Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Incorporating texture features in a computer-aided breast lesion diagnosis system for automated three-dimensional breast ultrasound

Haixia Liu
Tao Tan
Jan van Zelst
Ritse Mann
Nico Karssemeijer
Bram Platel

SPIE.

Incorporating texture features in a computer-aided breast lesion diagnosis system for automated three-dimensional breast ultrasound

Haixia Liu,[a,b,*] Tao Tan,[a] Jan van Zelst,[a] Ritse Mann,[a] Nico Karssemeijer,[a] and Bram Platel[a]
[a]Radboud University Medical Centre, Department of Radiology and Nuclear Medicine, 6525 GA Nijmegen, The Netherlands
[b]University of Nottingham Malaysia Campus, School Of Computer Science, Room BB79, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

Abstract. We investigated the benefits of incorporating texture features into an existing computer-aided diagnosis (CAD) system for classifying benign and malignant lesions in automated three-dimensional breast ultrasound images. The existing system takes into account 11 different features, describing different lesion properties; however, it does not include texture features. In this work, we expand the system by including texture features based on local binary patterns, gray level co-occurrence matrices, and Gabor filters computed from each lesion to be diagnosed. To deal with the resulting large number of features, we proposed a combination of feature-oriented classifiers combining each group of texture features into a single likelihood, resulting in three additional features used for the final classification. The classification was performed using support vector machine classifiers, and the evaluation was done with 10-fold cross validation on a dataset containing 424 lesions (239 benign and 185 malignant lesions). We compared the classification performance of the CAD system with and without texture features. The area under the receiver operating characteristic curve increased from 0.90 to 0.91 after adding texture features ($p < 0.001$). [DOI: 10.1117/1.JMI.1.2.024501]

Keywords: three-dimensional automated breast ultrasound; computer-aided diagnosis; texture features; local binary patterns; gray level co-occurrence matrix texture features; Gabor filters; three-dimensional texture features.

Paper 14020RR received Mar. 4, 2014; revised manuscript received Jun. 22, 2014; accepted for publication Jun. 26, 2014; published online Jul. 25, 2014.

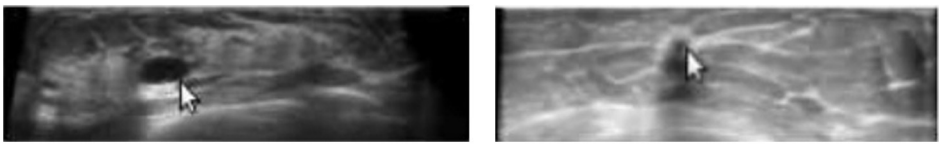
1 Introduction

Breast cancer is the most frequently diagnosed form of cancer in women. More than 1.3 million women all over the world are diagnosed with breast cancer each year.[1] Mammography has been introduced as a primary modality for breast cancer screening. The cancer mortality rate was reduced due to the screening[2,3] and also due to advanced treatments.[4,5] The breast cancer mortality rate reduction due to screening was estimated at between 16% and 48%.[6] Although it is cost-effective (for women 40 to 79 years old),[7] mammography screening still has limitations in women with dense breasts defined as category 3 and 4 by the American College of Radiology (ACR).[8,9] Therefore, supplemental modalities such as magnetic resonance imaging (MRI) and ultrasound (US) are recommended. US, as a complementary tool to mammography, has shown its strength in detecting small, early-stage invasive cancers in dense breasts.[10,11] Hand-held ultrasound (HHUS) has limitations such as operator dependency and inability to image and store three-dimensional (3-D) volumes of the breast. To overcome these problems, automated 3-D breast ultrasound (ABUS) was developed. This modality makes it possible to visualize large sections of the breast from the skin surface to the chest wall at once and store entire breast volumes on a picture archiving and communication system, enabling temporal comparison of current studies with relevant priors. Moreover, multiplanar reconstructions are possible in coronal and sagittal planes, which provides additional information. ABUS compresses the breast using a membrane and a transducer. Figure 1 shows two samples of two-dimensional (2-D) slices obtained from 3-D breast volume generated by ABUS, and Fig. 2 gives an example of coronal section from one breast, which is not available in HHUS. However, as the amount of acquired image data dramatically increases, the time needed by a radiologist to read these images increases consequently. Moreover, many benign lesions that remain undetected on mammograms are detected in the US images. It is often difficult for clinicians to decide correctly on either biopsy or follow-up for mass-like breast lesions seen on ultrasonographic images. Therefore, assistance by a computer-aided diagnosis (CAD) system would be useful to differentiate between benign lesions and breast cancers more efficiently and accurately.

