A superpixel-based framework for automatic tumor segmentation on breast DCE-MRI

Ning Yu[a], Jia Wu[b], Susan P. Weinstein[b], Bilwaj Gaonkar[b], Brad M. Keller[b], Ahmed B. Ashraf[b], YunQing Jiang[a], Christos Davatzikos[b], Emily F. Conant[b], Despina Kontos[b*]

[a]School of Engineering and Applied Science, [b]Dept. of Radiology, University of Pennsylvania
*Contact Author - em: Despina.Kontos@uphs.upenn.edu ph: 215-615-0827 fx: 215-614-0266

ABSTRACT

Accurate and efficient automated tumor segmentation in breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is highly desirable for computer-aided tumor diagnosis. We propose a novel automatic segmentation framework which incorporates mean-shift smoothing, superpixel-wise classification, pixel-wise graph-cuts partitioning, and morphological refinement. A set of 15 breast DCE-MR images, obtained from the American College of Radiology Imaging Network (ACRIN) 6657 I-SPY trial, were manually segmented to generate tumor masks (as ground truth) and breast masks (as regions of interest). Four state-of-the-art segmentation approaches based on diverse models were also utilized for comparison. Based on five standard evaluation metrics for segmentation, the proposed framework consistently outperformed all other approaches. The performance of the proposed framework was: 1) 0.83 for Dice similarity coefficient, 2) 0.96 for pixel-wise accuracy, 3) 0.72 for VOC score, 4) 0.79 mm for mean absolute difference, and 5) 11.71 mm for maximum Hausdorff distance, which surpassed the second best method (i.e., adaptive geodesic transformation), a semi-automatic algorithm depending on precise initialization. Our results suggest promising potential applications of our segmentation framework in assisting analysis of breast carcinomas.

Keywords: Breast DCE-MRI, Tumor Segmentation, Superpixel, Graph-cuts

1. INTRODUCTION

Breast cancer is the most commonly diagnosed form of cancer in women and is the second leading cause of cancer deaths in women in the United States[1]. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a commonly utilized breast imaging technique for diagnosing and staging breast cancer[2-4]. Proper delineation of the tumors is essential for accurate clinical assessment of tumor extent on DCE-MRI. For example, accurate tumor segmentation from longitudinal images of the same patient can efficiently describe the morphological changes of tumors over time, consequently providing assessment of pathologic response to neoadjuvant chemotherapy[5, 6]. However, manual tumor segmentation on a large volume of MR image data, coupled with the potential complicated structure and distribution of tumors, tends to be a very time-consuming and operator-dependent task. As a result, the automation of tumor segmentation has become highly desirable in medical imaging community.

Agner *et al.*[7] summarized the two prevalent segmentation models used in the medical imaging domain: pixel-wise models[8-10] and shape-based models[11-13]. Recently, superpixel-based models[14-16] in conjunction with pixel-wise intensity information and geometric shape information have been applied to tumor segmentation. Lucchi *et al.*[14] used superpixel boundary appearance and graph-cuts to segment irregular-shaped cells on electron microscopy (EM) images, but did not consider pixel-wise label diversity present within each superpixel in DCE-MRI images. Wang and Yushkevich[15] applied label fusion to transfer supervoxel-wise labels from a multi-atlas, followed by a voxel-wise refinement. This method utilized spatial information only during supervoxel generation, not considering the geometry, texture or spatial topology of the supervoxels. A recent algorithm by Ashraf *et al.*[16] modeled the topology of superpixels as a multi-channel Markov random field (MRF) and converged the superpixel labeling through loopy belief propagation (LBP). This method, however, focused only on kinetic MRI features which, while reasonably robust, requires a sufficient number of post-contrast images.

Our study was motivated by the three superpixel-based methods above[14-16], and by combining the strengths we contributed to addressing existing deficiencies within segmentation on breast DCE-MRI. An automatic segmentation framework was proposed, incorporating mean-shift smoothing, superpixel-wise classification, pixel-wise graph-cuts partitioning, and morphological refinement. This framework was validated by manually segmented ground truth and compared with four state-of-the-art methods on a set of 15 DCE-MR images.

