A Statistical Approach to Modelling Breast Tissue Appearance in Mammograms

Zhili Chen[1,3]
zzc09@aber.ac.uk

Erika Denton[2]
Erika.denton@nnuh.nhs.uk

Reyer Zwiggelaar[1]
rrz@aber.ac.uk

[1]Department of Computer Science
Aberystwyth University, Aberystwyth, UK

[2]Department of Breast Imaging, Norfolk and
Norwich University Hospital, Norwich, UK

[3]Faculty of Information and Control Engineering
Shenyang Jianzhu University, Shenyang, China

Abstract

In this paper, we present an approach to build models of breast tissue appearance in mammograms. Mammographic tissue is modelled based on a statistical analysis of local appearance. We investigate five strategies by using different types of local features, covering aspects of intensity, texture, and geometry. A visual dictionary is generated to summarise local tissue appearance with descriptive "words". The global appearance of the breast is represented as an occurrence histogram over the dictionary. The resulting histogram models can be applied to breast density classification. The validity is qualitatively and quantitatively evaluated using the full MIAS database. The consensus of three experts according to the BIRADS criterion is used as the classification ground truth. We test the performance of each individual strategy and the combination of all strategies. The results indicate that our approach has potential for mammographic risk assessment.

1 Introduction

Many studies have indicated that there is a strong correlation between breast tissue density/patterns and breast cancer risk. Therefore, modelling mammographic tissue appearance is beneficial for the qualitative perception of breast tissue patterns, quantitative analysis of breast density, and automated mammographic risk assessment. Recently, a variety of approaches have been developed to characterise breast tissue in mammograms [4, 7, 8, 12]. In general, intensity information is used to describe tissue density and texture information is used to represent tissue patterns. In this paper, we present a statistical approach to model breast tissue appearance over the whole breast. We focus on local tissue appearance in terms of intensity, texture, and geometry. Local features are extracted from the neighbourhood of breast tissue pixels and are statiscally analysed to build overall models of breast tissue.

2 Methodology

The schema of our method is shown in Fig. 1. Firstly, as a preprocessing step, the breast region is segmented using the approach in [1]. Subsequently, features are extracted from the

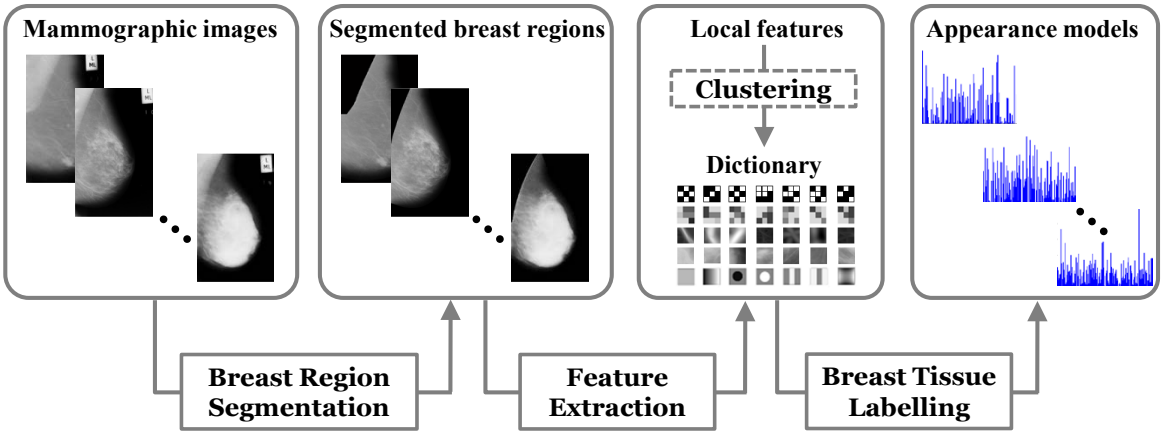

Figure 1: The schema of modelling breast tissue appearance based on the statistical analysis.

local neighbourhood of breast tissue pixels within the breast region. After that, a visual dictionary of local tissue appearance is generated in two ways: one is aggregating an exhaustive configuration of a certain type of local features; the other is performing an initial training step to "learn" clusters of local features. Finally, each breast tissue pixel is labelled by searching for the nearest "word" in the dictionary. Each mammographic image is represented as an occurrence histogram of the "words" in the dictionary. We investigate five different strategies for local features, including local binary patterns, local greylevel appearences, local geometric structures, joint filter responses, and raw image patches, respectively.