New methods have been developed to improve the performance of the CAD system. Several 3-D US CAD systems have been described in the literature. Sahiner et al.[12] developed computer algorithms to automatically delineate mass boundaries and extract features on the basis of segmented mass shapes and margins. They constructed a computer classifier to merge features into a malignancy score, which was used by radiologists, and came to the conclusion that the use of a computer algorithm may improve radiologists' accuracy in distinguishing malignant from benign breast masses on 3-D US volumetric images. Moon et al.[13] developed a CAD system using speckle features and concluded that combining speckle and morphological features

*Address all correspondence to: Haixia Liu, E-mail: liu.haixia@yahoo.com

0091-3286/2014/$25.00 © 2014 SPIE



(a) (b)

Fig. 1 Two-dimensional transversal slices are demonstrated. (a) One benign lesion and (b) one malignant lesion, with arrows pointing to the lesions.

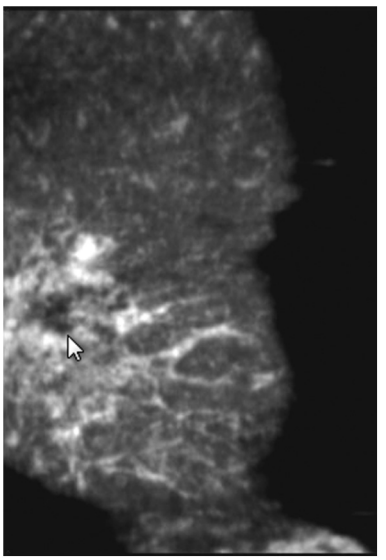


Fig. 2 A coronal slice with a malignant lesion pointed out by an arrow.

yielded an area under the receiver operating characteristic curve (AUC) that was significantly better than the AUC of the morphological features alone. Moon et al.[14] further implemented shape, orientation, margin, lesions boundary, posterior acoustic, and echo patterns features for computer-aided analysis. These combined features achieved the best performance.

Texture features have been applied to the classification of benign and malignant breast masses in MRI images and mammograms. Nagarajan et al.[15] showed that the dynamic texture quantification of the lesion enhancement patterns can significantly improve the performance of automated lesion classification. Hussain[16] proposed a method to classify mass regions by building an ensemble classifier that employs Gabor features. Their findings showed that the classifier based on the Gabor features achieved the best result. Nanni et al.[17] stated that local ternary patterns and an ensemble of local phase quantization are promising texture feature descriptors for mass classification in mammograms. Other studies have shown that the texture features can distinguish malignant from benign lesions in 2-D US.[18,19]

Previously, Tan et al.[20] investigated the performance of 11 features extracted from ABUS volumes for distinguishing malignant from benign masses. These 11 features are variance of intensities, entropy, average intensity, margin contrast, volumetric height-to-width ratio, sphericity, compactness, posterior acoustic behavior, and spiculation. Most of these features are mathematical descriptions of the features listed in the breast imaging reporting and data system standardized lexicon for US. Motivated by the discriminant power of texture features in MRI, mammography, and 2-D breast US images, in this work, we extended the feature pool of the existing CAD system with a set of texture features that have been already demonstrated as effective descriptors for recognizing patterns.[21] Specifically, we studied the utilization of texture features for the classification between benign and malignant lesions in 3-D breast US volumes.