2. METHODS

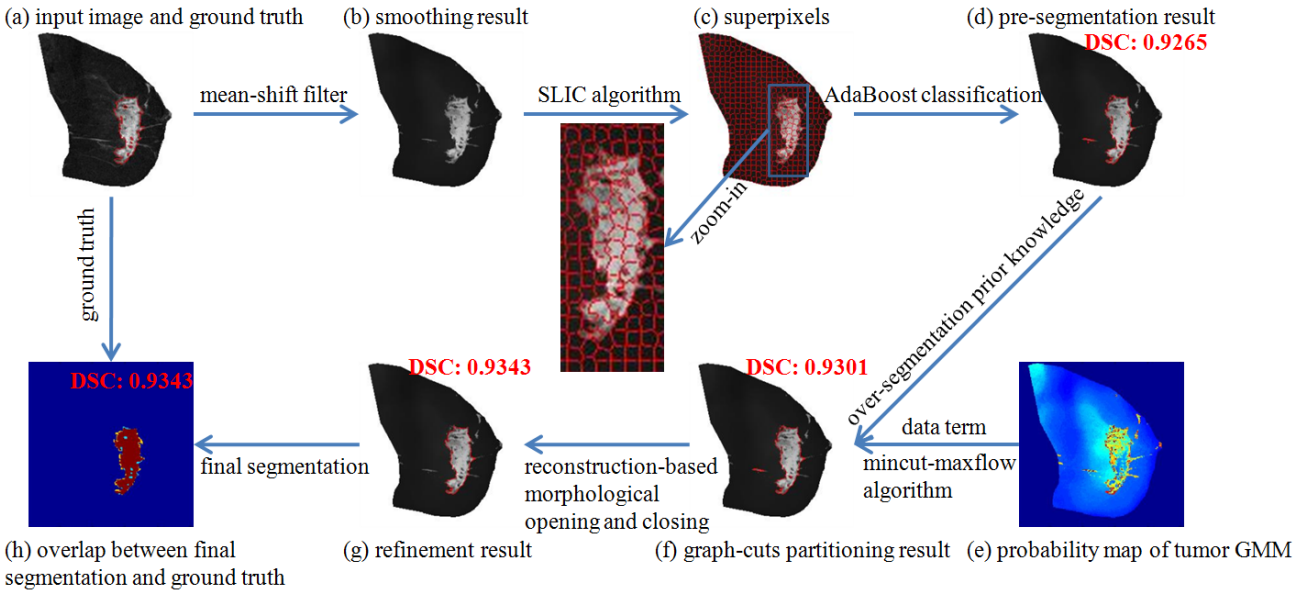

Fig. 1. The proposed framework overview. (a) Input subtraction image with manually segmented tumor contour in red. (b) Image after mean-shift smoothing. (c) Superpixels from SLIC. (d) AdaBoost classification as pre-segmentation result. (e) Probability map of tumor Gaussian Mixture Model. (f) Graph-cuts partitioning result. (g) Final segmentation result after morphological refinement. (h) Overlap between final automatic segmentation and manual segmentation. Brown region indicates overlap (true positive); cyan region indicates over-segmentation (false positive); and yellow region indicates under-segmentation (false negative).

2.1 Dataset

Our study was conducted on a longitudinal breast DCE-MRI dataset from 15 women with biopsy-proven stage 2 or 3 breast tumors obtained from the ACRIN 6657 I-SPY trial, a prospective clinical trial evaluating MR imaging for prediction of response to neoadjuvant chemotherapy. We only focused on the initial MRI examination of each patient which was performed within four weeks prior to the initiation of chemotherapy. T1-weighted 3D fat-suppressed gradient-echo sequences were done before (pre-contrast image) and after (post-contrast image) injection of a gadolinium-based contrast agent. The scans were performed in the sagittal direction with 2 mm or less thick contiguous slices. The in-plane pixel spacing ranged from 0.7031 to 0.7813 mm, and each slice consisted of 256 by 256 pixels. The post-contrast phase was acquired between 135 and 150 s. The proposed framework was directly performed on the subtraction images obtained by subtracting pixel-wise intensity values in the pre-contrast images from those in the post-contrast images. To avoid negative intensity values, we added a uniform offset of 500 to all subtraction images.

