Snakules: A Model-Based Active Contour Algorithm for the Annotation of Spicules on Mammography

Gautam S. Muralidhar*, *Student Member, IEEE*, Alan C. Bovik, *Fellow, IEEE*, J. David Giese,
Mehul P. Sampat, *Member, IEEE*, Gary J. Whitman, Tamara Miner Haygood, Tanya W. Stephens, and
Mia K. Markey, *Senior Member, IEEE*

Abstract—We have developed a novel, model-based active contour algorithm, termed *"snakules"*, for the annotation of spicules on mammography. At each suspect spiculated mass location that has been identified by either a radiologist or a computer-aided detection (CADe) algorithm, we deploy *snakules* that are converging open-ended active contours also known as snakes. The set of convergent *snakules* have the ability to deform, grow and adapt to the true spicules in the image, by an attractive process of curve evolution and motion that optimizes the local matching energy. Starting from a natural set of automatically detected candidate points, *snakules* are deployed in the region around a suspect spiculated mass location. Statistics of prior physical measurements of spiculated masses on mammography are used in the process of detecting the set of candidate points. Observer studies with experienced radiologists to evaluate the performance of *snakules* demonstrate the potential of the algorithm as an image analysis technique to improve the specificity of CADe algorithms and as a CADe prompting tool.

Index Terms—Active contours, computer-aided detection, mammography, snakes, snakules, spicules.

I. INTRODUCTION

BREAST cancer manifests as various findings on mammography—microcalcifications, masses (spiculated and nonspiculated), and architectural distortions. Spiculated masses are characterized by a pattern of radiating lines known as spicules that emanate from a central mass. Spiculated masses have a much higher risk of malignancy than nonspiculated

*G. S. Muralidhar is with the Department of Biomedical Engineering, The University of Texas, Austin, TX 78172 USA (e-mail: gautam@mail.utexas.edu).

J. D. Giese and M. K. Markey are with the Department of Biomedical Engineering, The University of Texas, Austin, TX 78172 USA (e-mail: davidgiese@mail.utexas.edu; mia.markey@mail.utexas.edu).

A. C. Bovik is with the Department of Electrical and Computer Engineering, The University of Texas, Austin, TX 78712 USA (e-mail: bovik@ece.utexas.edu).

M. P. Sampat is with the Department of Neurology, University of California, San Francisco, CA 94143 USA (e-mail: mehul.sampat@ucsf.edu).

G. J. Whitman, T. M. Haygood, and T. W. Stephens are with the Department of Diagnostic Radiology, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030 USA (e-mail: gwhitman@mdanderson.org; tamara.haygood@mdanderson.org; tstephens@mdanderson.org)

masses and calcifications and hence it is crucial to detect spiculated masses [1]. CADe systems have been developed to assist radiologists in detecting signs of early breast cancer [2]–[4]. Studies have shown CADe systems to perform consistently better on microcalcifications than on masses, and in particular, the detection performance on spiculated masses is not optimal [5]–[7]. This has prompted several research groups to develop CADe systems specifically designed to detect spiculated masses [8]–[13].

Most CADe systems are comprised of two stages: a high sensitivity stage to detect suspect lesion locations on the mammogram and a high specificity stage to reduce the number of false positive (FP) candidates that do not correspond to actual lesions. The final outcome of a CADe system is usually a set of marks (also referred to as prompts) on the mammogram identifying suspect lesion locations.

CADe systems designed to detect spiculated masses usually employ strategies to detect patterns of converging lines. For instance, Karssemeijer and te Brake deployed derivatives of Gaussians at multiple scales to estimate orientations of pixels towards detecting spiculated masses [9]. The pixel orientation map was used to compute features that were sensitive to patterns of converging lines, and these features were used for classifying suspect locations as possible spiculated masses. Zwiggelaar *et al.* proposed a statistical approach based on factor analysis to describe oriented patterns of linear structures characteristic of spiculated masses [13]. The appearance of the central mass was statistically modeled using local-scale orientation signatures extracted from recursive median filtering and approximated using principal component analysis. More recently, our group proposed a model-based framework for the early detection of spiculated masses on mammography [8]. What makes our approach novel is that we measured physical properties of spiculated masses on a number of mammograms and developed a statistical model from these measurements. The statistical model serves as the basis for determining the parameters of a novel class of pattern matching filters, which we term spiculated lesion filters that are deployed to aid in the detection of spiculated masses and architectural distortion. Other methods for detecting spiculated masses have included the use of directional wavelets and multiresolution analysis of mammograms [10], [11].

Most CADe systems for spiculated masses are developed around the central theme of detecting radial patterns of converging lines. Consequently most of these methods achieve high sensitivity in detecting suspect spiculated mass locations.

However, a mammogram invariably contains other normal linear structures that are superimposed on one another and resemble a pattern of converging lines. Such locations are routinely marked as suspect locations by the detection algorithms resulting in a high FP rate. For example, the oriented pattern method proposed by Zwiggelaar *et al.* achieved a sensitivity of 80% at 0.014 false positives per image (FPI) for lesions having diameters exceeding 16 mm in a data set containing 29 spiculated masses and 100 normals [13]. However, Zwiggelaar *et al.* reported that the specificity of their orientation pattern method fell below acceptable levels on smaller spiculated masses [13]. The method proposed by Karssemeijer and te Brake achieved a sensitivity of 80 % at 0.4 FPI on a data set containing 19 stellate distortions (spiculated masses and architectural distortions) and 31 normals [9]. However, the attained FPI rate was 1 at a sensitivity of approximately 90% [9]. Likewise, the model-based detection of spiculated masses proposed by our group achieved a sensitivity of 80% at approximately 0.75 FPI on a dataset comprising 50 spiculated masses and 50 normals [8], but the FPI rate increased to 2.7 at a sensitivity of 88%. As pointed out by Zwiggelaar *et al.*, the specificity of CADe systems designed to detect oriented patterns of converging lines could be improved if the radiating spicules could be discriminated from other linear structures in the mammogram [13]. This calls for methods that could accurately capture the profile of spicules and other linear structures. Another problem that exists in the evaluation of these different CAD systems (e.g., [8], [9], and [13]) is that each system has been evaluated on a different, limited dataset of spiculated masses and it is hard to judge how these systems would perform on a single, large dataset of spiculated masses and normals. Even clinical studies that have evaluated the performance of commercial CADe systems have found that the detection performance of these systems on stellate distortions is less than acceptable. For example, Baker *et al.* [14] found that for detection of architectural distortion, one commercial system achieved a per-image sensitivity of 38% at 0.70 FPI while another achieved a per-image sensitivity of 21% at 1.27 FPI. These numbers highlight the need for methods to improve the overall performance of CADe systems designed to detect spiculated masses and stellate distortions in general.

Another problem that exists with CADe systems for spiculated masses is the prompting strategy. As discussed by Astley [15], in the case of spiculated masses, it is not always clear where to place the symbolic prompt. A symbolic prompt could be placed at the center of the mass, or at the point of convergence (focus) of the spicules [15]. Yet, both these placements of the symbolic prompt might not be optimal in terms of helping a radiologist's efficiency in arriving at a diagnostic decision. CADe systems usually do not explicitly annotate spicules and hence it is left to the radiologist to determine if the symbolic prompt corresponds to a spiculated mass or not. While most radiologists can detect obvious spicules, it would be beneficial if algorithms were developed to detect and explicitly annotate spicules. Such algorithms might help improve clinical productivity.

In this paper, we present *"Snakules"*—a model-based active contour algorithm for the annotation of spicules on mammography. At each suspect spiculated mass location that has been identified by either a radiologist or a CADe algorithm, we de-ploy *snakules* that consist of converging open-ended active contours also known as snakes [16]. The set of convergent *snakules (snakes that seek spicules)* have the ability to deform, grow, and adapt to the true spicules in the image, by an attractive process of curve evolution and motion that optimizes the local matching energy. The algorithm is model-based in that statistics of physical measurements of spiculated masses collected from mammograms are used to detect the candidate *snakule* points from where the snakes originate.

It is important to note that other groups have tried to extract properties specific to spicules to improve specificity of CADe systems. Zwiggelaar *et al.* demonstrated the use of cross-sectional intensity profiles as a basis for classifying linear structures seen on a mammogram with particular emphasis on correctly recognizing spicules and ducts [17]. However, Zwiggelaar *et al.* do not explicitly seek to capture the path of linear structures such as spicules (by trying to accurately trace the path of a linear structure); rather, they collect cross-sectional profile information from each linear structure detected on a mammogram using line detection operators and classify them into anatomical types by using a classifier trained on ground truth and cross-sectional information. Karssemeijer and te Brake have exploited the property of convergence of linear structures to detect breast abnormalities (stellate distortions) without explicitly trying to capture the path of linear structures [9]. Qian *et al.* have shown that directional wavelets in conjunction with contour tracing can be used to trace spicules explicitly [18]. While the algorithm developed by Qian *et al.* is able to detect spicules of different sizes, it suffers from some drawbacks, most notably translation/rotational dependence of the decomposition used and *ad hoc* spicule aggregation rules with lack of model-based design mechanisms. There are other groups that have attempted to deploy snake-like devices to segment solid masses and use feature extraction strategies to classify the masses (e.g., [19]–[21]); however, to the best of our knowledge, our approach is the first attempt aimed at explicitly capturing the path of spicules using deformable active contours in a bid to use them for distinguishing spiculated masses from other breast structures.