2 Materials and Methods

2.1 *Dataset*

The ABUS volumes used in this study were provided by four different institutes, including Radboudumc (Nijmegen, The Netherlands), Falun Central Hospital (Falun, Sweden), the Jules Bordet Institute (Brussels, Belgium), and the Jeroen Bosch Ziekenhuis (Hertogenbosch, The Netherlands). Two types of ABUS systems were used: SomoVu ABUS system developed by U-systems [Sunnyvale, California (on November 9, 2012, the healthcare division of General Electric-GE, announced the acquisition of U-Systems)] and ACUSON S2000 automated breast volume scanning system developed by Siemens (Mountain View, California). In total, we collected 424 lesions from 165 patients, including 239 benign lesions and 185 malignant lesions. For volume processing and analysis, the original ABUS volumes were resampled to 0.6-mm cubic voxels. The average diameters of all lesions, benign lesions, and malignant lesions in our dataset are 12.53 ± 6.94 mm, 14.54 ± 6.72 mm, and 11.31 ± 6.79 mm, respectively.

Due to the use of different types of ABUS systems and the variations of scan parameter settings, the intensity levels of tissue types (fatty tissue, dense tissue, etc.) varied between images. Therefore, image intensities were normalized according to the intensities of segmented dense tissue and fatty tissue.[20]

2.2 *CAD System*

Previously, we have developed a CAD system[22] for the classification of benign and malignant lesions in 3-D ABUS volumes. After the intensity normalization on the volumes, lesions were segmented automatically based on spiral scanning algorithm.[23] Examples of this segmentation are shown in Figs. 3 and 4. Spiculation patterns were added to the existing CAD system,[20] and it was shown that incorporating spiculation patterns in coronal planes perpendicular to the transducer resulted in a significant progress in the classification performance. In this study, we analyzed the lesion classification performance changes after adding texture features in ABUS. By incorporating texture features in the existing CAD system, we were able to analyze the contributions from different texture features. The amount of texture features we implemented greatly exceeds the number of samples we have in our dataset. To deal with this issue, we trained a separate classifier for each type of texture features. These feature-oriented classifiers (FOCs) merged each type of texture features to likelihood values, which were added to the 11 features of the existing CAD system.

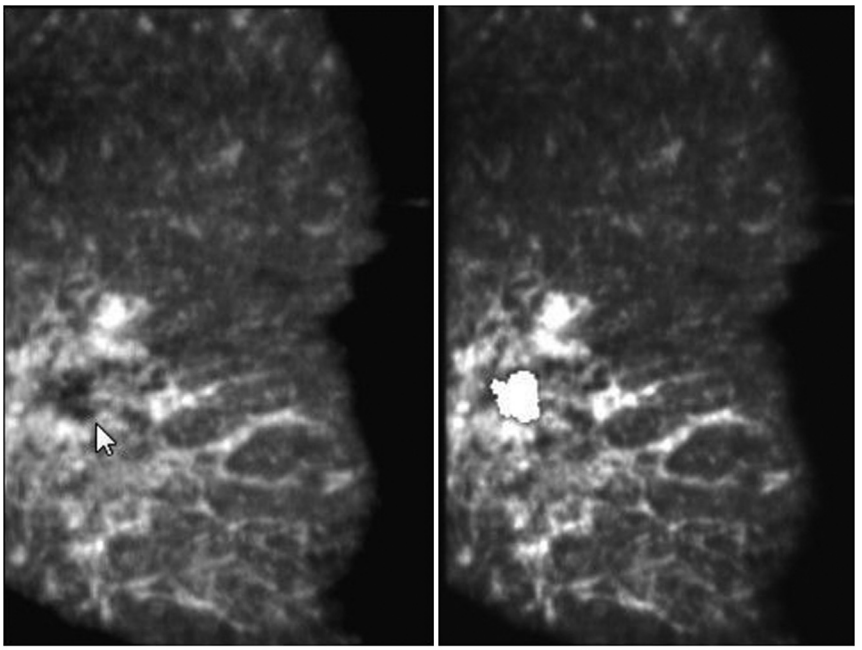


Fig. 3 Coronal slices with a malignant lesion (a) pointed out by the arrow and its segmentation (b) indicated by the white mask.

2.3 *Texture Features*

Texture features play an important role in medical image analysis for classification tasks. In this study, in particular, we investigated three important types of texture features, including local binary patterns (LBPs),[24] gray level co-occurrence matrix (GLCM)-based features,[25] and the Gabor filters.[26] Specifically, we integrated 2-D texture features computed from 2-D cross sections into the existing system.