2.2 Mean-shift smoothing

A mean-shift smoothing filter[17] was first applied to minimize the inherent artifacts introduced by MR imaging. Unlike traditional filters, a mean-shift filter is a discontinuity-preserving smoothing approach which adaptively decreases the amount of smoothing near edges and preserves local structure (Fig. 1b). We empirically set the spatial kernel bandwidth to 32 and intensity kernel bandwidth to 8 for our experiments.

2.3 Superpixel generation

After smoothing, a Simple Linear Iterative Cluster (SLIC) algorithm[20] was used to aggregate nearby pixels into superpixels. Here we took two advantages of superpixels that 1) they converge to local optimum to adhere to local boundaries; and 2) they enrich discriminative and robust local features from geometric and textural information more than pixel-wise intensities. Compared to the normalized-cuts-based superpixel generation algorithm[21] used by Ashraf *et al.*[16] and the graph-merging-based algorithm[22] used by Wang and Yushkevich[15], Achanta *et al.*[20] validated the prime advantages of SLIC are that it is a spatially localized version of k-means clustering which is faster, adheres to boundaries better, and generates superpixels with more identical sizes to be more suitable for scale-variant feature representation. As a follow-up work, Lucchi *et al.*[14] used SLIC to generate superpixels on EM images and demonstrated the computational

advantage and promising segmentation results. Fig. 1c and the zoom-in show the superpixels generated from SLIC in our case, where we set the region size of superpixels to 8 and the spatial regularizer to 0.005 (i.e., a relatively small value) so as to ensure superpixels do not straddle image boundaries.

2.4 Feature extraction and classification

In order to robustly distinguish tumor regions from others, we extracted three types of features from each superpixel: intensity, geometric and textural features. As the vast majority of tumors enhance on breast DCE-MRI, it is straightforward to distinguish enhancing regions from non-enhancing regions via the use of intensity-based features. Statistical intensity features included the mean, standard deviation, skewness, kurtosis, range, and entropy of pixel intensities within each superpixel, the difference between means of pixel intensities inside and outside each superpixel, the means and standard deviations of gradients of the two orthogonal directions within each superpixel, and the intensity and gradient histograms of each superpixel. Since we computed histograms with 5 bins, 35 intensity features were extracted.

Geometric features were applied in order to distinguish blood vessels as outliers from other tissues because the superpixels effectively delineated the narrow shape of vessels. Geometric features included: 1) the *compactness* as the ratio of each superpixel perimeter (in pixels) and area (in pixels), 2) the *axis ratio* as the ratio of lengths (in pixels) of minor and major axis of the ellipse that has the same normalized second central moments of each superpixel, 3) the *eccentricity* of that ellipse, 4) the *irregularity* as the ratio of perimeters (in pixels) of the area-equivalent circle and each superpixel, 5) the *solidity* as the ratio of each superpixel area (in pixels) and the convex hull area (in pixels), and 6) the *extent* as the ratio of each superpixel area (in pixels) and the bounding box area (in pixels). Since each feature was a scalar, 6 geometric features were extracted.

Texture information was utilized to distinguish enhancing fibroglandular tissues (outlier) from other tissue types because they are less uniform in texture. We extracted a subset of Haralick features[23] from the bounding box of each superpixel including the *contrast*, *correlation*, *energy* and *homogeneity* of each co-occurrence matrix with position offsets of 1, 5, 9 (in pixels) and direction offsets of 0, 45, 90, 135 degrees separately. Therefore, there were 12 co-occurrence matrices being computed with the number of bins as 5 for 2D histogram counting (i.e. each co-occurrence matrix is 5 by 5), and, as a result, 48 textural features were extracted. In total, 89 features were computed in the proposed framework.

Finally, an AdaBoost classifier[24], combined with 256 binary decision trees pf maximal depth of 2, was trained to identify and select those superpixels containing tumor based on the aforementioned feature space (Fig. 1d). Note that in some case the training superpixels contained an admixture of pixels. As a result, we only selected those with more than 90% dominant labeling pixels as training superpixels.