Local Binary Patterns (LBP) were first proposed in [6]. Local appearance of breast tissue is encoded into a set of binary values. The greylevel value of the centre pixel is substracted from the local neighbourhood, and a binary label is assigned to each neighbouring pixel according to the difference sign. The resulting binary pattern is transformed into a unique LBP number by $LBP = 1 + \sum_{p=0}^{P-1} s(g_p - g_c)2^p$, where g_c is the greylevel value of the centre pixel, g_p denotes the greylevel value of the p^{th} pixel in the local neighbourhood, and $s(x) = 1$ if $x \geq 0$ else $s(x) = 0$. Thus, each LBP number corresponds to a unique local binary pattern, and all possible patterns comprise the visual dictionary of local appearance. An LBP histogram is populated by counting the occurrences of LBP numbers at every pixel.

The Local Greylevel Appearance (LGA) based approach was presented in [12]. Breast tissue appearance is modelled by analysing the joint greylevel distribution of the local neighbourhood. The local greylevel appearance is transformed into a unique LGA number by $LGA = 1 + \sum_{i,j} N_g{}^{counter(i,j)} I(i,j)$, where N_g is the greylevel resolution, $counter(i,j)$ is the sequence number of pixel (i,j) within the neighbourhood, and $I(i,j)$ is the greylevel value of pixel (i,j). Thus, each LGA number corresponds to a unique local greylevel appearance, and all possible greylevel appearances comprise the visual dictionary. An LGA histogram containing the combination of LGA numbers and corresponding occurrences is generated.

Basic Image Features (BIF) were defined in [2]. A second-order family of six Gaussian derivative filters are used to analyse local appearance of breast tissue with respect to multi-scale geometric structures. Mammographic images are convolved with the Gaussian filter bank at multiple scales. Seven BIFs are defined, each corresponding to a distinct type of local geometric structures. For each breast tissue pixel, the corresponding geometric structure is determined according to the largest BIF computed with its local neighbourhood. The configuration of the seven geometric structures across multiple scales is encoded into a unique BIF column, and all possible configurations comprise the visual dictionary. Breast tissue

appearance is modelled by counting the occurrences of BIF columns over the whole breast.

In this paper, we use two types of textons. The first type of textons are generated by filtering a set of training mammograms with a filter bank and clustering the joint filter responses. The cluster centres are considered as textons. We use the MR8 filter bank defined in [10], where only eight filter responses are retained. The aggregated filter responses of breast tissue pixels over all the training mammograms are clustered using the classic K-Means algorithm. The resulting textons comprise the dictionary of local tissue appearance. Breast tissue appearance is modelled by a frequency histogram of textons. The second type of textons are generated by clustering raw image patches extracted from the training mammograms [11]. The procedure of learning textons is similar as described above. A frequency histogram of textons based on source image patches constructs the appearance model of breast tissue. It has been demonstrated in [11] that the image-patch based textons can provide superior performance to the filter-response based textons for the sake of texture classification. We refer to the two modelling strategies based on the MR8 and image-patch textons as Texton I and Texton II, respectively. Note that the preprocessing step in [10, 11] is not performed in Texton I and Texton II, and the filter responses in Texton I are not normalised by Weber's law as in [10, 11], in order to retain the original intensity correlation between mammograms.

3 Experiments and Results

In our experiments, the Mammographic Image Analysis Society (MIAS) database [9] was used, which contains 322 Medio-Lateral Oblique (MLO) mammograms taken from 161 women. The original spatial resolution is $50\mu m \times 50\mu m$ per pixel. Due to memory and efficiency reasons, we downsampled the full resolution to $800\mu m \times 800\mu m$ per pixel. Three experts classified 321 available mammograms (mdb295ll was excluded for historical reasons) into four Breast Imaging Reporting and Data System (BIRADS) density categories [5]. The consensus between individual classification decisions was considered as the ground truth. We built appearance models for each mammogram using the five strategies described above.