II. MATERIALS AND METHODS

A. Data Set

The data set for this study consisted of 52 spiculated masses identified on mediolateral oblique views that were randomly selected from the digital database for screening mammography (DDSM) [22], [23]. Cases from two different scanners, LUMISYS (scanner resolution 50 μm per pixel) and HOWTEK (scanner resolution 43.5 μm per pixel), and a range of density and subtlety ratings were represented in this cohort. Table I summarizes the statistics of the dataset. Regions of interest (ROIs) were extracted from each case and the experiments were conducted on the ROIs. Each ROI was defined such that the central mass and all the spicules were clearly included. Out of the 52 ROIs, 21 ROIs that were digitized using the LUMISYS scanner were selected and used to collect measurements of physical properties of spiculated masses from experienced radiologists (measurement set in Table I) [24]. These measurements

TABLE I
STATISTICS OF THE DATASET FROM DDSM USED IN THIS STUDY (SUBTLETY RATINGS WERE OBTAINED FROM DDSM, WHERE SUBTLE = 1, AND OBVIOUS = 5)

Number of cases for each scanner		Number of cases for each density rating				Number of cases for each subtlety rating				
LUMISYS	HOWTEK	Density 1	Density 2	Density 3	Density 4	Subtlety 1	Subtlety 2	Subtlety 3	Subtlety 4	Subtlety 5
Statistics of the measurement set										
21	0	7	7	7	0	0	0	1	5	15
Statistics of the development set										
5	6	4	2	3	2	1	1	2	2	5
Statistics of the evaluation set										
11	9	4	6	4	6	0	2	5	5	8

included the major axis of the central mass region, the width of a spicule, length of a spicule, and the number of spicules around the central mass. We have previously shown that experienced radiologists can reliably measure these physical properties of spiculated masses with a high degree of interobserver agreement [24]. These measurements on real spiculated masses provided valuable information when designing the *snakules* algorithm. The remaining 31 ROIs were broken down into two sets: a set of 11 ROIs (development set in Table I) was used for the development and validation of the *snakules* algorithm and a set of 20 ROIs was reserved exclusively for the evaluation of the algorithm (evaluation set in Table I). The ROIs on which the *snakules* were deployed were sub-sampled by a factor of four (along both the row and the column dimension). This was done to reduce the computation time. Institutional review board approval was obtained prior to this study.

B. Detecting Candidate Snakule Points

The first step in the *snakules* algorithm is to detect a set of points on the image from where spicules originate in the region around the suspect spiculated mass location identified by either a CADe algorithm or a radiologist. These candidate points represent locations on the image where *snakules* would be deployed. The problem of detecting the candidate *snakule* points can be mathematically stated as the problem of finding the set

$$C = \left\{ \begin{array}{l} (x,y) \ni (x,y) \in N, |\theta_{x,y} - \psi_{x,y}| < R/r_{x,y}, \psi_{x,y} \in k, \\ \forall (p,q) \in N, |\theta_{p,q} - \psi_{p,q}| < R/r_{p,q}, \psi_{p,q} \in k : r_{x,y} \le r_{p,q} \end{array} \right\}$$
(1)

where (x,y) is a candidate *snakule* point, N represents a neighborhood of pixels under consideration around the suspect spiculated mass location, (x_c, y_c) represents another point in the neighborhood N, θ is the dominant pixel orientation at a point in N (e.g., $\theta_{x,y}$ is the dominant pixel orientation at (x,y)), ψ is the direction of a point in N with respect to (x_c, y_c) (e.g., $\Psi_{x,y}$ is the direction of the point (x,y) with respect to (x_c, y_c)), R is the radius of a circular disk centered on (x_c, y_c) and towards which the pixel at a given location is directed, r is the Euclidean distance between a point and (x_c, y_c) (e.g., $r_{x,y}$ represents the Euclidean distance between (x,y) and (x_c, y_c)), and k represents the k th orientation bin. The direction ψ at a point, (e.g., (x,y)), is computed as $\psi_{x,y} = \tan^{-1}((y_c - y)/(x_c - x))$.

The condition $|\theta_{x,y} - \psi_{x,y}| < R/r_{x,y}$ is the same as defined by Karssemeijer and te Brake [9], in that we consider pixels in a

neighborhood around the suspect spiculated mass location that are directed towards a circular disk of radius R centered on the suspect spiculated mass location. However, this condition alone will not suffice to detect the candidate *snakule* origin points. This condition will yield all the points in the neighborhood N that are directed towards the suspect spiculated mass location. However, we require those pixel points that are not only directed towards the suspect spiculated mass location, but are also closest to the mass. This is captured by the condition on the second line of (1), which ensures that of all the pixels that are directed towards the central mass region and whose directions with respect to the suspect spiculated mass location fall in the same orientation bin k, only the point that is closest to the suspect spiculated mass location will be selected. These two conditions yield a set of pixel locations whose dominant orientations are directed towards the central mass region and are closest to the central mass.

The detection of candidate *snakule* points is carried out on *steerable filtered-Radon enhanced* ROIs rather than on the ROIs cropped directly from the mammograms. The primary motivation behind this is to mitigate the effects of noise and clutter caused due to overlapping out-of-plane tissue structures, which is a common occurrence in mammography due to the projection of 3-D breast structures onto a 2-D image plane. Radon enhancement of spiculated lesions on mammograms is explained in detail in [8]. We briefly review the procedure here.

The Radon transform of a continuous function $f(x,y)$ is defined as

$$g(\rho, \theta) = \int_{y=-\infty}^{y=+\infty} \int_{x=-\infty}^{x=+\infty} f(x,y)$$
$$\times \delta(\rho - x\cos(\theta) - y\sin(\theta)) \, dx dy \quad (2)$$

where ρ and θ are the parameters of the Radon domain and δ is the Dirac Delta function. Extending the above definition to images, the integral denotes that the value of $g(\rho, \theta)$ for any given (ρ, θ) is the integrated density of the image $f(x,y)$ along the line $\rho - x\cos(\theta) - y\sin(\theta) = 0$. Thus, an n-pixel thick line in the image space $f(x,y)$ maps to n-points located along a column in the Radon domain. It is possible to selectively enhance local peaks comprising a certain number of points in the Radon domain, which corresponds to a linear structure in the image space of the same thickness as the number of points in the Radon domain. The enhancement of local peaks was achieved through a simple linear column filter whose design was based on the average width of the spicules. Once the peaks had been enhanced, we used a Hamming window to attenuate the high
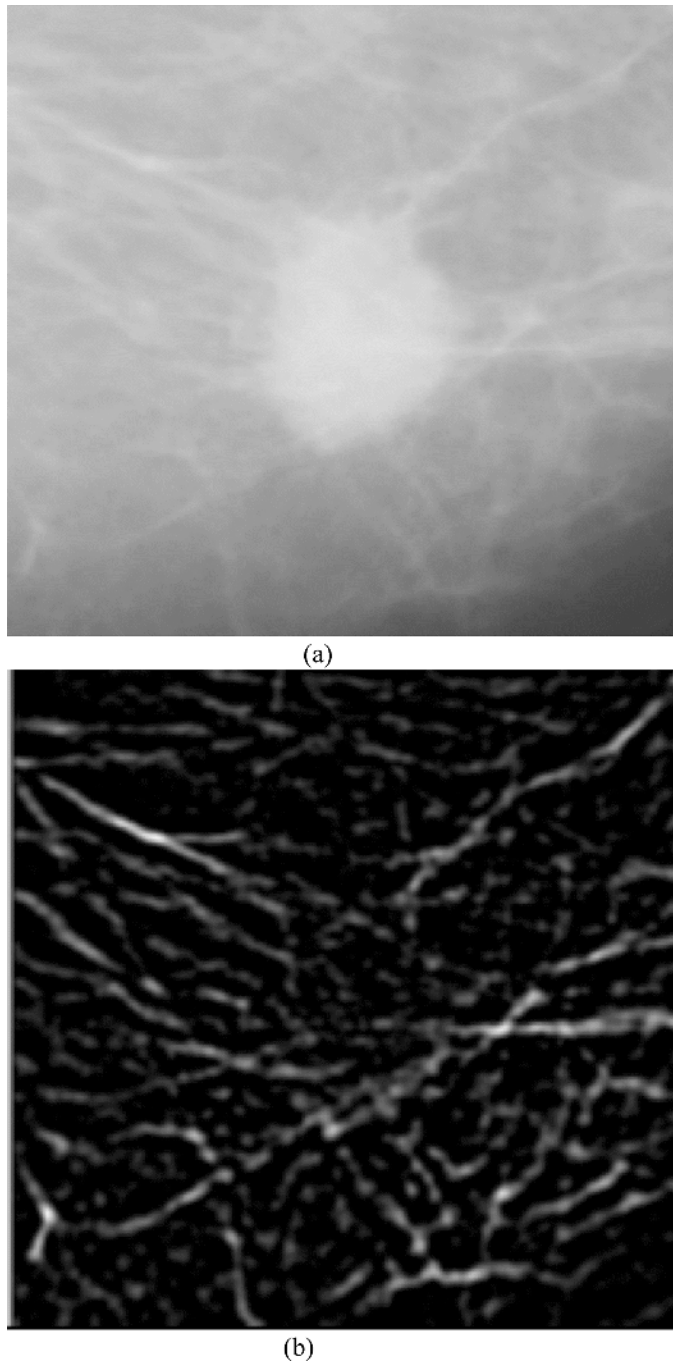
(a)

(b)

Fig. 1. (a) An ROI from a mammogram depicting a spiculated mass. (b) Radon enhanced ROI depicting enhanced curvilinear structures.

frequency noise in the Radon projections (the set of line integrals). Then the inverse Radon transform was computed using the filtered back projection technique to yield a ROI with enhanced linear structures and minimum clutter. Fig. 1(a) and (b) shows a ROI extracted from a mammogram depicting a spiculated mass, and the Radon enhanced ROI. It is evident from Fig. 1(b) that the Radon enhanced ROI has curvilinear structures enhanced with suppression of clutter due to overlapping out-of-plane tissue structures.

