Combining Edge-based and Region-based Segmentations for Diagnosing Masses in Mammograms

Y. Zhang[1], N. Tomuro, J. Furst, D. Raicu[2]
[1] East-West University, Computer Information Science Department, Chicago, IL 60604, USA
[2] DePaul University, College of Computing and Digital Media, Chicago, IL 60604, USA

Keywords: Mammogram, Mass Segmentation, Mass Segmentation Evaluation

Purpose

Mass segmentation identifies a mass region from its background and captures the mass contour from a suspicious area. From the detected contour, the mass shape and spiculation features can be computed for classification. Previous studies have shown that improving the mass segmentation can significantly improve the accuracy of mass diagnosis [1]. The research presented in this paper is an ongoing project for developing an image-based CADx system to classify suspicious masses in mammograms as malignant or benign. In this paper, we propose to combine edge-based and region-based segmentations to improve overall segmentations in mammograms for mass classification.

Methods

1. Data Description and Mass ROI Extraction

In this work, all mass ROI images were extracted from the Digital Database for Screening Mammography (DDSM) from the University of South Florida [2]. For each suspicious mass, we extracted a rectangle image as a mass ROI, which includes the suspicious mass and its surrounding area. A total of 543 mass ROI images were used, where 272 instances were benign and 271 instances were malignant.

2. Build Three Edge-based Segmentors and Three Region-based Segmentors

Our previous research demonstrated that using multiple weak segmentors was an effective method to generate a strong mass segmentation for mammograms. Three edge-based segmentors were built from three enhanced images (using gamma corrections $\gamma = 1, 2, 5$, and Gaussian filter $\sigma = 5$) [3].

In addition, a region-based segmentation was developed as a complementary method. The method includes three steps: 1) apply a Gaussian mask to the mass ROI to enhance its center area and suppress the intensities of its surroundings; 2) use a variable threshold to generate possible mass regions and identify the mass region by morphological operation; 3) smooth the border of mass regions by a dilation operation. Applying three Gaussian masks of sigma values ($\sigma = 1$, $\sigma = 0.5$ and $\sigma = 0.25$), three region-based mass segmentation results were generated for each mass ROI. All procedures are implemented by using Matlab.

3. Mass Feature Extraction

After six segmentors were built (three edge-based and three region-based), 14 shape features were computed from a detected mass contour, they are: area, convex, perimeter, circularity, compactness, solidity, convex, roughness, equivalent diameter, elongation, major axis length, minor axis length, eccentricity and extent [4]. In addition, a segmentor success indicator was included as a binary feature. For those ROI images, whose contour could not be identified, their shape features were set as average values, so that they will have no influence in classifications. A total of 90 shape features were computed from six segmentors.

4. Evaluation of Segmentations

Segmentation success rates were used to evaluate the performance of each segmentor, where overlapping ratio of the identified mass region over the ROI central area was measured to indicate each segmentation result as successful or unsuccessful. However, a success rate cannot fully reveal the effectiveness of the segmentation for classification. In this study, we further investgate the effectiveness of segmentation for the mass classification. We fed the shape features extracted from six segmentors to a logistic regression (LR) model and obtained a set of feature regression weights. Since some regression weights are negative, we computed the feature regression weight as the square root of the absolute value of the regression weight of the model. Then, the average regression weights from each segmentor were computed to measure the effectiveness of segmentation for mass classification. In a LR model, features with larger regression weights (absolute values) indicate that they have more influence for classifications. Note that the model was built using 10-fold cross-validation, and all input features were normalized.

Results

Table 1 displays the individual segmentation rates and average regression weights from each segmentor. Generally, segmentors with higher segmentation success rates also have higher regression weights, which indicates that these segmentors are more effective for mass classification. The combination of two segmentation methods achieved a 99% overall segmentation success rate, which was significantly higher than the results by any single method. In the combined models, the average regression weights from the edge-based and region-based methods are close (0.50 vs. 0.66), the two methods were comparably effective in modeling the shapes of masses for classification.

Conclusion

Combination of region-based segmentors and edge-based segmentors is an effective approach for improving overall mass segmentation in mammograms for mass classification.

References

[1] Jiang L, Song E, Xu X, Ma G, Zhang B (2008). Automated Detection of Breast Mass Spiculation Levels and Evaluation of Scheme Performance. Acad Radiol.

[2] Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP (2001). The Digital Database for Screening Mammography. Proceeding of the 5th International Workshop on Digital Mammography. 212-218

[3] Zhang Y, Tomuro N, Furst JD, Raicu DS (2011). Multiple Weak Segmentors for Strong Mass Segmentation in Mammogram. 2011 SPIE Medical Imaging Conference.

[4] Choras R (2008). Shape and Texture Feature Extraction for Retrieval Mammogram in Databases. Information Tech. Biomedicine. 47:121-128

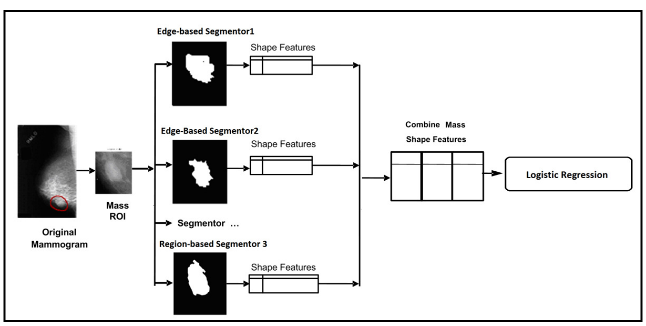

Figure 1. Overall Framework

Table 1: Three Edge-based and Three Region-based Segmentations

	Three Edge-based Segmentations			Three Region-based Segmentations		
Image Enhancement	γ =1	γ = 2	γ = 5	σ = 1	σ =0.5	σ=0.25
Segmentation Rate	66%	73%	77%	82%	93%	85%
Avg. Regression Weight	0.56	0.41	0.54	0.53	0.63	0.81
Avg. Regression Weight	0.50			0.66		
Overall Segment Rate	81%			95%		
Overall Segment Rate	99%					