When modelling breast tissue appearance, to eliminate bias caused by mammogram edges and the breast-background boundary, we only focused on the pixels with neighbourhoods entirely located within the breast region, which can be identified automatically based on the pre-segmented breast region. For LBP, we used a 3×3 neighbourhood. For LGA, we also used a 3×3 neighbourhood and reduced the greylevel resolution to 16. For BIF, we convolved mammographic images with the Gaussian filter bank at four scales. For the two texton based methods, at the texton learning stage, we randomly selected 40 mammograms from the database as the training set. For texton II, we extracted 3×3 image patches and rearranged the pixels in row order for the K-means clustering. For each individual method, we learned 160 textons from the training set. Thus, the number of "words" generated for each type of local feature was 2^9, 16^9, 7^4, 160 and 160, respectively

For qualitative evaluation, we display the results in the form of label maps. To generate a label map, at each breast tissue pixel, we searched for the nearest "word" to the extracted local feature across the dictionary, and labelled each pixel according to the response "word". Fig. 2 shows the resulting label maps of example mammograms covering the four BIRADS categories. For LBP, LGA and BIF, the "words" were sorted according to the order of LBP number, LGA number and BIF column. For texton I and texton II, the "words" were sorted according to the magnitude value of textons in ascending order before labelling the pixels. It is shown that LBP is sensitive to noise and small textures, and there are no homogeneous