The dominant orientation θ at each pixel of the Radon enhanced ROI was computed by filtering the Radon enhanced ROI

with a set of steerable quadrature filter pairs comprised of the fourth derivative of a Gaussian and its Hilbert transform [25]. The fourth derivative of a Gaussian and its Hilbert transform provide good angular resolution and can be used for analyzing multiple oriented structures at a single location as described in [25]. The dominant orientation energy response at each pixel of the Radon enhanced ROI was computed as described in [25] and this was used to find the strongest pixel locations in the neighborhood of a suspect spiculated mass location whose dominant orientations are directed towards the central mass region. It is important to note that the steerable filters are applied on the Radon enhanced ROI and not directly in the Radon domain, and the processing is carried out at a single scale.

To test the hypothesis that the estimate of the dominant orientation from the Radon enhanced ROI will be more accurate than from the ROI directly, we simulated linear structures of different widths, with each linear structure oriented at an angle $90°$ and superimposed on real mammographic backgrounds of different densities. The mammographic backgrounds were ROIs of size 256×256 pixels, and were cropped from the same set of 21 mammograms used in the measurement study [24], and from regions of the mammogram that did not contain the mass. Each background had one linear structure of length 72 pixels superimposed on it at an angle of $90°$. The intensity profile along the length of the structure was set to a constant value equal to the maximum intensity of the mammographic background. This ensured that the linear structure was at least as bright as the background. We considered three different widths (5, 6, and 7 pixels), and each of these linear structures was superimposed on all the 21 backgrounds. The backbone of each linear structure was centered close to the middle column of each mammographic background ROI. Fig. 2 illustrates examples of simulated linear structures superimposed on real mammographic backgrounds and the corresponding Radon-enhanced ROIs. The column filter used in the Radon-enhancement was designed to match the width of the linear structure. The dominant orientation was computed from each ROI as well as its Radon enhanced version. The error in the estimation of the dominant orientation was computed at each point along the backbone of the linear structure by comparing the estimated orientation with the ground truth ($90°$) for both the cases. The error was accumulated across all 21 ROIs for each width of the linear structure. Cumulative distribution curves of the dominant orientation error were generated (Fig. 3), and the standard deviation of the error (σ) was computed (Fig. 3). The Wilcoxon sign rank (WSR) [26] test was also performed to assess if the dominant orientation errors computed from the ROIs and their Radon enhanced versions were significantly different for each of the three different widths of the linear structure. The resulting p-values are also summarized in Fig. 3. These results suggest that the estimate of the dominant orientation is more accurate for different widths of the linear structure when performed on the Radon enhanced ROI rather than on the ROIs cropped directly from the mammograms. We acknowledge that our experimental setup is fairly simple, in that we have considered only a single orientation and a single synthetic linear structure, with no intersections or junctions. However, we sought to evaluate the accuracy in the estimation of dominant orientation on mammographic
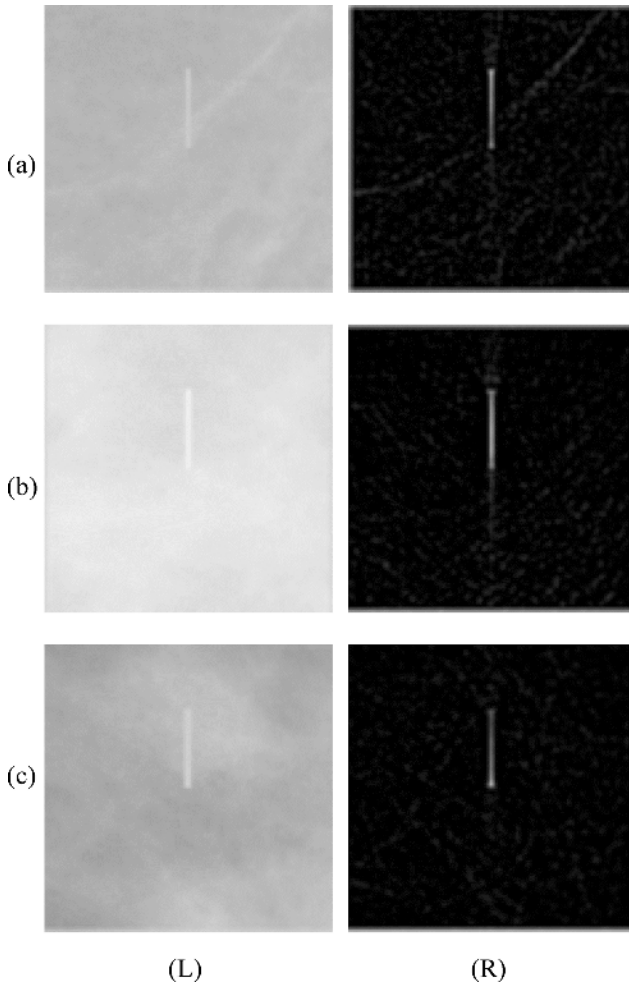
(L)   (R)

Fig. 2. (a) (L) 5 pixel wide (b) (L) 6 pixel wide (c) (L) 7 pixel wide simulated linear structures, all oriented at angle $90°$ and superimposed on real mammographic backgrounds. Also shown are the corresponding Radon enhanced ROIs (R).
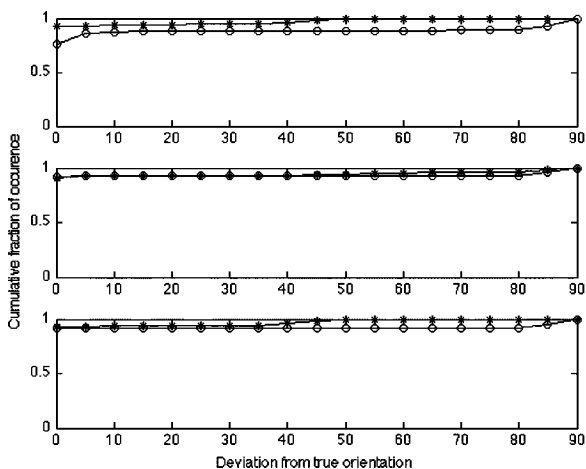

Fig. 3. Cumulative distributions curves of dominant orientation error (o = ROI, ∗ = Radon enhanced ROI). Top row: five pixels wide linear structure (LS), $\sigma = 27.39$ (ROI), $\sigma = 10.60$ (Radon enhanced ROI), WSR p-value <0.05; Middle row: six pixels wide LS, $\sigma = 23.21$ (ROI), $\sigma = 19.17$ (Radon enhanced ROI), WSR p-value <0.05; and Bottom row: seven pixels wide LS, $\sigma = 24.18$ (ROI), $\sigma = 11.12$ (Radon enhanced ROI), WSR p-value <0.05.

backgrounds in the simplest possible scenario to help us decide whether to estimate dominant orientations from the ROIs

directly or their Radon enhanced versions in the more complex scenario of real data. In Fig. 9(a) and (b), we also show the results of the overall *snakules* algorithm on an ROI when the dominant orientation was estimated from the Radon enhanced ROI and the ROI cropped directly from the mammogram, respectively. These results also suggest that the quality of the annotations is better when the dominant orientation is estimated from the Radon enhanced ROI rather than from the ROI directly.

The set C of candidate *snakule* points could contain outliers in the sense that there are points that are farther away from the suspect spiculated mass location than the other points are. This could arise due to the choice of the heuristic for determining the size of the neighborhood N. To eliminate such outliers, we use a popular distance-based outlier detection technique described in [27]. According to the definition of a distance-based outlier, an object O in a dataset T is an $DB(p, D)$ outlier if at least fraction p of the other objects in the data set lie greater than a distance D from the object O. This is a simple, yet powerful technique [27] to detect outliers in a dataset. In our implementation, D represents the 2-D Euclidean distance of a candidate snakule point from the remaining candidate points and p is set to a value of 50%. In other words, we required at least half the number of remaining candidate snakule points to lie at a distance greater than D for any given point to be considered an outlier.

C. Parameter Design

The choice of the parameter values used in the detection of the candidate *snakule* points, where possible, is based on prior measurements of physical properties of spiculated masses. These measurements were collected from experienced radiologists on 21 ROIs scanned using the LUMISYS scanner and our previous study had shown that these measurements could be reliably carried out with a good degree of inter-observer agreement [24].

