
[9]W. K. Pratt, *Digital Image Processing* (Wiley, New York, 1978), pp. 113–120.
[10]A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).
[11]D. Lee, "Coping with discontinuities in computer vision: Their detection, classification and measurement," IEEE Trans. Pattern Anal. Machine Intell, PAMI-12, 1990, pp. 321–344.
[12]S. M. Kay, *Modern Spectral Estimation: Theory and Application* (Prentice-Hall, Englewood Cliffs, NJ, 1988).
[13]W. J. Dixon, *BMDP Statistical Software Manual* (University of California, Berkeley, CA, 1990).
[14]C. E. Metz, "ROC methodology in radiologic imaging," Invest. Radiol. **21**, 720–733 (1986).
[15]J. A. Swets and R. M. Pickett, *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory* (Academic, New York, 1982).
[16]D. McNicol, *A Primer of Signal Detection Theory* (Allen and Unwin, London, 1972).
[17]M. D. Murphey, "Digital skeletal radiography: Spatial resolution requirements for detection of subperiosteal resorption," Am. J. Roentgenol. **152**, 541–546 (1989).