SNAKULES: SNAKES THAT SEEK SPICULES ON MAMMOGRAPHY

Gautam S. Muralidhar[1], Alan C. Bovik[2], Mia K. Markey[1]

[1]The University of Texas Department of Biomedical Engineering, Austin, 78712 USA.
[2]Department of Electrical and Computer Engineering, The University of Texas at Austin, 78712 USA.

ABSTRACT

We present a new method called "*snakules*" for the annotation of spicules on mammography. *Snakules* employs parametric open-ended snakes that are deployed in a region around a suspect spiculated mass location that has been identified by a radiologist or a computer-aided detection (CADe) algorithm. The set of convergent *snakules* 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. Our results from an initial observer study involving an experienced radiologist demonstrate the strong potential of the method as an image analysis technique to improve the specificity of CADe algorithms.

***Index Terms*—** Snakules, computer-aided detection, mammography, snakes, active contours.

1. INTRODUCTION

Breast cancer manifests as various findings on mammography – microcalcifications, masses (spiculated and non-spiculated) 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 non-spiculated 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]. 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.

Studies have shown that the detection performance of most CADe systems on spiculated masses is not optimal (e.g., [3]). One of the major reasons for the poor performance is that most CADe systems for spiculated masses are developed around the central theme of detecting radial patterns of converging lines (e.g., [4]-[5]) 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. Hence, most CADe systems for spiculated masses suffer from a high FP rate.

The performance of CADe systems on spiculated masses can be improved by developing sophisticated image analysis techniques to extract physical properties specific to spicules. These physical properties can then be used in the reliable classification of suspect locations identified on the image as cancerous or non-cancerous. Towards this goal, we have developed "*snakules*" – an algorithm that employs parametric open-ended snakes (active contours) [6] to annotate spicules on mammography. The set of convergent *snakules* are automatically initialized at a set of candidate points on the image from where the spicules most likely originate. *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. The parameters of the algorithm are based on actual physical-measurements of spiculated masses that have been collected on mammograms.

It is important to note that other groups have tried to extract properties specific to spicules to improve specificity of CADe systems (e.g., [7]); however our approach is the first attempt to explicitly capture spicules in a bid to use them to distinguish spiculated masses from other breast structures.

2. PROPOSED ALGORITHM

2.1. Detecting candidate snakule points

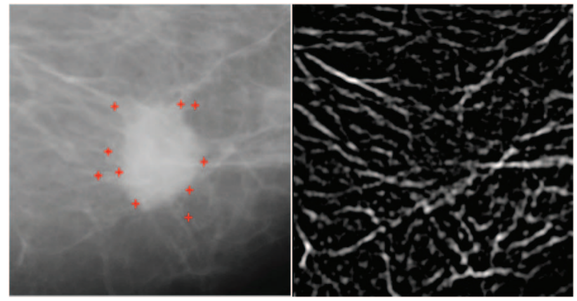
The first step in the *snakules* algorithm is to detect a set of points on the image from where spicules originate in the region around the suspect spiculated mass location identified by either a CADe algorithm or a radiologist. These candidate points represent locations on the image where *snakules*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

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

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), $\theta_{x,y}$ is the dominant pixel orientation at location (x, y), $\psi_{x,y}$ is the direction of the location (x, y) with respect to (x_c, y_c) and is computed as $\psi_{x,y} = \tan^{-1}\left(\left(y_c - y\right)/\left(x_c - x\right)\right)$, R is the radius of a circular disk centered on (x_c, y_c) and towards which the pixel at location (x, y) is directed, $r_{x,y}$ is the Euclidean distance between (x, y) and (x_c, y_c), and k represents the k^{th} orientation bin.

The condition $\left|\theta_{x,y} - \psi_{x,y}\right| < R/r_{x,y}$ is the same as defined in [5], 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* regions of interest (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 3D breast structures onto a 2D image plane. Radon enhancement of spiculated lesions on mammograms is explained in detail in [4]. 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 [8]. The choice of the parameters R and N, and the number of orientations bins considered around the suspect spiculated mass location are



(a) (b)

Fig. 1: (a) A ROI from a mammogram depicting a spiculated mass and candidate *snakule* points (denoted by a '+') (b) Radon enhanced ROI depicting enhanced curvilinear structures.

based on the average mass radius, range of spicule length, and the number of spicules respectively that have been computed from measurements of spiculated masses collected on mammograms [9]. Fig. 1(a), and 1(b) show a ROI extracted from a mammogram depicting a spiculated mass along with the detected candidate *snakule* points, and the Radon enhanced ROI depicting enhanced curvilinear structures respectively.

