CS 229 Project Milestone Report: Automated Segmentation of Breast Density

Rebecca L. Sawyer

Abstract—**This project is to develop an automated breast density segmentation system. Breast density has been shown to be a key indicator of breast cancer risk. Currently the primary computational system for analyzing breast density is a semi-automated segmentation system that requires a radiologist to set thresholds for the segmentation algorithm. This results in reader variability, which reduces the validity of the density assessment. I propose a fully automated segmentation system. The system will utilize superpixels and neighborhood information to create visual dictionaries from training data and use this information to segment image patches (small superpixels) as dense or not dense.**

I. INTRODUCTION

Breast cancer affects approximately 1 in 8 women in the United States, and it is the leading cause of cancer deaths amongst women worldwide [1]. It has be shown that early detection is an important factor in cancer survival. Mammography is beneficial for early detection of breast cancer [2]. Currently, the American Cancer Society recommends that women with average risk for breast cancer get yearly screening mammograms to detect potentially malignant findings at an early stage [3].

Of the mammographic risk factors for breast cancer, breast density has been shown to be a significant predictor. Women with increased breast density are four to six times more likely to develop breast cancer [4,5]. In current practice, the overall mammographic breast density is reported by the radiologist qualitatively on a four-point scale using the Breast Imaging, Reporting and Data System (BIRADS) [6]. Although this standard has specific reporting guidelines, it is qualitative and subject to the inherent variability of radiology reporting [7-9]. It has been suggested that quantitative breast density can provide a more robust predictor of mammography risk.

Recently, there has been growing interest in reproducible quantitative methods for estimating breast density. Initial methods were based on hand-drawn segmentations of the breast and dense tissue. However, with the development of digital mammography, more semi-automated approaches are being studied. In these methods, the radiologist outlines the area of the breast parenchyma and the total area of the breast, or semi-automatic segmentation of the mammogram is performed by radiologists adjusting software parameters [4]. The current state-of-the-art method for semi-automated segmentation is the Cumulus software [10,11]. This method requires the user to participate in producing the segmentation, which is time-consuming and disrupts the workflow. My goal was to develop a fully automated method of quantifying breast density. This would eliminate user variability and provide an efficient means to quantify breast cancer risk.

My previous efforts towards this goal also involved an unsupervised feature learning approach to classify image patches has dense or not dense. This method involved the extraction of 5x5 pixel image patches and then clustering of the patches using k-means. The final feature vector for each patch was the distance of the patch to each cluster, resulting in a vector of length k. The patch classification was then performed using L_1-regularized logistic regression classifier (LASSO). The results were evaluated by two metrics: per-pixel accuracy and the jaccard index as shown in Equation 1

$$JaccardIndex = \frac{predicted \bigcap true}{predicted \bigcup true} \qquad (1)$$

where R_1 and R_2, in our case, correspond to masks of the actual density and the predicted density. The average per-pixel accuracy was approximately 99%, which is not surprising considering the ratio of non-dense patches to dense patches. The average Jaccard index, however, was 0.605. This shows that there is room for improvement in breast density segmentation.

I made alterations to improve upon my previous work by making the following changes: (1) using superpixels instead of square patches and (2) incorporating neighbor information into the patch classification. I will explain more about these changes and why each of them could improve classification in the next section. Results of this work will be shown in Section III, and conclusions will follow.

II. MATERIALS AND METHODS

I built a system that takes an input of a cranio-caudal view of a breast and outputs a binary mask delineating dense regions as described in Figure 1. It works by dividing the image into superpixels using vlfeat-0.9.14 software package and characterizes these superpixels. The features are then used to train a LASSO classifier using glmnet software to estimate the probability of density given the features. Whole patches are classified as dense if the probability is higher than a given threshold that is determined during training. Each patch is then mapped to original breast image to create a density mask.

A. Data Set

I obtained 19 cranio-caudal mammography images that have been segmented by an expert radiologist using Cumulus software. The Cumulus company developed an algorithm that uses thresholding-based segmentation to distinguish dense and non-dense breast tissue [10,11]. The images are scanned film files and have been subsequently normalized to have equal optical
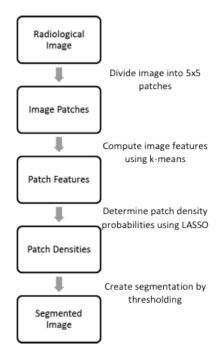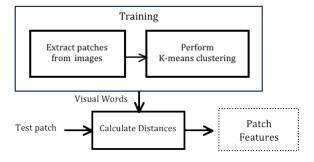
Fig. 1.   System Overview

Fig. 2.   Feature Extraction Flowchart

formed in step size S, a free parameter to optimize via cross-validation. Next, the algorithm computes a modified version of k-means that calculates distances (or "searches" for points of its class) in a $2Sx2S$ region around each centroid. The distance equations are shown below:

$$d_c = \sqrt{(g_i - g_j)^2} \qquad (2)$$

$$d_s = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \qquad (3)$$

$$D = \sqrt{d_c^2 + (\frac{d_s}{S})^2 m^2} \qquad (4)$$

where d_c is the distance between colors, d_s is the distance between columns and rows of the image, and D is the overall distance calculation. Note that d_c would normally consist of differences between r, g, and b values. However, since mammograms are grayscale images, this is simplified to the distance between the grayscale values. Also, of note is the parameter m in Equation 4. This parameter is used to control the tradeoff between spatial and intensity distances. It is another parameter to be chosen via cross validation.