2.3.1 *2-D cross sections*

We used a spiral-scanning-based dynamic programming technique, which was originally introduced by Wang et al.[23] for lesion segmentation. Based on the segmented masks, we calculated the mass centers to obtain the 2-D cross sections. In our experiment, 2-D LBP features were extracted from the segmented lesions indicated by white masks in Figs. 3 and 4. Two-dimensional GLCM features were computed inside a bounding box that frames a segmented lesion, that is to say we also included the surroundings of the lesions. Gabor features were extracted from a region with the lesion centered in each plane. The square region, where the Gabor features were calculated, is defined in such a way that the side length of the square equals to $2r$, where r is the referenced radius ($\pi r^2 = A$, where A represents the area of the segmented breast lesion). (We also compared the classification results that were based on the bounding box, where we computed the 2-D GLCM features with the results based on regions defined in this way. We found out that the later one performed better than the bounding box).

2.3.2 *Local binary pattern*

As a statistical and a structural textural features, LBPs have proven to be a useful descriptor in pattern recognition and texture feature classification.[24] Matsumoto et al.[19] studied the texture features of solid masses in 2-D breast US images and their

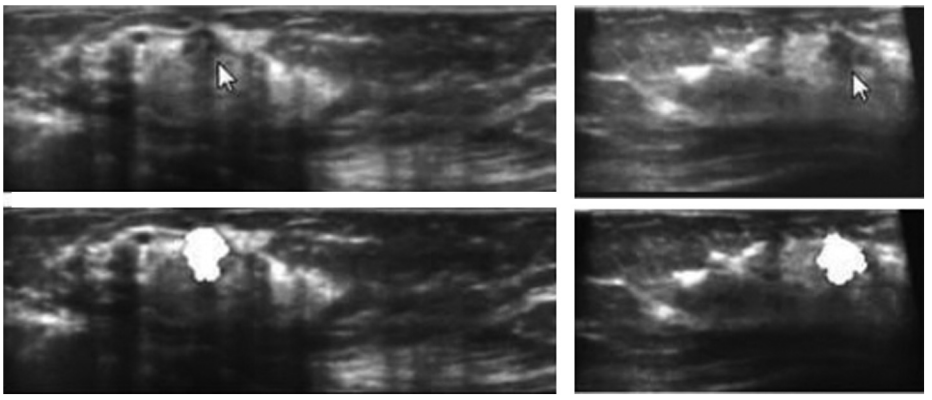


Fig. 4 Transversal slices with malignant lesions (top) pointed out by the arrows and their segmentations (bottom) indicated by the white masks.

result (with AUC value 0.80) indicated that the texture features represented by LBP are useful for lesion classification.

The main idea of the LBP algorithm is to count the occurrences of certain patterns, which are represented by binary numbers. For a local center point in a 2-D image, a pattern can be defined by thresholding its intensity with the intensity of its neighborhood. This process resulted in eight (take eight neighbors, for example) binary numbers, which are called LBPs. Considering possible rotations of the neighbors surrounding the local center point, there are only 36 unique patterns for eight-neighborhood LBP features. The normalized occurrences (that constitute a histogram) of these unique binary patterns are used as texture features. 108 LBP features were obtained from the three orthogonal planes.

2.3.3 *Gray level co-occurrence matrix*

GLCM has been widely used for image texture analysis. Alvarenga et al.[27] showed that the contrast representing spatial and gray level differences obtained from GLCM is an effective parameter in discriminating malignant from benign masses in breast US images. Recently, Gómez et al.[18] investigated the performance of using co-occurrence statistics in a CAD system to classify malignant and benign lesions and they obtained an AUC value of 0.87.