The design of the linear column filter used to create the Radon enhanced ROI was based on the average spicule width. An analysis of the distribution of the spicule width (in pixels) measured by two experienced radiologists (GJW and TWS) [24] revealed a lognormal distribution for spicule width as illustrated in Fig. 4. We used the measurements made by the more senior radiologist (GJW) for choosing the parameter values. The average spicule width was found to be 968.42 ± 615.28 μm from the spicule width distribution of the more senior radiologist [Fig. 4(a)]. This value of average spicule width translated to five pixels for sub-sampled ROIs extracted from mammograms scanned using LUMISYS and six pixels for sub-sampled ROIs extracted from mammograms scanned using HOWTEK. We chose to set the parameter value for the average spicule width to be equal to six pixels since the *snakules* algorithm was to be deployed on cases from both the scanners. Consequently, we used a column filter $[-3, 1, 1, 1, 1, 1, 1, -1.5, -1.5]^T$ to enhance local peaks that are comprised of six points in the Radon domain. It is important to note that since the filter coefficients sum to 0, the response of the filter to slowly varying values in the Radon domain will be close to 0 or very small.

The choice of the parameter R, the radius of the circular disk as defined in (1), was based on physical measurements of the major axis of 21 spiculated masses. The median radius of a spiculated mass was found to be 6.9 ± 3.98 mm from the measure-
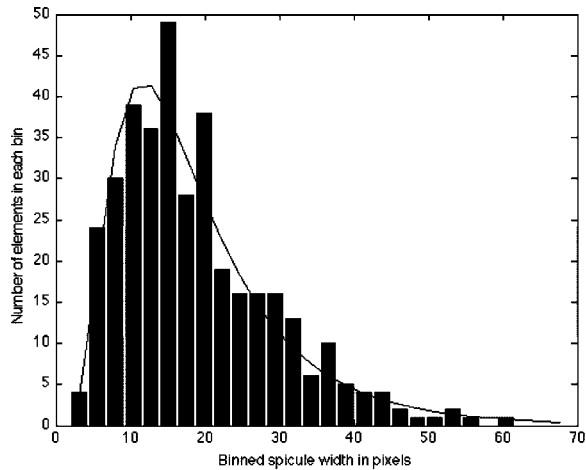
(a)
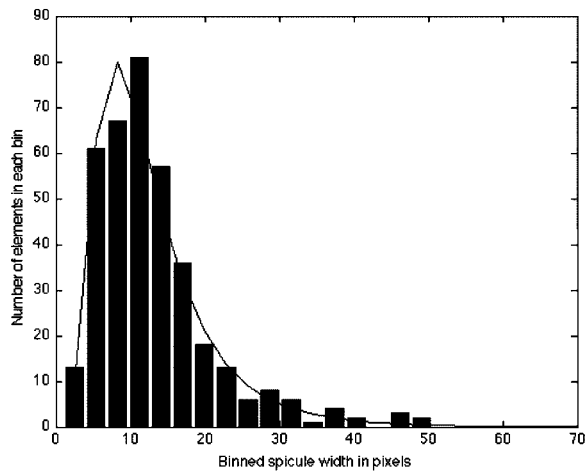
(b)

Fig. 4. (a) Spicule width distribution (GJW), average spicule width $= 968.42 \pm 615.28$ μm, mean squared error (MSE) of lognormal fit $= 0.65$. (b) Spicule width distribution (TWS), average spicule width $= 641.25 \pm 392.69$ μm, MSE of lognormal fit $= 0.29$.

ments of the major axis performed by the more senior radiologist. This value of median radius translated to 35 pixels for sub-sampled ROIs extracted from mammograms scanned using LUMISYS and 40 pixels for sub-sampled ROIs extracted from mammograms scanned using HOWTEK. We chose to set the parameter R to be equal to 40 pixels since the *snakules* algorithm was to be deployed on cases from both the scanners.

The space around the suspect spiculated mass location (x_c, y_c) as defined in (1) was divided into 24 orientation bins with a bin width of $15°$. From the measurements of the physical properties of spiculated masses, we found the average number of spicules to be equal to 17.57 ± 6.01. Our choice for the number of orientation bins (24) was based on this finding from the measurements of physical properties of spiculated masses. Finally, the neighborhood N in (1) was defined as a circular torus whose inner radius was set to a value of 31 pixels and the outer radius was set to a value of 99 pixels. These numbers, though somewhat arbitrary, were set to account for the variation in the mass sizes of different spiculated masses and the varying lengths of their spicules. Fig. 5 illustrates the detection of candidate *snakule* points with the parameters set as described in this section.
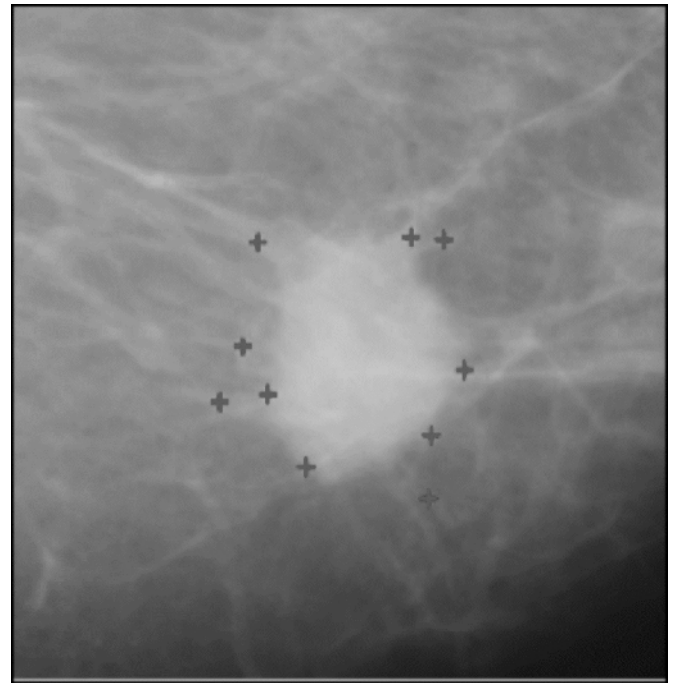

Fig. 5. Detection of candidate *snakule* points (each candidate point identified on the ROI is denoted by a "+").

D. Snakule Evolution and Growth

Once the candidate *snakule* points are identified, we deploy open-ended, parametric snakes, originating from these points. A parametric snake or active contour [16] is a parametric curve $v(s) = [x(s), y(s)]^T, s \in [0, 1]$ that evolves through the image to minimize the energy functional

$$E_{\text{total}} = \int_0^1 \left[\frac{1}{2} (\alpha |v'(s)^2| + \beta |v''(s)|^2) + E_{\text{ext}}(v(s)) \right] ds \tag{3}$$

where $v'(s)$ and $v''(s)$ are the first and second derivatives of $v(s)$, representing continuity and curvature of the contour respectively. The weighting parameters α and β represent the relative importance of the continuity and curvature of the contour. The external energy, $E_{\text{ext}}(v(s))$, typically arises from the image and draws the contour towards features of interest such as edges.

A snake that minimizes the net energy has to satisfy the Euler–Lagrange equation $\alpha v'' - \beta v'''' - \nabla E_{\text{ext}}(v) = 0$, which can be expressed as the force-balanced equation $F_{\text{int}} + F_{\text{ext}} = 0$, where $F_{\text{int}} = \alpha v'' - \beta v''''$ is the internal force that controls the contour's continuity and curvature, and $F_{\text{ext}} = -\nabla E_{\text{ext}}(v)$ is the external force arising from the image that draws the contours towards edges or lines. Many groups have focused on developing new formulations for the external force F_{ext} (e.g., [28], [29]) and these formulations are primarily geared at improving the capture range of the contour and reducing the impact of image noise on the contour evolution. We use one such formulation for the external force called *vector field convolution (VFC)* for the evolution of *snakules* [29]. The VFC force is computed as the convolution of a user-defined vector field kernel with a feature map (e.g., edge map) generated from the image [29]. The vector field kernel is defined such that all

vectors in the field point towards the kernel origin [29]. Hence, when the kernel origin is a feature of interest such as an object boundary, then all the vectors in the vector field point towards the object boundary, thereby causing the evolving contour to move towards the object boundary. A significant advantage in using the VFC force as opposed to standard formulations of external forces or more sophisticated formulations such as the gradient vector flow field (GVF) [28] is that the VFC force is robust to spurious edges and noise in the image and provides a large capture range. This is particularly important when evolving snakes on images such as mammograms, due to the large amount of clutter present in these images. Additionally, it is possible to show that the standard edge-based external force and the GVF force can be expressed as special cases of the VFC force [29].

Instead of using a standard edge map as a feature map, we use the *Radon enhanced ROI* as the feature map. The Radon enhanced ROI has the nice property that the linear structures are enhanced and clutter from spurious edges in the mammogram is reduced, which would otherwise manifest on a traditional edge map. Further, the Radon enhanced ROI has the property that it is nonnegative and has a larger value near the enhanced curvilinear structures [Fig. 1(b)], and hence these curvilinear structures contribute more to the VFC force than the homogeneous regions of the ROI. This property will cause the VFC external force to attract the evolving contour close to the curvilinear structures of interest. The VFC force replaces the standard external force $F_{\text{ext}} = -\nabla E_{\text{ext}}(v)$ in the force balance equation $\alpha v'' - \beta v'''' + f_{\text{VFC}}(v) = 0$. The snake update equations remain the same and are as given in [29].