The combination of spatial distance and intensity measures results in pixel groups that are more homogeneous than uniformly extracted square patches. The result is that these sets of pixels will be more accurately classified since each set is classified together.

2) Superpixel Features: The original system was modified to extract superpixels and to obtain SIFT features and statistical features (mean and variance). The features are then clustered using k-means in order to obtain a dictionary of visual words to use for classification. From there the system classifies using LASSO.

The SIFT algorithm first finds local maxima and minima by convolving the image with Gaussians at various scales and then taking the difference of the images. Maxima and minima are then found by comparison with neighboring pixels. This is repeated at each level near the maxima and minima of the previous level until the maxima or minima no longer appears. The maxima and minima that remain (considered stable) correspond to key locations in the image. The next step in to characterize the image and each of these key locations. This is accomplished first my computing the gradient magnitude, $M_{i,j}$, and the orientation, $R_{i,j}$, at each pixel, $A_{i,j}$:

$$M_{i,j} = \sqrt{(A_{i,j} - Ai + 1, j)^2 + (Ai, j - Ai, j + 1)^2} \quad (5)$$

$$R_{i,j} = \arctan 2(A_{i,j} - Ai + 1, j, A_{i,j+1} - A_{i,j}) \quad (6)$$

The gradient magnitudes are thresholded at 0.1 times the maximum possible gradient value in order to obtain limited effect by illumination changes. Next, each image is a assigned an orientation. This is done so that the image descriptors will be invariant to rotation. An orientation histogram is created by applying a Gaussian-weighted window. The weights are multiplied my the thresholded gradient magnitude, then binned to form a histogram in locations corresponding to $R_{i,j}$. Then, key location descriptors are obtained [14].

density values. The radiologist performing the segmentation is a trained expert in Cumulus and has over 18 years experience using semi-automated thresholding based density segmentation experience. I use these segmentation results as ground truth for training and evaluation of our system.

B. Image Processing

The images are analyzed by first segmenting the image into superpixels for analysis. These superpixels provide a robust way to describe local areas of an image. The superpixels are then characterized using real-valued features that are derived entirely from the data Figure 2. Once features are extracted for each patch, they are used to train a LASSO classifier to predict their density state [12].

1) Superpixel Extraction: Superpixels are groupings of pixels related by distance as well as intensity values. The algorithm I used to extract superpixels is the simple linear iterative clustering (SLIC) algorithm [13]. SLIC works by first initializing cluster centroids of column, row, intensity values at equally spaced points in a grid across the image. The grid is

After obtaining SIFT features, k-means is used to transform these features into a dictionary of visual words, using each cluster centroid as an individual word.

C. Superpixel Classification

The authors [15] performed similar image processing using the quick shift algorithm for obtaining superpixels, then finding SIFT features. They found success using support vector machines (SVM) for classification and then incorporating neighborhood information using conditional random fields (CRFs). Mathematically, $P(c|G; w)$ is the conditional probability of the set of class label assignments c given the adjacency graph $G(S, E)$ and a weight w:

$$-log(P(c|G; w) = \sum_{s_i \in S} \Psi(c_i|s_i) + w \sum_{s_i, s_j \in E} \Phi(c_i, c_j|s_i, s_j) \tag{7}$$

Here, S refers to the collection of superpixels in a given image. Also, we have:

$$\Psi(c_i|s_i) = -log(P(c_i|s_i)) \tag{8}$$

The calculation of $\Psi(c_i|s_i)$ comes from the SVM. And finally, $\Phi(c_i, c_j|s_i, s_j)$ is a measure of pairwise edge potentials:

$$\Phi(c_i, c_j|s_i, s_j) = \frac{L(s_i, s_j)}{1+ \parallel s_i - s_j \parallel} 1[c_i \neq c_j] \tag{9}$$

where $1[c_i \neq c_j]$ is the indicator function and $\parallel s_i - s_j \parallel$ is the norm of the color difference between superpixels. $L(s_i, s_j)$ is the shared boundary length between superpixels s_i and s_j.

My density segmentation system was modeled after this system. However, I tested my system with both SVM and LASSO.

D. Evaluation

The system was eveluted via leave-one-out cross-validation. Two metrics were used to describe the system: (1) per-pixel accuracy and (2) the Jaccard index as defined in Equation 1, which provides a measure of the amount of overlap of the actual dense region and the predicted dense region.

III. RESULTS

The system was tested on 19 675x925 pixel mammograms with two types of features: (1) SIFT features and (2) statistical features and two different classifiers: SVM and LASSO. Unfortunately, do to the large size of the images the SIFT feature extraction resulted in extremely long run times of several days on the Barley cluster. These jobs were aborted, and results were not able to be obtained. Additionally, the SVM classifier resulted in majority rule classification. Therefore, the results are not included here.