GLCM has been defined as a positive $N \times N$ matrix, where N is the number of gray levels in an image.[28] The elements in the matrix are generated by counting the number of occurrences of voxel pairs (with gray levels i and j) given a certain distance and direction. The minimum distance can be set to one voxel and the maximum distance can be set to the half of the image size. The following four directions are commonly used in the literature:[29–31] horizontal, vertical, left and right diagonals. Haralick[25] has extracted many statistical features known as Haralick texture features based on symmetric and normalize GLCM. There are six popularly used Haralick texture features: energy, entropy, inverse difference moment, inertia, cluster shade, and cluster prominence. In this study, 24 2-D GLCM texture features were generated corresponding to four directions and six types of Haralick texture features described above. 72 features were extracted from the three orthogonal views.

2.3.4 *Gabor filters*

Gabor filter, named after Dennis Gabor,[32] is a linear filter used for image processing. In the spatial domain, a 2-D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. There are successful cases that the Gabor filters have been used as texture descriptors in US images.[33–35] Chen et al.[35] investigated the diagnostic performance of texture variance in 3-D US images and concluded that the texture extraction with Gabor filter is more accurate than auto-correlation.

Gabor features can be represented by Gabor filter responses generated by the complex-conjugation between a set of Gabor kernels and an image. Each Gabor kernel is a product of a Gaussian envelope and a complex plane wave. Different frequencies and orientations constitute a diversity of complex sinusoidal plane waves. In our study, we investigated a group of Gabor filters specified by the combination of five scales of frequencies and eight orientations. Figure 5 illustrates a series of Gabor kernels used in our study. We used the mean and the standard deviation of the magnitude outputs of the Gabor-filtered image as the texture features. In total, 240 Gabor features were generated and integrated to our existing CAD system.

2.4 *Classification Scheme*

The number of features generated based on the texture descriptors is very large compared with the existing set of 11 features. In total, 420 texture features were generated, including 108 LBP features, 72 GLCM features, and 240 Gabor features. To deal with this large amount of features, we trained a separate support vector machine (SVM) classifier for each group of texture features, combining the information of all features in a group into a single-likelihood value. We used an SVM with a radial basis function kernel and optimized the cost and gamma parameters in a nested cross-validation loop to prevent bias. We refer to

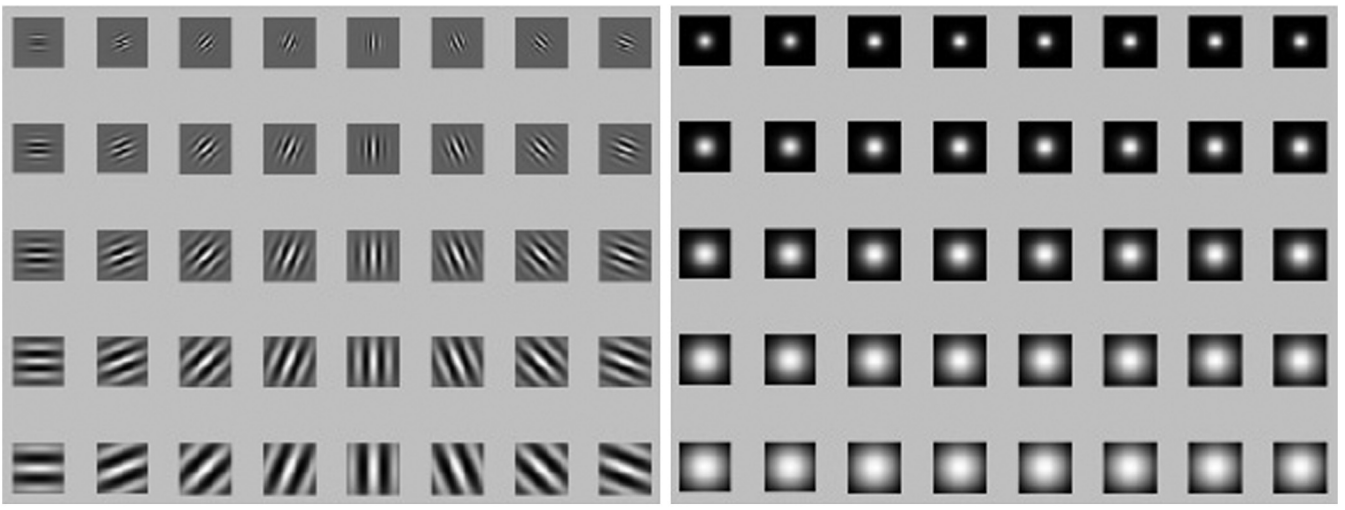


Fig. 5 Demonstrations of Gabor filters: the real part of the used 40 Gabor filters (a) and the magnitude output (b).

these classifiers, merging the features of a certain group to one likelihood value, as FOCs.