2.2. Snakule evolution

Once the candidate *snakule* points are identified, we deploy open-ended, parametric snakes, originating from these points. A parametric snake or active contour [6] is a parametric curve $v(s) = \left[x(s), y(s)\right]^T, s \in [0,1]$ that evolves through the image to minimize the energy functional

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

where $v'(s)$ and $v''(s)$ are the first and second derivatives of $v(s)$ representing continuity and curvature of the contour respectively, and α and β are weighting parameters representing the relative importance of the continuity and curvature of the contour. $E_{ext}\left(v(s)\right)$ represents the external energy, typically arising from the image (e.g., gradient map) that draws the contour towards features of interest such as edges.

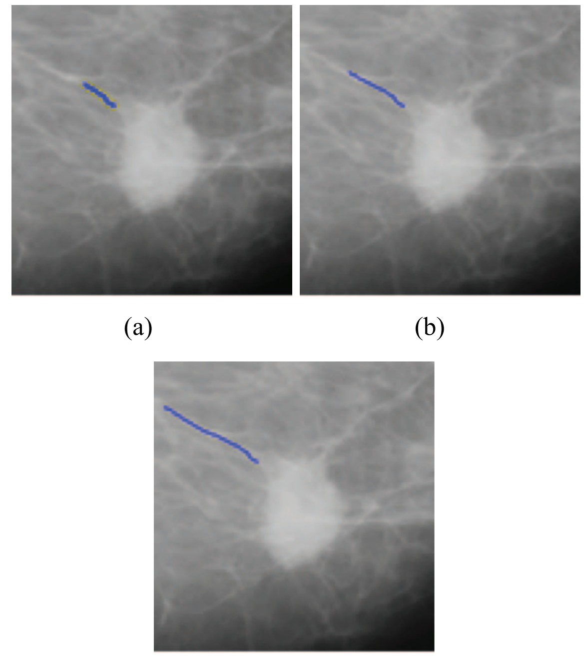
A snake that minimizes the net energy has to satisfy the Euler-Lagrange equation $\alpha v'' - \beta v'''' - \nabla E_{ext}(v) = 0$, which can be expressed as the force-balance equation $F_{\text{int}} + F_{ext} = 0$, where $F_{\text{int}} = \alpha v'' - \beta v''''$ is the internal force that controls the contour's continuity and curvature, and

$F_{ext} = -\nabla E_{ext}(v)$ is the external force arising from the image that draws the contours towards edges or lines. We used the *vector field convolution (VFC)* force for the evolution of *snakules* [10]. 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. A significant advantage in using the VFC force 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.

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 non-negative 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.

2.3. Snakule growth

We adopted an approach in which the *snakules* evolved and grew iteratively until a stopping criterion was met. 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 [11] as a robust alternative to dropping long snakes to trace open-ended curvilinear structures in images and we adopted a similar strategy. In the first iteration, a short *snakule* segment is initialized as a straight line of $length \approx 10$ 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 next iteration begins in which the deformed curve is extended in the tangent direction by introducing another short *snakule* segment, which is a straight line of $length \approx 10$ 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. Fig. 2 illustrates three iterations of a growing *snakule.*



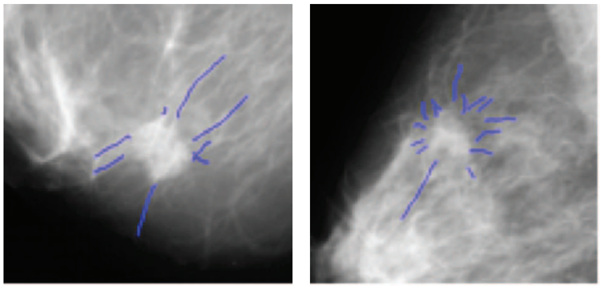
(a) (b)

(c)

Fig. 2: (a) – (c) Three iterations of a growing *snakule* illustrating the growth of a *snakule* until a stopping criterion is met.

3. EVALUATION METHODOLOGY

To evaluate the *snakule* annotation results, we have conducted an initial observer study involving an experienced radiologist. Two randomly chosen sets of 10 distinct spiculated masses obtained from the digital database for screening mammography [12] was 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 *snakules* algorithm was used to annotate one set of ROIs, while the other set was annotated by an experienced, non-clinical observer (GSM) using a freehand annotation tool. The two sets of annotations were then presented to the radiologist in a randomized order. This study was blinded in that the radiologist did not know anything about the techniques used to annotate the ROIs. The radiologist was informed that the objective of the study was to evaluate the quality of two annotation techniques. The main goal of the study was to judge if the quality of the annotations performed by the *snakules* method on a randomly chosen set of spiculated masses was equivalent to the quality of annotations performed by an experienced observer on another randomly chosen set of spiculated masses. The radiologist was asked to report the number of linear structures that were annotated, but did not correspond