Results as shown in Table 1 were obtained with LASSO classification using statistical features with and without CRF computations included. The results with CRF regularization were significantly lower than those without. Figure 3 shows examples of segmentation accomplished with my system. The

	Mean Accuracy	Mean Jaccard Index
Original	0.894	0.605
Modified without CRF	0.906	0.584
Modified with CRF	0.811	0.302

TABLE I
MEAN ACCURACY AND MEAN JACCARD INDEX OF THE ORIGINAL SYSTEM, THE MODIFIED SYSTEM WITHOUT CRFs, AND THE MODIFIED SYSTEM WITH CRFs.

Fig. 3. Top left: True density mask; Top right: Prediction from original system, Jaccard = 0.89; Bottom left: Prediction from modified system without CRFs, Jaccard = 0.82, Bottom right: Prediction from modified system with CRFs, Jaccard = 0.35

first image shows the true segmentation as obtained from the radiologist. The second, third, and fourth images show segmentations from the square-patch-based system from my prior work, the modified system without CRFs, and the modified system with CRF, respectively. It is clear from these images that use of CRFs result in over-segmentation.

Overall, the modified system without CRFs showed comparable results than the original system.

IV. Conclusion

Breast density is a key indicator of breast cancer, and as such, breast density segmentation of mammograms is important for assessing cancer risk. My prior work showed some promising results, but modifications are necessary for further improvement. In that view, I have altered the system to obtain superpixels instead of square patches with the intention of classifying more homogeneous patches of the image. Second, I implemented CRF regularization on the predictions obtained from LASSO.

The current modifications to the original breast density segmentation algorithm unfortunately have not resulted in improvement. However, there are many further improvements that can be made to increase segmentation accuracy. First, there are several parameters that can be optimized in the current system: the number of visual words in the dictionary (k in k-means), the size of the superpixels (s in the search window of the SLIC algorithm), and the weight giving the tradeoff between spatial and intensity components of the SLIC algorithm (w). Next, implementing the system in a programming language other than the current language (Matlab) could result in faster processing time. This would allow for obtaining smaller superpixels as well as extracting more features to describe the superpixels. Further, other classifiers may result in improved classification.

V. Acknowledgements

VI. References

[1] Jemal A., et al., Global Cancer Statistics, CA Cancer J Clin. 2011 Mar-Apr;61(2):134.

[2] Nystrm L, Andersson I, Bjurstam N, Frisell J, Nordenskjld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet. 2002Mar.16;359(9310):90919.

[3] Smith RA, Saslow D, Sawyer KA, Burke W, Costanza ME, Evans WP, et al. American Cancer Society Guidelines for Breast Cancer Screening: Update 2003. CA: A Cancer Journal for Clinicians. John Wiley & Sons, Ltd; 2003;53(3):14169.

[4] Harvey JA, Bovbjerg VE. Quantitative Assessment of Mammographic Breast Density: Relationship with Breast Cancer Risk1. Radiology. Radiological Society of North America; 2004;230(1):2941.

[5] Nelson HD, Zakher B, Cantor A, Fu R, Griffin J, O'Meara ES, et al. Risk Factors for Breast Cancer for Women Aged 40 to 49 Years. Ann. Intern. Med. 2012Jan.1;156(9):63548.

[6] Baker JA, Kornguth PJ, Floyd CE. Breast imaging reporting and data system standardized mammography lexicon: observer variability in lesion description. AJR Am J Roentgenol. 1996Apr.;166(4):7738.

[7] Boyer B, Canale S, Arfi-Rouche J, Monzani Q, Wassef K, Balleyguier C. Variability and errors when applying the BIRADS mammography classification. European journal of radiology. 2012Apr.5.

[8] Jackson SL, Taplin SH, Sickles EA, Abraham L, Barlow WE, Carney PA, et al. Variability of interpretive accuracy among diagnostic mammography facilities. J. Natl. Cancer Inst. 2009Jun.3;101(11):81427.

[9] Beam CA, Layde PM, Sullivan DC. Variability in the Interpretation of Screening Mammograms by US Radiologists: Findings From a National Sample. Arch Intern Med. 1996Jan.22;156(2):20913.

[10] Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. Automated analysis of mammographic densities. Phys Med Biol. 1996May;41(5):90923.

[11] Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. Phys Med Biol. 1994Oct.;39(10):162938.

[12] Tibshirani RT. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological). Blackwell Publishing for the Royal Statistical Society; 1996Jan.1;58(1):26788.

[13] Achanta, Radhakrishna, et al., SLIC Superpixels Compared to State-of-the-art Superpixel Methods, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.34, no.11, pp.2274-2282, Nov. 2012.

[14] Lowe, D.G., "Object recognition from local scale-invariant features," Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on , vol.2, no., pp.1150-1157 vol.2, 1999.

[15] Fulkerson, Brian, et al., "Class segmentation and object localization with superpixel neighborhoods," Computer Vision, 2009 IEEE 12th International Conference on , vol., no., pp.670-677, Sept. 29 2009-Oct. 2 2009.