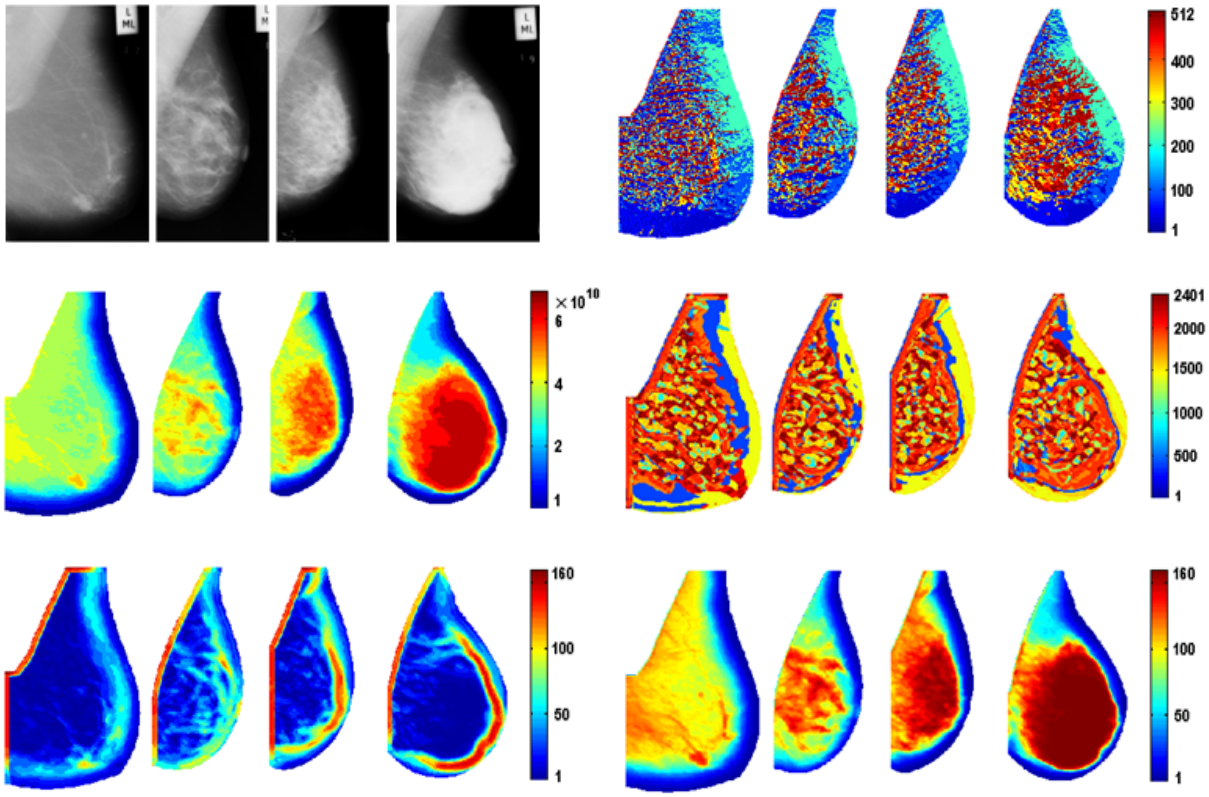

Figure 2: Example mammograms and resulting label maps using different strategies: top row: original mammograms and LBP; middle row: LGA and BIF; bottom row: Texton I and Texton II. For each batch, from left to right corresponds to an increasing BIRADS category.

regions obtained within the breast area. Both LGA and Texton II indicate realistic segmentation with respect to tissue density. Texton I has strong responses to the boundary between dense and fatty tissue regions, and relatively homogeneous regions are obtained within the dense/fatty tissue. For BIF, cell-like regions are obtained within the breast area, corresponding to different types of geometric structures.

For quantitative evaluation, we applied the resulting models in the form of occurrence histograms (L_1 normalised) to breast density classification. As described in Section 2, the occurrence histograms were populated by spanning the full dictionary. This could result in sparse histograms in the real observation of breast tissue. Therefore, we avoided using the full range of the histograms for the classification. We sorted the histogram bins in descending order to select the dominant "words" from the dictionary. We chose the most frequently used "words" occupying 99% occurrences as the dominant "words". Thus, the histogram bins corresponding to the "words" never or rarely referenced were removed. In our experiments, it was indicated that the feature space dimensionality was significantly reduced after compressing the visual dictionary, especially for LBP, LGA, and BIF.

A leave-one-woman-out methodology was used for the evaluation. We applied the multi-class boosting algorithm for the classification. We chose to use the Gentleboost algorithm [3] to train a classifier for each strategy, where the weak learners used in each of serial rounds are decision stumps which are regarded as degenerate decision trees with a single node. The created classifiers can provide continuous-valued outputs, which are interpreted as posterior probabilities. For a testing mammogram with a histogram-like model x, the classifier output corresponding to the class B_j is the probability of x to belong to B_j, denoted by $p_{B_j}(x)$. To

Table 1: Confusion matrices for breast density classification using different strategies.

LBP

Truth \ BIRADS	I	II	III	IV	CA
I	72	13	2	0	83%
II	14	60	27	2	58%
III	7	13	66	8	70%
IV	0	5	21	11	30%

LGA

Truth \ BIRADS	I	II	III	IV	CA
I	69	15	3	0	79%
II	19	47	37	0	46%
III	1	25	64	4	68%
IV	1	4	5	27	73%

BIF

Truth \ BIRADS	I	II	III	IV	CA
I	74	13	0	0	85%
II	14	63	25	1	61%
III	1	27	54	12	57%
IV	0	2	19	16	43%

Texton I

Truth \ BIRADS	I	II	III	IV	CA
I	72	15	0	0	83%
II	12	73	18	0	71%
III	2	26	55	11	59%
IV	0	4	11	22	59%

Texton II

Truth \ BIRADS	I	II	III	IV	CA
I	76	10	1	0	87%
II	11	71	18	3	69%
III	1	23	64	6	68%
IV	1	2	14	20	54%

Comb.

Truth \ BIRADS	I	II	III	IV	CA
I	79	8	0	0	91%
II	11	78	14	0	76%
III	1	19	70	4	74%
IV	1	1	11	24	65%

make the sum of the outputs over all the four BIRADS classes equal to 1, they are normalised by $P(B_j \mid x) = e^{p_{B_j}(x)} / \sum_{c=1}^{4} e^{p_{B_c}(x)}$, where $P(B_j \mid x)$ is the normalised value of $p_{B_j}(x)$.