We adopted an approach that lets the *snakules* grow and subsequently evolve to trace the spicules. This was motivated by the fact that the true length of a spicule is not known beforehand. Dropping snakes much longer than the actual spicule brings about instability as the snakes evolve towards structures that are not spicules. The idea of growing snakes was first described by Berger [30] as a robust alternative to dropping long snakes to trace open-ended curvilinear structures in images and we adopted a similar strategy. The initial *snakule* segment is deployed as a short straight line of $\text{length} \approx 10$ pixels at the detected candidate *snakule* point. This choice of *snakule* segment length was arbitrary, and the only consideration was to prevent instability in the snake evolution process by keeping the segment relatively short. The initial orientation of this straight line is the same as the dominant pixel orientation θ computed from steerable filter analysis. This short *snakule* segment then deforms towards the enhanced spicule under the influence of the VFC force. Once the deformation is complete, the deformed curve is extended in the tangent direction by introducing another short snakule segment, which is a straight line of length ≈ 10 pixels. The new segment then deforms and the iterative process of *snakule* growth and deformation continues until a stopping criterion is met. We used a curvature-based stopping criterion, in which the growth of a *snakule* was stopped at a point where the curvature of the snake exceeded a $30°$ limit. We used a measure of curvature described in [31] that is bounded, with values in the interval $[0, \pi]$. The $30°$ curvature limit was based on excitatory-inhibitory models of human perception of

contour grouping [32]. The values of α and β in (3) have been set to 0 and 1, respectively. A zero value of α at a point along the snake could result in a discontinuity at that point. However, since each *snakule* segment that deforms is a relatively short segment (length ≈ 10 pixels), discontinuities are rare. A value of 1 for β places an emphasis on the segment being free of corners, which is a desirable property for tracking curvilinear segments. Fig. 6 illustrates three iterations of a growing *snakule*. It is important to note that our strategy of growing *snakules* is similar to how a radiologist would annotate a spicule on a mammogram using an electronic interface such as a stylus or a mouse. Finally, once the evolution and the growth of a snakule has stopped, we perform a simple postprocessing operation to check if the current snakule has annotated a linear structure that has already been annotated by any of the other snakules. This could happen if two snakules originate from relatively close points, and end up being attracted towards the same linear structure. To prevent this redundancy, at every point on the current snakule, we check if there is at least one point in any of the other snakules that lies within a square window of size 3×3 centered at the point on the current snakule. If there exists one such point in the 3×3 window, then the current snakule is discarded. The size of the window is kept deliberately small and we end up only discarding snakules that are very close to any of the other snakules.

E. Evaluation of Snakule Annotations

To evaluate the performance of *snakules* for annotating spicules, we conducted an observer study with three experienced radiologists (GJW, TMH, TWS) who judged the quality of the annotations. Two randomly chosen sets of 10 distinct spiculated masses were used for this study. Each set consisted of ROIs manually centered on a spiculated mass that was cropped from the mammogram, and neither of these sets had been previously used during the development and the initial qualitative validation of the algorithm. The evaluation of the *snakules* algorithm was performed in two phases. In the first phase, the *snakules* algorithm was used to annotate one set of ROIs, while the other set was annotated by an experienced, nonclinical observer (GSM) using the NIH ImageJ freehand annotation tool. The images were not edited prior to annotating using the free hand annotation tool and were annotated sequentially. No magnification was used on any of the images and the annotations were performed in one sitting. The two sets of annotations were then presented to the three radiologists in a randomized order. The three radiologists have a collective experience of more than 40 years in interpreting mammograms. In the second phase, which was conducted six months later, the set of 10 ROIs that were previously annotated using *snakules*, were manually annotated by the nonclinical observer (GSM), and the set of ROIs that were previously manually annotated were annotated using *snakules*. The two sets of ROIs were again presented to the three radiologists in a randomized order. Having the two phases six months apart ensured that the radiologists did not remember the cases, and the annotations. Thus, at the end of the two-phase study, there were 20 ROIs, each of which was annotated using both the methods (*snakules* and manual). The two phases of the study were blinded in that the
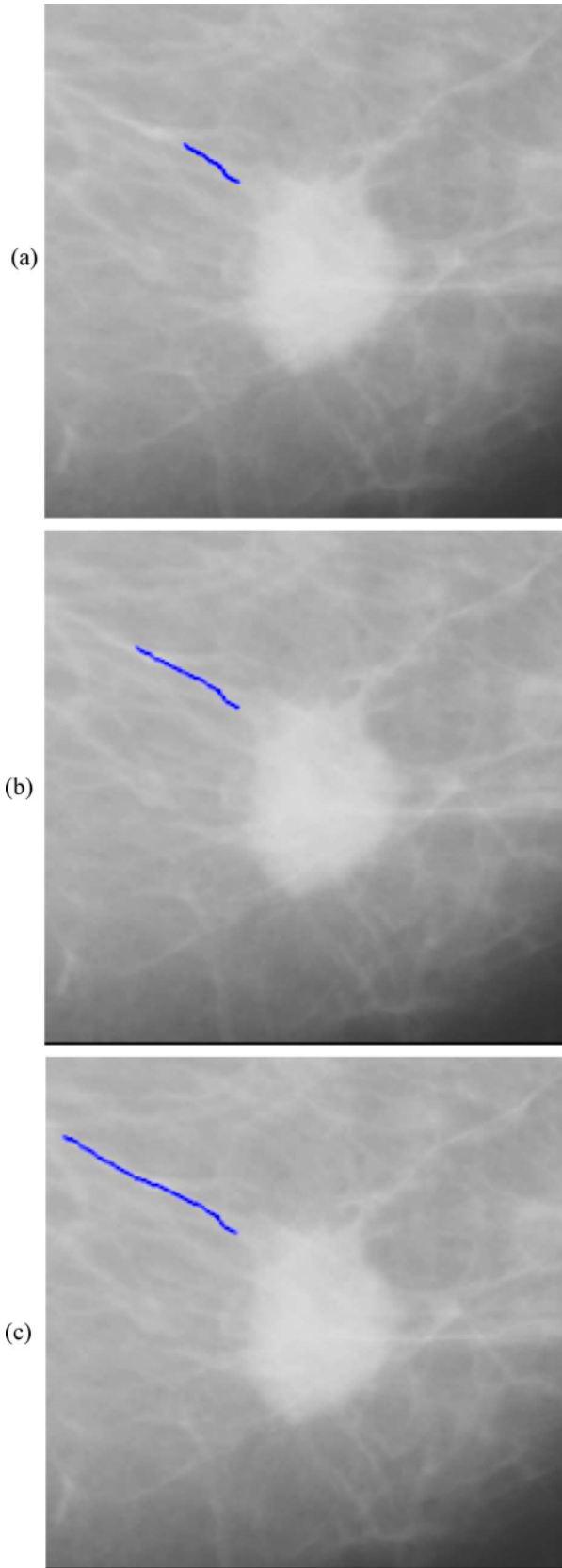
Fig. 6. (a)–(c) Three iterations of a growing *snakule* illustrating the growth of a *snakule* until a stopping criterion is met.

radiologists did not know anything about the techniques used to annotate the ROIs, and the radiologists were not allowed

to discuss with one another. The main goal of the study was to judge if the quality of the annotations performed by the *snakules* method on a set of spiculated masses was equivalent to the quality of annotations performed by an experienced observer.

In each phase, the radiologists were presented with a questionnaire in which they were asked to report the following quantities for each of the 20 annotations that were presented in a randomized order. 1) The number of linear structures that were annotated but did not correspond to actual spicules (i.e., the number of FP annotations) 2) The number of obvious spicules that were missed (i.e., the number of false negatives (FN)) 3) Of the obvious spicules correctly annotated (in other words, the true positives (TP)), the number of spicules that had their lengths correctly identified. The radiologists were specifically instructed to focus on obvious spicules, since it is nearly an impossible task for either a human or a computer vision algorithm to identify all the spicules around a spiculated mass.

For each radiologist, we computed the following measures from the three quantities reported by them:

$$P = \#TP/(\#TP + \#FP)$$
$$R = \#TP/(\#TP + \#FN)$$
$$L_{\text{acc}} = (\#\text{TP whose length was correctly}$$
$$\text{identifed})/(\#\text{ofTP}) \tag{4}$$

where P is the precision of the annotation method defined as the fraction of all the annotations accumulated from the 20 cases that correspond to a true spicule, R is the recall of the annotation method defined as the fraction of obvious spicules correctly identified and annotated across all 20 cases, and L_{acc} is the fraction of correctly annotated spicules whose length was correctly identified across all 20 cases by the annotation method (in other words, the accuracy of correctly identifying the true spicule length). For a good annotation method, all the three measures described above in (4) should be close to 1.