(a) (b)

Fig. 3: (a) – (b) Examples of ROIs annotated using *snakules*.

to true spicules, the number of obvious spicules missed and the number of true spicules whose length was also correctly identified. From these quantities, three performance measures were computed: average precision and recall scores, and fraction of true spicules whose length was correctly identified. An equivalence test for binomial random variables using the continuity corrected simple asymptotic interval (SAIC) at a lenient, moderately lenient, and strict criterion (denoted by Δ: 0.45, 0.35, and 0.25 respectively) for rejecting the null hypothesis of *no equivalence* [13] was performed on each of the three measures described above to statistically assess the performance of the *snakules* method.

4. RESULTS AND CONCLUSIONS

Fig. 3 illustrates examples of ROIs annotated using *snakules*. Qualitatively speaking, the *snakule* annotations look very encouraging on these ROIs. The average precision and recall scores, and fraction of true spicules whose length was correctly identified was found to be 0.67, 0.72, and 0.80 respectively for the *snakules* annotations and 0.86, 0.81 and 0.76 respectively for the manual annotations. The results of the equivalence tests are encouraging. Firstly, at a lenient criterion (Δ = 0.45), the quality of *snakules* annotations was found to be statistically equivalent to the quality of manual annotations in terms of all three performance measures described earlier. Secondly, at a moderately lenient criterion (Δ = 0.35), only the precision of *snakules* was found to be non-equivalent. This suggests that if there is an evident spicule on the mammogram, the *snakules* method finds it with a reasonably high probability. Also, at the same criterion, the fraction of true spicules whose length was correctly identified was found to be equivalent. This suggests that the *snakules* method can be reliably used to annotate obvious spicules and extract meaningful physical properties of spicules such as length that can be used as a feature in the classification of linear structures evident on a mammogram. Finally, at a strict criterion (Δ = 0.25), none of the performance measures computed from the evaluation of *snakules* annotations were found to be equivalent to the manual annotations. However, we have to bear in mind that the *snakules* annotations are being compared to manual annotations made by an experienced observer, which presents a very high bar to judge the quantitative results of an automated algorithm.

To conclude, we have developed a new algorithm called "*snakules*" for the annotations of spicules on mammography. We foresee *snakules* being a useful device for extracting valuable features specific to spicules and as a tool that helps radiologists in visualizing spiculated masses on mammograms.

5. 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," *American Journal of Roentgenology,* 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*, 2nd ed, A. C. Bovik, Ed.: Academic Press, 2005, pp. 1195-1217.

[3] S. K. Yang, W. K. Moon, N. Cho, J. S. Park, J. H. Cha, S. M. Kim 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.

[4] M. P. Sampat, A.C. Bovik, G.J. Whitman, M.K. Markey, "A Model-based Framework for the Detection of Spiculated Masses on Mammography," *Medical Physics*, vol. 35, pp. 2110-2123, 2008.

[5] N. Karssemeijer and G. M. te Brake, "Detection of Stellate Distortions in Mammograms," *IEEE Trans. Med. Imag.*, vol. 15, 1996.

[6] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *International Journal of Computer Vision*, vol. 1, pp. 321-331, 1987.

[7] 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, p. 1077, 2004.

[8] W. T. Freeman and E. H. Adelson, "The Design and Use of Steerable Filters," *IEEE Trans. Pattern Anal. Mach*. Intell., vol. 13, pp. 891-906, 1991.

[9] M. P. Sampat, G. J. Whitman, T. W. Stephens, L. D. Broemeling, N. A. Heger, A. C. Bovik et al., "The Reliability of Measuring Physical Characteristics of Spiculated Masses on Mammography," *British Journal of Radiology*, vol. 79, pp. S134-S140, 2006.

[10] B. Li and S. T. Acton, "Active Contour External Force using Vector Field Convolution for Image Segmentation," *IEEE Trans. Image Proc.*, vol. 16, pp. 2096-2106, 2007.

[11] M. O. Berger, "Snake Growing," in *Proceedings, First European Conference on Computer vision*, Antibes, France, 1990.

[12] M. Heath, K. W. Bowyer, and D. Kopans, "Current Status of the Digital Database for Screening Mammography," in *Digital Mammography*, Academic Publishers, 1998, pp. 457-460.

[13] L. Barker, H. Rolka, D. Rolka, and C. Brown, "Equivalence Testing For Binomial Random Variables: Which Test to Use?" *The American Statistician*, vol. 55, pp. 279-287, 2001.