After the use of FOC for each of the three types of texture features, every feature group is represented by a single-likelihood value. By adding the existing 11 features, 14 features were obtained for each lesion. The final classification was done in a cross-validation manner, and it should be noted that to avoid bias the FOCs were optimized in a nested manner for each training-fold of this cross-validation loop. This implies that the FOCs themselves never get to see any of the data in each test-fold of the outer cross-validation loop.

We did not perform feature selection in any experiments in this work. For each type of texture features, we used all available features to obtain malignancy likelihood as a new feature.

2.5 *Evaluation*

To investigate the benefits of incorporating texture features, the classification results were compared with and without texture features. The discriminative performance was evaluated by computing the AUC. We compared the different texture features contributions to the CAD system by generating AUC values after adding each FOC.

2.5.1 *p-Value computations and Bonferroni corrections*

The significance of the diagnosis difference before and after utilizing different texture features was reflected by p-value. In this experiment, we used bootstrapping[36] to compute a p-value with 1000 bootstrap replications. We performed replacement sampling from the original cases and it should be noted that the total number of one set of cases equals to the amount of cases in the original set. For each new set, the AUC values were calculated for all different experiments. The difference in AUC values was generated for each of the 1000 new sets and p-values were defined as the fraction of the differences in AUC values that are negative or zero. In our experiment, we proposed four main hypotheses to predict the CAD classification performance changes before and after adding different texture features, which are base features (BF) + GLCM classifier, BF + LBP classifier, BF + Gabor classifier, and BF + FOCs. The more tests we perform on a set of data, the more likely we are to reject the null hypothesis when it is true, which could lead to Type I error, also known as false-positive error.[37] To correct this, we performed Bonferroni corrections[37] in such a way that we measured each significance level by comparing each p-value with 0.0125 (0.05/4) instead of 0.05 by default.

Table 1 Classification performance of computer-aided diagnosis system using different sets of features.

Feature(s)	Number of features	AUC	Std. Dev
Base features (BF)	11	0.90	0.01
BF + Gabor classifier	12	0.90	0.01
BF + LBP classifier	12	0.91	0.01
BF + GLCM classifier	12	0.91	0.02
BF + FOCs	14	0.91	0.01

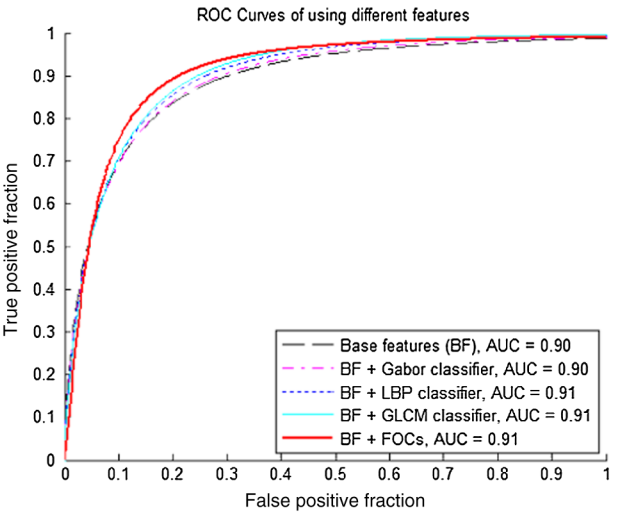


Fig. 6 Receiver operating characteristic (ROC) curves of using different features. The AUC values of using base features (BF), BF + Gabor classifier, BF + LBP classifier, BF + GLCM classifier, and BF + feature-oriented classifiers (FOCs) are 0.90, 0.90, 0.91, 0.91, and 0.91, respectively.