2.5 Graph-cuts partitioning

With the pre-segmentation results, graph-cuts model[25] was used to partition the pixel-wise undirected graph into two labels (i.e., tumor and background). The nodes of the graph correspond to all of the individual pixels inside superpixel-based pre-segmentation while the edges correspond to a four-connectivity neighborhood system. Mathematically, an energy function taking the form in Eq. 1 is minimized:

$$E(f) = \lambda \underbrace{\sum_{p \in P} D(f_p \mid I_p)}_{\text{data term}} + \underbrace{\sum_{\{p,q\} \in N} V(f_p, f_q \mid I_p, I_q)}_{\text{smoothness term}} ; \quad (1)$$

where f_p, $f_q \in \{tumor, background\}$ are labeling variables, I_p, I_q are intensity values, P is the set of pixels, N is the neighborhood system, $D(f_p \mid I_p)$, contributing to the data term, measures the penalty of label f_p fitting to pixel p given observed data I_p, $V(f_p, f_q \mid I_p, I_q)$, contributing to the smoothness term, measures the penalty to assign different labels $\{f_p, f_q\}$ to each pair of $\{p, q\}$ in the neighborhood system given observed data $\{I_p, I_q\}$, and finally, λ is a trade-off factor controlling relative weight of the data term and the smoothness term.

The data penalty $D(f_p)$ is the logarithm probability of pixel p assigned to label f_p, which is the output of Gaussian Mixture Model (GMM). GMMs of the tumor and background labels were trained separately with 5 cluster centers. Pixels outside the superpixel-based pre-segmentation were assigned to background label with probability 1 and, therefore, were disregarded in the training process. The probability map of tumor GMM is shown in Fig. 1e. The smoothness penalty $V(f_p, f_q \mid I_p, I_q)$ is positively correlated with the difference between intensities I_p and I_q of adjacent pixel p and q as long as they are assigned to different labels:

$$V(f_p, f_q \mid I_p, I_q) = \begin{cases} e^{-\frac{(I_p - I_q)^2}{\sigma^2}}, & if \quad f_p \neq f_q \\ 0, & otherwise \end{cases}; \tag{2}$$

where σ is the standard deviation of image intensities inside pre-segmentation estimated from training images.

Tumor segmentation (Fig. 1f) was accomplished by minimizing Eq. (1) through mincut-maxflow algorithm[26], where λ was empirically set as 20.

2.6 Morphological refinement

Lastly, tiny, narrow and irregular segments were removed using reconstruction-based morphological opening and closing[12] (with radius 2 in pixels) in order to reinforce the robustness of the proposed framework. If there is prior knowledge that desired tumors appear regular and concentrated, this step becomes optional for the sake of efficiency. Fig. 1g displays an example of the final segmentation result. And Fig. 1h indicates the overlap between the automatic segmentation and ground truth.

3. EXPERIMENTS

3.1 Ground truth

One 2D slice containing the maximal tumor area was selected from each DCE-MRI volume for experimental validation. Manually-segmented tumor masks (as ground truths) and breast masks (as regions of interest) of those slices were performed by a board certified breast imaging radiologist using the ITK-SNAP software[27].

3.2 Evaluation metrics

The evaluation criteria of performance included area-based metrics: 1) Dice similarity coefficient (DSC)[7]; 2) pixel-wise accuracy[14]; 3) VOC score[14], and boundary-based metrics: 4) mean absolute difference (MAD)[7]; and 5) maximum Hausdorff distance (MHD)[7]. Higher values of DSC, accuracy, and VOC score, and lower values (in mm) of MAD and MHD indicate better agreement between automatic segmentation and ground truth.

3.3 Comparative methods

For comparison, four state-of-the-art methods for medical image segmentation adapted to the specific task of tumor segmentation were also evaluated on our dataset, as summarized in Table 1. All supervised methods were trained and tested in a leave-one-woman-out cross-validation strategy.

Table 1. The model and automation characteristics of the proposed framework and four comparative methods.