An equivalence test for binomial random variables [33] was performed on each of the three measures described above in (4) to statistically assess the quality of annotations performed by the *snakules* method against the quality of annotations performed by the manual method. An equivalence test was the appropriate choice since we were primarily interested in establishing equivalence of the two methods rather than the difference between the two methods. We made a few assumptions regarding our experimental design in order to perform the equivalence test for binomial random variables. Firstly, for assessing the equivalence of precision score of the two methods, we made an assumption that each annotation in the set of all annotations accumulated from the 20 ROIs was an independent and identically distributed (i.i.d) Bernoulli trial. The outcome of each trial was a true spicule (success) or a false spicule (failure) as judged by the radiologist. Secondly, for assessing the equivalence of recall score of the two methods, we made an assumption that each spicule in the set of all obvious spicules (spicules that were evident to a radiologist) that were accumulated from the 20 ROIs was an i.i.d Bernoulli trial. The outcome of each trial was a successful annotation (success) or a missed annotation (failure). Finally, a similar assumption of a set of i.i.d Bernoulli trials was

made to statistically assess the equivalence of the accuracy of length of true spicules identified by the two methods. We acknowledge that these were simplifying assumptions as the detection of a spicule from a dense mammographic background might not be an i.i.d Bernoulli trial and might depend on factors such as presence of other spicules, and the mammographic background in itself. Given these assumptions, it was possible to express the three performance measures of the two methods as probabilities of success of two binomial random variables and the effectiveness of the *snakules* method was judged by an equivalence test for binomial random variables.

For testing equivalence of the quality of annotations performed by *snakules* against the quality of annotations performed by an experienced observer, the null hypothesis is that the quality of annotations performed by *snakules* is *not* equivalent to the quality of annotations performed by an experienced observer. The alternate hypothesis is that the quality of annotations performed by the two methods is equivalent to one another. The null and alternate hypothesis can be mathematically defined as [33]

$$H_0 : |p_S - p_M| > \Delta$$
$$H_1 : |p_S - p_M| < \Delta \qquad (5)$$

where H_0 and H_1 denote the null and alternate hypotheses, respectively, p_S and p_M represent the population binomial probabilities of success [could be one of the three measures described in (4)] of the *snakules* and manual methods, respectively, Δ is a positive quantity specified by the experimenter, and $|\ |$ is used to denote the absolute difference.

A confidence interval approach was used to decide if the null hypothesis could be rejected [33]. There are many variants of the confidence interval approach and these are based on considerations of prior knowledge of the population binomial probabilities, sample size and statistical power (the probability of finding a statistically significant result when there exists one) [33]. We used the continuity corrected simple asymptotic interval (SAIC) to decide if the null hypothesis of no equivalence has to be rejected or not. According to SAIC, the null hypothesis H_0 is rejected if

$$\hat{p}_S - \hat{p}_M \pm Z_\alpha [\{ \hat{p}_S (1 - \hat{p}_S)/n_S$$
$$+ \hat{p}_M (1 - \hat{p}_M)/n_M + (1/n_S + 1/n_M)/2 \}]^{0.5} \quad (6)$$

is completely contained in the interval $[-\Delta, +\Delta]$ [33]. In (6) $\hat{p}_S, \hat{p}_M$, and n_S, n_M are the sample binomial probabilities of success and the sample sizes of the *snakules* and the manual annotation methods respectively, Z_α is $(1 - \alpha)$ 100th percentile of the standard normal distribution, and α is the nominal significance level. We used the SAIC confidence interval since when there is no prior knowledge of the population probabilities and when the sample sizes are moderate (50—a few hundreds), the SAIC test runs a lesser risk of being anti-conservative, and offers reasonable statistical power. The decision to use SAIC was based on the recommendations made in [33]. In our analysis, the value of α was set to 0.05 and we considered three values for Δ : 0.25, 0.35, and 0.45. The choice of the parameter Δ affects the outcome of the test. For example, if Δ were set to

0.25, then the *snakule* and manual annotation methods are considered equivalent if their probabilities of success were within 0.25 of each other, and not equivalent for a larger difference in the probabilities. Consequently, a smaller value for Δ makes the interval tighter and represents a strict criterion for rejecting the null hypothesis of no equivalence, while a larger value for Δ represents a more lenient criterion for rejecting the null hypothesis. The three values we used for Δ, i.e., 0.25, 0.35, and 0.45 represented a strict, moderately lenient, and lenient criterion, respectively. For assessing inter and intraobserver variability in clinical tasks such as radiographic measurements, a Δ value of 0.25 is considered to be a strict criterion, while a Δ value of 0.30 is considered to be a more lenient criterion, and if the variability is within these values, it is considered acceptable [24]. However, to the best of our knowledge, no prior literature exists on comparing effectiveness of spicule annotation algorithms, and hence we used three values for Δ to represent a strict, moderately lenient, and lenient criterion. Further, our choice of Δ might seem liberal, for instance if $\Delta = 0.45$ and the two methods turned out to be equivalent, then the methods are equivalent even if their probabilities of success are within 0.45 of each other. However, it is important to note that we are comparing the annotations performed by an automated algorithm against those performed by an experienced observer. The manual annotations made by an experienced observer and evaluated by three radiologists can be treated as the *gold standard* and they provide a high bar for assessing the effectiveness of an automated algorithm.

III. Results

Fig. 7 and Fig. 9(a) illustrate examples of ROIs annotated using *snakules*, while Fig. 8 illustrates examples of ROIs annotated manually. Qualitatively speaking, the *snakule* annotations look encouraging on the ROIs illustrated in Fig. 7, and Fig. 9(a). The *snakule* annotations also suggest that the method is fairly robust to variation in the size of the central mass. Fig. 10 illustrates the variation in annotations with the initial location (x_c, y_c) as defined in (1). The choice of initial location does affect the quality of annotations by generating some false positives, but as evident in Fig. 10, the prominent spicules are usually annotated.

Table II summarizes the three performance measures described in (4) and the results of the statistical tests for equivalence. The following conclusions can be made from Table II: 1) Firstly, when an obvious spicule is present on the mammogram, the *snakules* method annotates the spicule with a high probability of success as evident from the recall scores (maximum recall$= 0.84$). This was consistent across the evaluations by all the three radiologists. Further, the recall scores of the *snakule* and manual annotations were found to be statistically equivalent at a strict criterion ($\Delta = 0.25$) for all the three radiologists, which was encouraging. 2) Secondly, as evident from Table II, the *snakules* method annotates spicules with a reasonably good precision (maximum precision$= 0.76$) and this was again consistent across the evaluations by all the three radiologists. When compared to the recall, the precision of the *snakules* method is lower, suggesting that the method annotates a few curvilinear structures such as blood vessels that are directed towards the central mass region. Fig. 11 illustrates
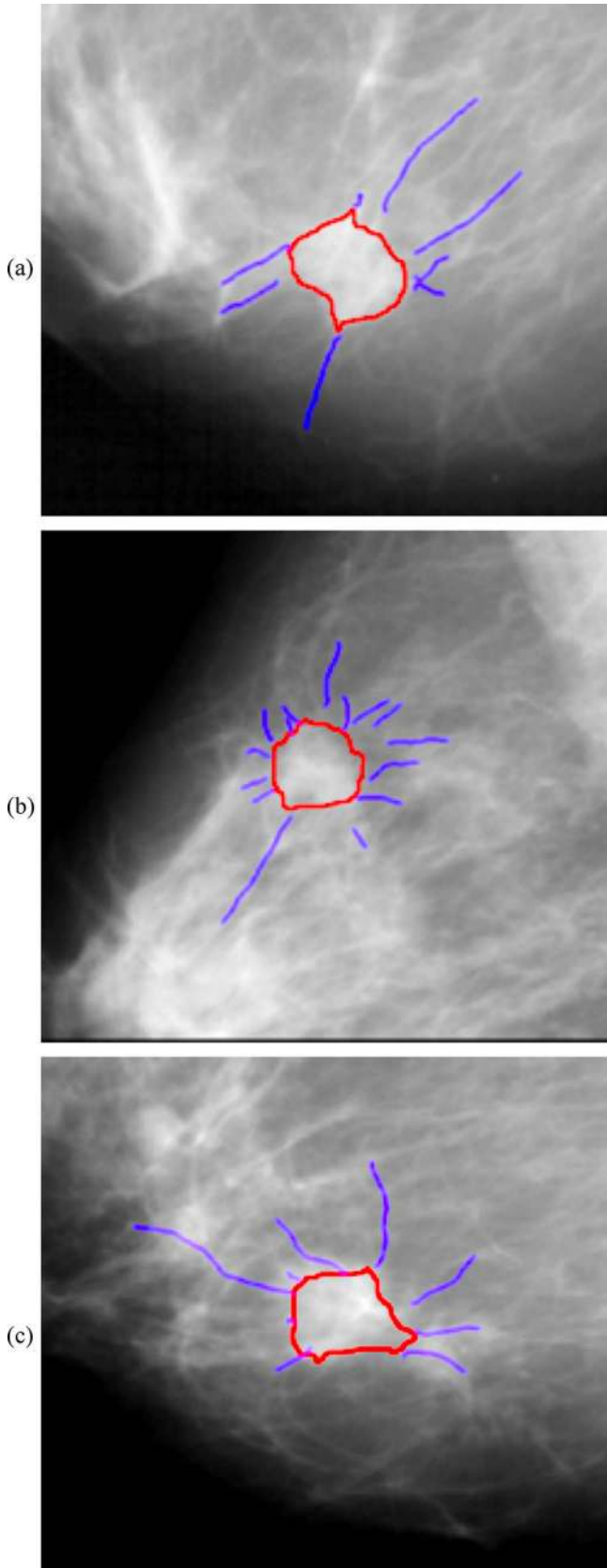
Fig. 7. (a)–(c) Examples of ROIs annotated using *snakules* (central mass annotated manually by GSM).
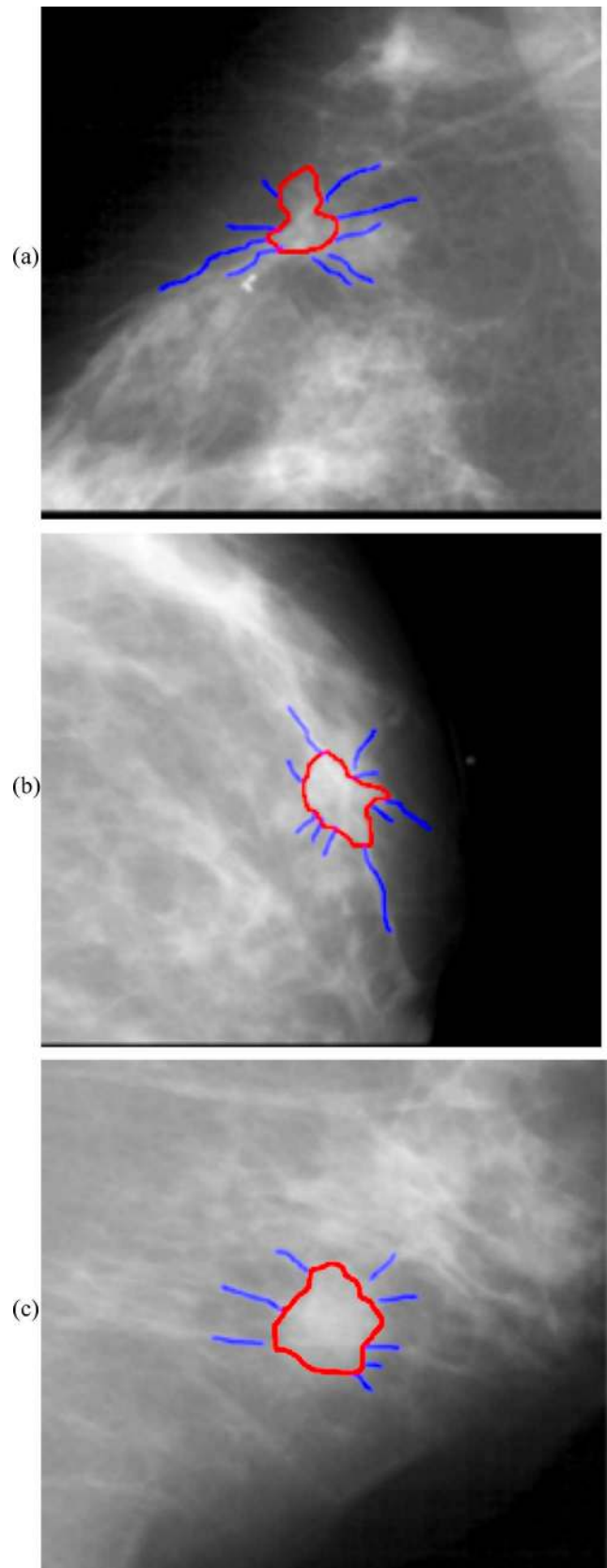
Fig. 8. (a)–(c) Examples of ROIs annotated manually by GSM.