3 Results

Table 1 shows AUC values using different types of features extracted from the ABUS dataset. By combining the features in sets and training separate classifiers for each of these sets and subsequently calculating the likelihoods from these classifiers as new features in our classification scheme, we obtained one of the best results (0.91). The AUC values of using base features (BF), BF + Gabor classifier, BF + LBP classifier, BF + GLCM classifier, and BF + FOCs are 0.90, 0.90, 0.91, 0.91, and 0.91, respectively. The ROC curves of the CAD system before and after adding texture features are shown in Fig. 6. The AUC value was 0.90 using the existing features, whereas the AUC value was increased to 0.91 after adding texture features with our FOCs scheme. The significance analysis test showed ($p < 0.001$, after Bonferroni corrections, each p-value is smaller than 0.05/4) that the difference is statistically significant. Classification performance of using different texture features is shown in Table 2. The AUC values of using Gabor, LBP, and GLCM texture features are 0.82, 0.81, and 0.85, respectively.

With respect to LBP, we found out that the best result was achieved when the distance between the neighborhood and the center point was set to 2 voxels (1.2 mm). For GLCM features, 8 voxels (4.8 mm) gave the best result. (We did not perform significance tests among all the distances. We have compared the result generated by all the possible distances and the result from one single distance. One single distance performed better. We also tried Adaboost classifier on the whole features that were

Table 2 Classification performance of each type of texture features.

Feature(s)	Number of features	AUC	Std. Dev
Gabor features	240	0.82	0.02
LBP features	108	0.81	0.02
GLCM features	72	0.85	0.02

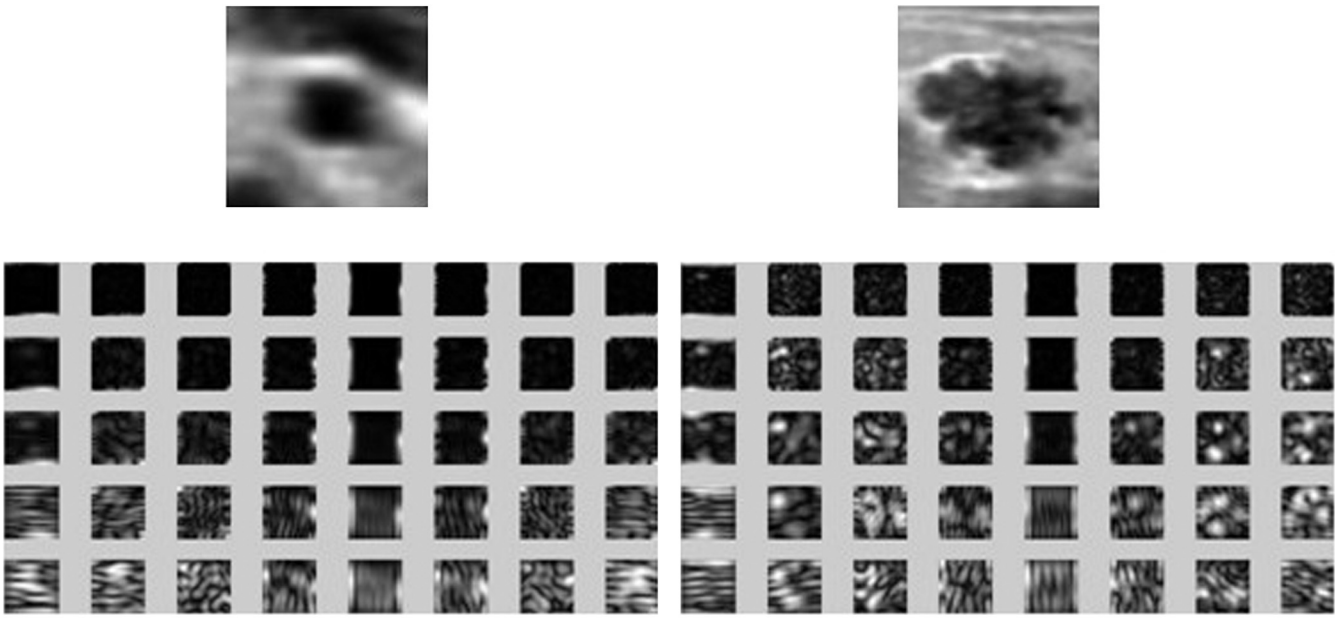


Fig. 7 An axial section through a benign (a) and a malignant (b) lesion together with the magnitude outputs of 40 Gabor filters for the respective lesions.