Method	*Model*	*Characteristic*
AdaFCM[10]	pixel-wise clustering + classification	automatic, supervised
Marker-controlled watershed[12]	morphological reconstruction + watershed	automatic, unsupervised
Adaptive geodesic transformation[13]	landmark-based distance map + thresholding	semi-automatic, unsupervised
Superpixel-based MRF[16]	superpixel + MRF + LBP	automatic, supervised
The proposed framework	superpixel-based classification + graph-cuts	automatic, supervised

Adaptive fuzzy c-means (AdaFCM)[10] is an adaptive strategy to select N-highest-intensity FCM[28] clusters as tumor regions via a dichotomous tumor-vs-non-tumor classifier. In our experiments, FCM was implemented with 25 clusters. To allow for a more fair comparison of the methods, we used the same intensity, geometric and textural features, as well as the same AdaBoost classifier used in our proposed framework (instead of support vector machine classifier utilized in the initial description of the algorithm[10]) in the implementation of AdaFCM used in this study.

Marker-controlled watershed segmentation[12] applies watershed transformation on the gradient image superimposed with markers, which are automatically selected from regional maxima after morphological reconstruction operation. In our implementation, a disk-shaped structuring element with radius 7 was utilized for reconstruction operation.

Adaptive geodesic transformation[13] is based on the geodesic distance map where intensity and gradient information are fused for geodesic transformation. It is adapted by the spatial variance of gradient weighted factor according to expert

landmarks. As a result, it is semi-automatic because of manually-initialized masks and landmarks, which are sensitive to the final results. In our experiments, the initial masks were manually selected while the expert landmarks were automated by the outcomes of AdaFCM[10]. In detail, gradient weighted factor was assigned 0 where pixels were labeled as tumor according to AdaFCM, and assigned 1 otherwise. The distance threshold is set to 30.9.

Superpixel-based MRF[16] is described in Sec 1. To allow for a fair comparison of the methods, we set the same region size of superpixels and the same number of clusters of GMMs as ours in the experiments. We also set the damping coefficient to 0.2.

3.4 Results

The proposed framework consistently outperformed the other four methods based on five selected metrics (Table 2). In comparison, the second best method, adaptive geodesic transformation, is a semi-automated algorithm in which results are dependent on the precision of manually-initialized masks. Moreover, the proposed framework performed most robustly on diverse samples as evidenced by the relatively small standard deviations in all metrics. Fig. 2 shows examples of segmentation results produced by all five methods. When tumors appeared to be irregular and heterogeneous, marker-controlled watershed, AdaFCM, and adaptive geodesic transformation tended to under-segment the tumor, while the superpixel-based MRF tended to over-segment the tumor. The proposed framework captured both the overlap and boundaries adequately.

Table 2. Statistical measures based on five metrics for performances of five tumor segmentation methods. **Boldface** indicates best performance and underline indicates second best performance among five.

Method	*DSC*[7]	*Accuracy*[14]	*VOC*[14]	*MAD*[7]	*MHD*[7]
AdaFCM[10]	0.72±0.27	0.88±0.23	0.61±0.28	2.52±7.98	18.46±28.15
Marker-controlled watershed[12]	0.65±0.24	0.82±0.30	0.52±0.22	4.51±12.63	16.01±34.11
Adaptive geodesic transformation[13]	0.78±0.14	0.95±0.04	0.66±0.18	1.67±2.71	36.84±39.77
Superpixel-based MRF[16]	0.69±0.17	0.88±0.10	0.56±0.19	2.64±3.02	19.02±12.97
The proposed framework	**0.83±0.12**	**0.96±0.03**	**0.72±0.16**	**0.79±1.43**	**11.71±12.44**