an example of one such ROI in which the radiologists reckoned that blood vessels had been annotated. The precision scores of

the *snakules* and manual methods were found to be statistically equivalent at a lenient criterion ($\Delta = 0.38, \Delta = 0.45$ is
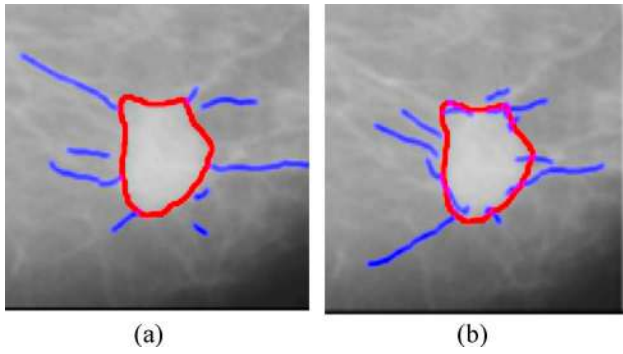
Fig. 9. (a) *Snakule* annotations when dominant orientation is computed from Radon enhanced ROI (b) *Snakule* annotations when dominant orientation is computed from the ROI cropped directly from the mammogram. Central mass was annotated manually by GSM.
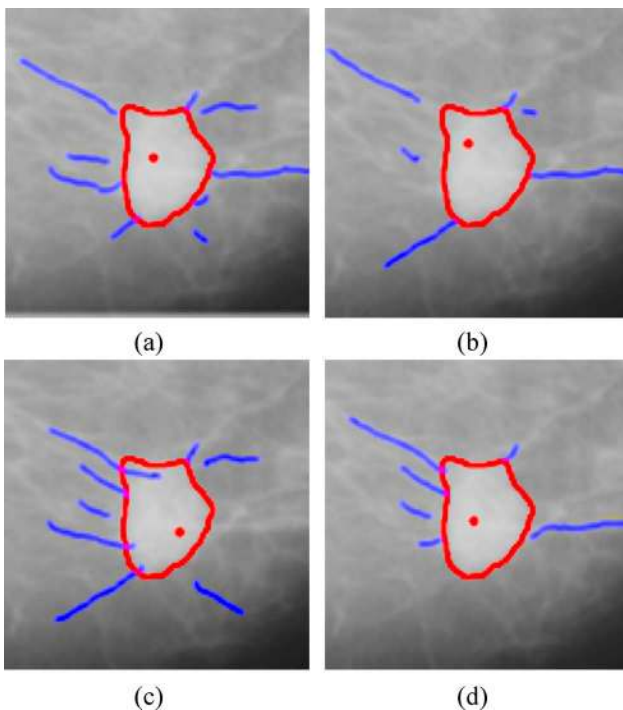

Fig. 10. (a)–(d) Variation in *snakule* annotations with the initial location indicated by a "+" (central mass annotated manually by GSM).
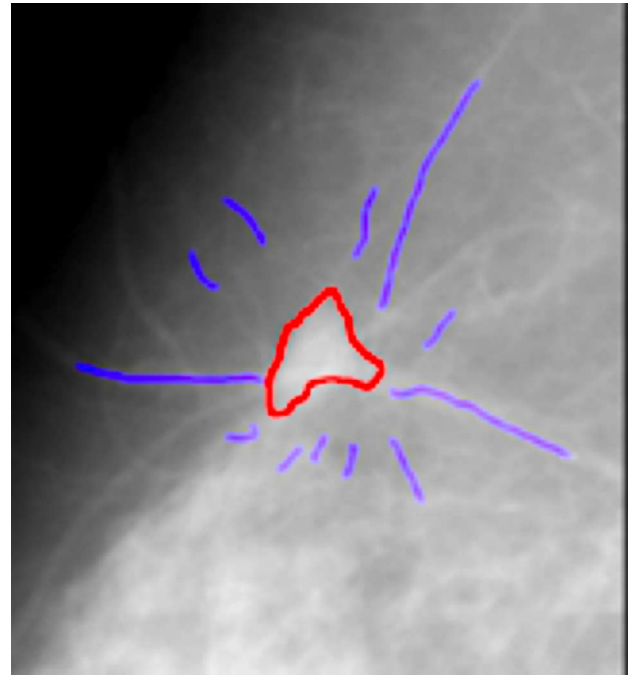

Fig. 11. Example of an ROI in which the *snakules* method has annotated quite a few curvilinear structures that are non-spicules but directed towards the central mass.

reported in Table II) for all the three radiologists. 3) Finally, the maximum accuracy of correctly identifying the true spicule length was 0.95 for the *snakules* method. However, the length accuracy value was most variable across the three radiologists (0.50, 0.79, and 0.95). Further, for one radiologist (TWS) the length accuracy scores of the *snakules* and manual methods were found to be statistically equivalent at a strict criterion ($\Delta = 0.25$), while for the other radiologist (TMH) the length accuracy scores were found to be statistically equivalent at $\Delta = 0.26$ ($\Delta = 0.35$ is reported in Table II). However, for the third radiologist (GJW), the length accuracy scores of the two methods were not found to be statistically equivalent at any of the three chosen Δ values. This suggests that while the radiologists are fairly consistent in identifying the obvious spicules and the false positive annotations, they are not very consistent in identifying the true length of the spicule. However, for two

out of the three radiologists (TWS and TH), the length accuracy scores of the *snakules* and manual methods were found to be statistically equivalent at $\Delta = 0.26$, which was encouraging from the point of view of extracting meaningful physical properties of spicules such as length that can subsequently be used as a feature in the classification of linear structures evident on a mammogram.