Table 1 shows the classification results obtained by using the five modelling strategies. The overall classification accuracy (CA) is 65%, 64%, 64%, 69%, and 72%, for LBP, LGA, BIF, Texton I, and Texton II, respectively. The best classification result is obtained by Texton II, which is followed by Texton I, providing the second-best result, while LGA and BIF perform worst among these five methods. In addition, we investigated the performance of combining the outputs of the five individual classifiers. A weighted average combination rule was used to compute the total probability for each class. For a testing mammogram, the total probability corresponding to the class B_j (denoted by $P_{sum}(B_j)$) is obtained by $P_{sum}(B_j) = \sum_{t=1}^{5} w_t P(B_j \mid x_t)$, where $P(B_j \mid x_t)$ is the output of the t^{th} classifier, and w_t is the corresponding weight value. We set $w_1 = 0.16$, $w_2 = 0.12$, $w_3 = 0.12$, $w_4 = 0.32$, and $w_5 = 0.28$ experimentally, but small variations provided similar results. The obtained overall CA is 78%, which indicates better performance compared with those obtained by the individual classifiers. The resulting confusion matrix is also shown in Table 1.

We compared the obtained results with publications where the BIRADS criterion was used for breast density classification. Petroudi *et al.* [8] modelled parenchymal patterns of the whole breast with a statistical distribution of textons. They obtained an overall CA of 76% for the Oxford Database. Oliver *et al.* [7] extracted morphological and texture features from dense and fatty tissue regions. They obtained an overall CA of 77% for the MIAS database, which increased to 86% when the Bayesian combination of the kNN classifier and the C4.5 decision tree was used. He *et al.* [4] developed a number of mammographic image segmentation methods for mammographic risk assessment. The best classification accuracy that they recently obtained was 75%. It is shown that our results are comparable to related publications. Note that the same database and the same classification ground truth were used in [4, 7], which enables a direct comparison.

4 Discussion and Conclusions

There are several parameters in the five modelling strategies. One common parameter in LBP, LGA, and Texton II is the size of the local neighbourhood of breast tissue pixels.

Using a large neighbourhood will drastically increase the feature space dimensionality. In LBP and LGA, the number of histogram bins grows exponentially as the neighbourhood size increases, which would raise the risk of overfitting the data and result in sparser histograms. A second parameter in LGA is the greylevel resolution, which also has a significant effect on the feature space dimensionality. For the texton based strategies, increasing the number of textons would also result in an increased risk of overfitting for the K-Means clustering. The scale parameters in the filter banks and the image spatial resolution involve the multi-resolution aspect, which could be further investigated.

In summary, we have presented an approach to modelling breast tissue based on a statistical analysis of local appearance. We generated a visual dictionary of generic breast tissue appearance by aggregating local features across different density classes. We investigated five types of local features and evaluated their performance quanlitatively and quantitatively. To our knowledge, this work is the first attempt to combine different modelling strategies for breast density classification. The experimental results indicate the validity of the approach.

References

[1] Z. Chen and R. Zwiggelaar. Segmentation of the breast region with pectoral muscle removal in mammograms. In *Proc. MIUA*, 2010.

[2] M. Crosier and L. D. Griffin. Using basic image features for texture classification. *IJCV*, 88(3):447–460, 2010.

[3] J. Friedman et al. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, 28(2):337–407, 2000.

[4] W. He et al. Mammographic image segmentation and risk classification using a novel texture signature based methodology. In *LNCS*, volume 6136, pages 526–533, 2010.

[5] American College of Radiology. *Illustrated Breast Imaging Reporting and Data System BIRADS*. Amer. College of Radiol, 1998.

[6] T. Ojala et al. Multiresolution grey-scale and rotation invariant texture classification with local binary patterns. *PAMI*, 24(7):971–987, 2002.

[7] A. Oliver et al. A novel breast tissue density classification methodology. *TITB*, 12(1): 55–65, 2008.

[8] S. Petroudi et al. Automatic classification of mammographic parenchymal patterns: A statistical approach. In *Proc. EMBC*, pages 798–801, 2003.

[9] J. Suckling, *et al*. The Mammographic Image Analysis Society digital mammogram database. In *International Congress Series 1069*, pages 375–378, 1994.

[10] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *IJCV*, 62(1):61–81, 2005.

[11] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *PAMI*, 31(11):2032–2047, 2009.

[12] R. Zwiggelaar. Local greylevel appearance histogram based texture segmentation. In *LNCS*, volume 6136, pages 175–182, 2010.