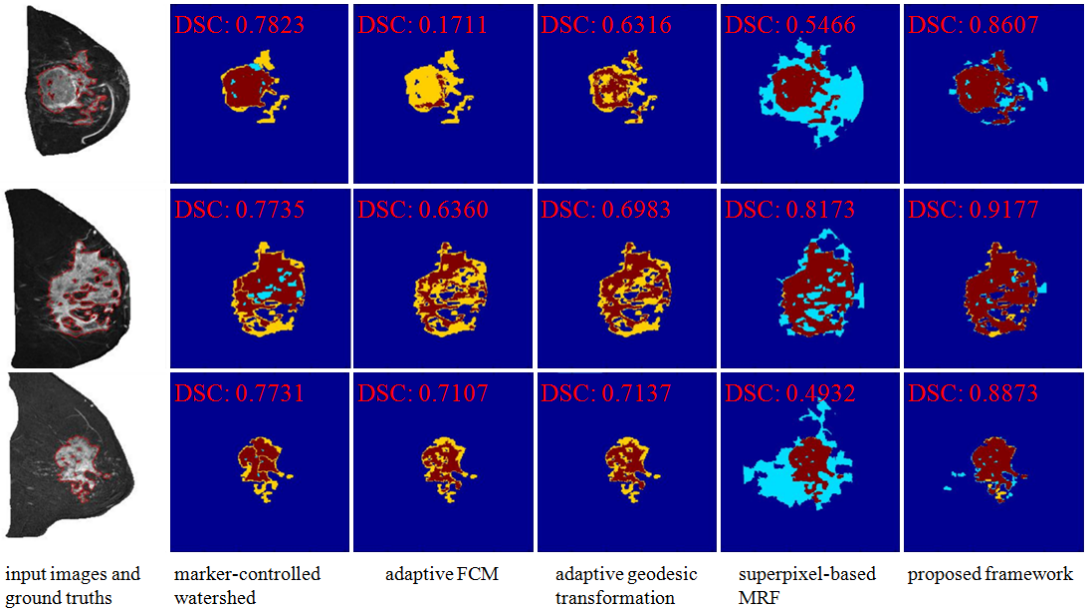

Fig. 2. Segmentation examples produced by five methods for three cases. The first column shows input subtraction images with ground truth contours in red. In the rest columns, brown, cyan, and yellow regions indicate overlap (true positive), over-segmentation (false positive), and under-segmentation (false negative) between automatic segmentation and ground truth, respectively. The Dice similarity coefficient (DSC) of each segmentation is also provided.

4. CONCLUSION

An automatic tumor segmentation framework was developed and validated, which leveraged intensity, geometric and textural information of superpixels, and the topology of pixel-wise graph for tumor segmentation for use in breast DCE-MRI. The proposed framework consistently outperformed four state-of-the-art segmentation methods in terms of superior capture of the overlap and boundaries of the tumors, demonstrating the potential as a means of improving clinical analysis and characterization of breast carcinomas.

5. ACKNOWLEDGMENTS

Support for this project was provided by funding from the University of Pennsylvania Abramson Cancer Center 2-PREVENT Center of Excellence Program and the Center for Biomedical Image Computing and Analytics (CBICA). The authors would like to thank Dr. Nola Hylton from UCSF for the permission to use the I-SPY data and the useful discussions on this research work.

6. REFERENCES

[1] R. Siegel, J. M. Ma, Z. H. Zou *et al.*, "Cancer Statistics, 2014," Ca-a Cancer Journal for Clinicians, 64(1), 9-29 (2014).

[2] S. H. Heywang, D. Hahn, H. Schmidt *et al.*, "Mr Imaging of the Breast Using Gadolinium-Dtpa," Journal of Computer Assisted Tomography, 10(2), 199-204 (1986).

[3] S. H. Heywang, A. Wolf, E. Pruss *et al.*, "Mr Imaging of the Breast with Gd-Dtpa - Use and Limitations," Radiology, 171(1), 95-103 (1989).

[4] W. A. Kaiser, and E. Zeitler, "Mr Imaging of the Breast - Fast Imaging Sequences with and without Gd-Dtpa - Preliminary-Observations," Radiology, 170(3), 681-686 (1989).

[5] E. L. Rosen, K. L. Blackwell, J. A. Baker *et al.*, "Accuracy of MRI in the detection of residual breast cancer after neoadjuvant chemotherapy," American Journal of Roentgenology, 181(5), 1275-1282 (2003).

[6] S. C. Partridge, J. E. Gibbs, Y. Lu *et al.*, "MRI measurements of breast tumor volume predict response to neoadjuvant chemotherapy and recurrence-free survival," American Journal of Roentgenology, 184(6), 1774-1781 (2005).

[7] S. C. Agner, J. Xu, and A. Madabhushi, "Spectral embedding based active contour (SEAC) for lesion segmentation on breast dynamic contrast enhanced magnetic resonance imaging," Medical Physics, 40(3), (2013).