IV. CONCLUSION

In this paper, we have presented a novel algorithm termed *snakules* for the annotation of spicules on mammography. The algorithm is model-based in that its design is guided by statistics of physical measurements of real spiculated masses on mammography. Certain generic aspects of the design such as the use of Radon transform to enhance linear structures and compute dominant orientation, and the VFC force field computed from the Radon enhanced ROI for active contour deformation could be useful for other medical image segmentations tasks. We foresee *snakules* being useful for a variety of tasks. Firstly, the results of our observer study demonstrate the reliability of *snakules* in annotating obvious spicules, and also suggest that *snakules* can be used for extracting features (e.g., length) specific to spicules, which can be subsequently used to distinguish true spiculated mass locations from nonlesion locations on a mammogram and improve the specificity of CADe systems. We have conducted a preliminary study that explores the potential of *snakules* for classifying candidate spiculated mass locations on mammography. The results from this study are promising [34] and we plan to carry out further analysis of *snakules* for the task of CADe FP reduction. Secondly, *snakules* can be used as a CADe prompting tool that explicitly marks spicules on a mammogram. Lastly, *snakules* can be used in conjunction with

TABLE II
PRECISION, RECALL, AND SPICULE LENGTH ACCURACY SCORES COMPUTED FROM RADIOLOGISTS' EVALUATIONS OF *SNAKULES* AND MANUAL ANNOTATIONS. ALSO SHOWN ARE THE CONCLUSIONS OF THE STATISTICAL TEST FOR EQUIVALENCE FOR ALL THREE MEASURES

Performance Measure	Radiologist	Sample binomial probabilities of success		Sample Size		Null hypothesis (no equivalence) rejected?		
		Snakules $\hat{p}_S$	Manual $\hat{p}_M$	*Snakules* n_S	Manual n_M	$\Delta = 0.25$	$\Delta = 0.35$	$\Delta = 0.45$
Precision	GJW	0.69	0.90	187	163	No	No	Yes
	TWS	0.76	0.97	187	163	No	No	Yes
	TMH	0.71	0.90	187	163	No	No	Yes
Recall	GJW	0.81	0.87	160	169	Yes	Yes	Yes
	TWS	0.84	0.88	169	180	Yes	Yes	Yes
	TMH	0.80	0.83	166	176	Yes	Yes	Yes
Accuracy of identifying the true spicule length	GJW	0.50	0.84	129	147	No	No	No
	TWS	0.95	0.98	142	158	Yes	Yes	Yes
	TMH	0.79	0.86	133	146	No	Yes	Yes

online annotation tools (e.g., [35]) to annotate spiculated masses on mammograms that are used for training radiology personnel (such as residents).

One aspect of the *snakules* method that needs improvement is the annotation of curvilinear structures that are non-spicules but are directed towards the central mass, as evident from the precision scores in Table II. One possible approach that could be adopted to address this issue is to use multiscale processing by varying the radius of the circular disk R in (1) during the detection of candidate *snakule* points. Another possible approach would be to annotate the central mass and restrict the detection of candidate snakule points to the annotated boundary of the central mass. However, as pointed out by Zwiggelaar *et al.* in [13], not all spiculated masses have a well-defined central mass boundary, which makes it hard to accurately estimate the boundary of the central mass. Alternate approaches could also include probabilistic inference of candidate snakule points using multiple initialization locations, and the use of multiscale curvilinear structure detection strategies (e.g., [36]) with prior information of spicule properties to filter false spicule candidate points.

REFERENCES

[1] L. Liberman, A. F. Abramson, F. B. Squires, J. R. Glassman, E. A. Morris, and D. D. Dershaw, "The breast imaging reporting and data system: Positive predictive value of mammographic features and final assessment categories," *Am. J. Roentgenol.*, vol. 171, pp. 35–40, 1998.

[2] M. P. Sampat, M. K. Markey, and A. C. Bovik, "Computer-aided detection and diagnosis in mammography," in *Handbook of Image and Video Processing*, A. C. Bovik, Ed., 2nd ed. New York: Academic, 2005, pp. 1195–1217.

[3] R. M. Nishikawa, "Current status and future directions of computer-aided diagnosis in mammography," *Computerized Med. Imag. Graphics*, vol. 31, pp. 224–235, 2007.

[4] R. A. Castellino, "Computer aided detection (CAD): An overview," *Cancer Imag.*, vol. 5, pp. 17–9, Aug. 23, 2005.

[5] S. K. Yang *et al.*, "Screening mammography–detected cancers: Sensitivity of a computer-aided detection system applied to full-field digital mammograms," *Radiology*, vol. 244, pp. 104–111, 2007.

[6] T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12 860 patients in a community breast center," *Radiology*, vol. 220, pp. 781–6, 2001.

[7] L. W. Burhenne *et al.*, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology*, vol. 215, pp. 554–62, 2000.

[8] M. P. Sampat, A. C. Bovik, G. J. Whitman, and M. K. Markey, "A model-based framework for the detection of spiculated masses on mammography," *Med. Phys.*, vol. 35, pp. 2110–2123, 2008.

[9] N. Karssemeijer and G. M. te Brake, "Detection of stellate distortions in mammograms," *IEEE Trans. Med. Imag.*, vol. 15, pp. 611–619, 5, Oct. 1996.

[10] L. Li, F. Mao, W. Qian, and L. P. Clarke, "Wavelet transform for directional feature extraction in medical imaging," in *Proc. Int. Conf. Image Process.*, 1997, vol. 3, pp. 500–503.

[11] S. Liu, C. F. Babbs, and E. J. Delp, "Multiresolution detection of spiculated lesions in digital mammograms," *IEEE Trans. Image Process.*, vol. 10, no. 6, pp. 874–884, Jun. 2001.

[12] R. M. Rangayyan and F. J. Ayres, "Gabor filters and phase portraits for the detection of architectural distortion in mammograms," *Med. Biol. Eng. Comput.*, vol. 44, pp. 883–894, 2006.

[13] R. Zwiggelaar *et al.*, "Model-based detection of spiculated lesions in mammograms," *Med. Image Anal.*, vol. 3, pp. 39–62, 1999.

[14] J. A. Baker *et al.*, "Computer-aided detection (CAD) in screening mammography: Sensitivity of commercial CAD systems for detecting architectural distortion," *Am. J. Roentgenol.*, vol. 181, pp. 1083–1088, 2003.

[15] S. M. Astley, "Evaluation of computer-aided detection (CAD) prompting techniques for mammography," *Br. J. Radiol.*, vol. 78, pp. S20–S25, 2005.

[16] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, pp. 321–331, 1987.

[17] R. Zwiggelaar, S. M. Astley, C. R. M. Boggis, and C. J. Taylor, "Linear structures in mammographic images: Detection and classification," *IEEE Trans. Med. Imag.*, vol. 23, no. 9, p. 1077, Sep. 2004.

[18] W. Qian, L. Li, L. Clarke, R. A. Clark, and J. Thomas, "Digital mammography: Comparison of adaptive and nonadaptive CAD methods for mass detection," *Acad. Radiol.*, vol. 6, pp. 471–80, 1999.

[19] T. Nakagawa, T. Hara, H. Fujita, K. Horita, T. Iwase, and T. Endo, "Radial-searching contour extraction method based on a modified active contour model for mammographic masses," *Radiol. Phys. Technol.*, vol. 1, pp. 151–161, 2008.

[20] J. Shi *et al.*, "Characterization of mammographic masses based on level set segmentation with new image features and patient information," *Med. Phys.*, vol. 35, pp. 280–90, 2008.

[21] Y. Yuan, M. L. Giger, H. Li, K. Suzuki, and C. Sennett, "A dual-stage method for lesion segmentation on digital mammograms," *Med. Phys.*, vol. 34, pp. 4180–93, 2007.

[22] M. Heath, K. W. Bowyer, and D. Kopans, "Current status of the digital database for screening mammography," in *Digital Mammography*. New York: Academic, 1998, pp. 457–460.

[23] M. Heath, K. W. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, Jr., "The digital database for screening mammography," in *Proc. 5th Int. Workshop Digital Mammogr.*, Toronto, ON, Canada, 2001, pp. 212–218.

[24] M. P. Sampat *et al.*, "The reliability of measuring physical characteristics of spiculated masses on mammography," *Br. J. Radiol.*, vol. 79, pp. S134–S140, 2006.

[25] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.

[26] B. Rosner, *Fundamentals of Biostatistics*, 6th ed. Belmont, CA: Duxbury, 2005.

[27] E. M. Knorr and R. T. Ng, "Finding intensional knowledge of distance-based outliers," in *Proceedings, 25th International Conference on Very Large Data Bases*, Edinburgh, U.K., 1999, pp. 211–222.

[28] X. Chenyang and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 359–369, Mar. 1998.

[29] B. Li and S. T. Acton, "Active contour external force using vector field convolution for image segmentation," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2096–2106, Aug. 2007.

[30] M. O. Berger, "Snake growing," in *Proc. 1st Eur. Conf. Comput. Vis.*, Antibes, France, 1990, pp. 570–572.

[31] D. J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," *CVGIP: Image Understand.*, vol. 55, pp. 14–26, 1991.

[32] S. C. Yen and L. H. Finkel, "Extraction of perceptually salient contours by striate cortical networks," *Vis. Res.*, vol. 38, pp. 719–741, 1998.

[33] L. Barker, H. Rolka, D. Rolka, and C. Brown, "Equivalence testing for binomial random variables: Which test to use?," *Am. Stat.*, vol. 55, pp. 279–287, 2001.

[34] G. S. Muralidhar, M. K. Markay, and A. C. Bovik, "Snakules for automatic classification of candidate spiculated mass locations on mammography," presented at the 2010 IEEE Southwest Symposium on Image Analysis and Interpretation, Austin, TX, 2010.

[35] Y. Zheng, M. Wu, E. Cole, and E. D. Pisano, "Online annotation tool for digital mammography," *Acad. Radiol.*, vol. 11, pp. 566–572, 2004.

[36] S. Berlemont and J. C. Olivo-Marin, "Combining local filtering and multiscale analysis for edge, ridge, and curvilinear objects detection," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 74–84, Jan. 2010.