[8] T. Twellmann, A. Meyer-Baese, O. Lange *et al.*, "Model-free visualization of suspicious lesions in breast MRI based on supervised and unsupervised learning," Engineering Applications of Artificial Intelligence, 21(2), 129-140 (2008).

[9] W. J. Chen, M. L. Giger, U. Bick *et al.*, "Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI," Medical Physics, 33(8), 2878-2887 (2006).

[10] B. M. Keller, D. L. Nathan, Y. Wang *et al.*, "Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation," Medical Physics, 39(8), 4903-17 (2012).

[11] K. H. Zhang, L. Zhang, H. H. Song *et al.*, "Active contours with selective local or global segmentation: A new formulation and level set method," Image and Vision Computing, 28(4), 668-676 (2010).

[12] K. Parvati, B. S. P. Rao, and M. M. Das, "Image Segmentation Using Gray-Scale Morphology and Marker-Controlled Watershed Transformation," Discrete Dynamics in Nature and Society, (2008).

[13] B. Gaonkar, L. Shu, G. Hermosillo *et al.*, "Adaptive geodesic transform for segmentation of vertebrae on CT images." Proc. SPIE Medical Imaging, 9035, 903516-903516-7 (2014).

[14] A. Lucchi, K. Smith, R. Achanta *et al.*, "A fully automated approach to segmentation of irregularly shaped cellular structures in EM images," Med Image Comput Comput Assist Interv, 13(Pt 2), 463-71 (2010).

[15] H. Wang, and P. A. Yushkevich, "Multi-atlas segmentation without registration: a supervoxel-based approach," Med Image Comput Comput Assist Interv, 16(Pt 3), 535-42 (2013).

[16] A. B. Ashraf, S. Gavenonis, D. Daye *et al.*, "A multichannel Markov random field approach for automated segmentation of breast cancer tumor in DCE-MRI data using kinetic observation model," Med Image Comput Comput Assist Interv, 14(Pt 3), 546-53 (2011).

[17] D. Comaniciu, and P. Meer, "Mean shift: A robust approach toward feature space analysis," Ieee Transactions on Pattern Analysis and Machine Intelligence, 24(5), 603-619 (2002).

[18] P. Perona, and J. Malik, “Scale-Space and Edge-Detection Using Anisotropic Diffusion,” Ieee Transactions on Pattern Analysis and Machine Intelligence, 12(7), 629-639 (1990).
[19] C. Tomasi, and R. Manduchi, "Bilateral filtering for gray and color images." 839-846.
[20] R. Achanta, A. Shaji, K. Smith *et al.*, “SLIC Superpixels Compared to State-of-the-Art Superpixel Methods,” Ieee Transactions on Pattern Analysis and Machine Intelligence, 34(11), 2274-2281 (2012).
[21] J. B. Shi, and J. Malik, “Normalized cuts and image segmentation,” Ieee Transactions on Pattern Analysis and Machine Intelligence, 22(8), 888-905 (2000).
[22] P. F. Felzenszwalb, and D. P. Huttenlocher, “Efficient graph-based image segmentation,” International Journal of Computer Vision, 59(2), 167-181 (2004).
[23] R. M. Haralick, Shanmuga.K, and I. Dinstein, “Textural Features for Image Classification,” Ieee Transactions on Systems Man and Cybernetics, Smc3(6), 610-621 (1973).
[24] R. Appel, T. Fuchs, P. Dollár *et al.*, "Quickly Boosting Decision Trees-Pruning Underachieving Features Early."
[25] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” Ieee Transactions on Pattern Analysis and Machine Intelligence, 23(11), 1222-1239 (2001).
[26] D. M. Greig, B. T. Porteous, and A. H. Seheult, “Exact Maximum a-Posteriori Estimation for Binary Images,” Journal of the Royal Statistical Society Series B-Methodological, 51(2), 271-279 (1989).
[27] P. A. Yushkevich, J. Piven, H. C. Hazlett *et al.*, “User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability,” Neuroimage, 31(3), 1116-1128 (2006).
[28] P. H. Wang, “Pattern-Recognition with Fuzzy Objective Function Algorithms - Bezdek,Jc,” Siam Review, 25(3), 442-442 (1983).