Table 3 Mean of the magnitude output generated from selected benign and malignant samples (benign/malignant) using Gabor filter banks shown in Fig. 5. O1-O8 represent eight orientations and f1-f5 represent five frequencies.

	O1	O2	O3	O4	O5	O6	O7	O8
f1	899.41/500.04	359.62/733.20	697.50/554.85	160.79/298.74	40.78/84.43	604.30/256.26	380.16/517.10	490.99/576.39
f2	160.19/299.31	40.74/84.44	293.36/166.60	416.74/247.13	487.28/584.05	159.79/298.45	40.72/84.36	523.75/173.66
f3	345.30/476.55	484.78/538.66	159.76/295.82	40.72/84.22	674.58/369.84	320.91/667.99	488.92/502.71	160.19/293.80
f4	40.75/84.12	547.91/190.00	450.11/503.72	501.06/527.01	160.88/294.34	40.78/84.13	321.36/166.12	523.75/173.66
f5	568.30/268.41	510.13/566.15	161.37/296.33	40.81/84.22	817.77/286.04	476.12/545.90	506.99/564.25	161.31/297.83

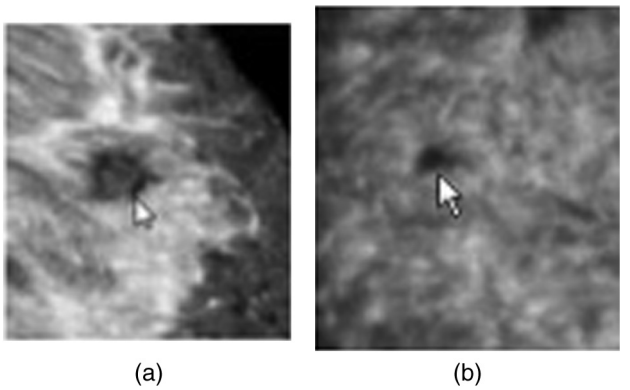
Table 4 Standard deviation of the magnitude output generated from selected benign and malignant samples (benign/malignant) using Gabor filter banks shown in Fig. 5. O1-O8 represent eight orientations and f1-f5 represent five frequencies.

	O1	O2	O3	O4	O5	O6	O7	O8
f1	222.13/322.68	33.57/225.55	9.99/213.99	0.79/71.53	0.05/24.14	104.26/115.07	23.66/164.53	9.85/156.43
f2	0.79/72.26	0.05/24.22	110.71/123.33	24.95/123.12	9.78/86.34	0.79/73.09	0.05/24.30	109.16/128.41
f3	42.78/116.08	9.74/173.73	0.79/74.31	0.05/24.32	202.26/352.47	59.26/206.39	9.84/227.54	0.79/74.45
f4	0.05/24.29	106.95/107.98	39.65/115.13	10.10/179.88	0.79/72.59	0.05/24.20	133.42/88.30	34.53/118.86
f5	10.28/98.02	0.79/70.52	0.05/24.11	168.38/139.37	25.18/148.15	10.20/164.24	0.79/70.43	0.05/24.09

based on all the possible distances, but the result was not better than using one single distance.) For Gabor filters, we used five frequency scales and eight orientations instead of specifying certain frequencies and orientations. Forty Gabor-filtered images are shown in Fig. 7, and selected Gabor features (mean and standard deviation) are listed in Tables 3 and 4.

4 Conclusion and Discussion

In this work, we extracted three sets of texture features based on LBP, GLCM, and Gabor filters for the classification of malignant and benign lesions in ABUS. Texture features were extracted from regions that were obtained by lesion-centered



(a) (b)

Fig. 8 Misclassified examples: cancer misclassified by the previous computer-aided diagnosis (CAD) but correctly by the current CAD (a) and cancer misclassified by the current CAD but correctly by the previous CAD (b).

segmentation algorithm from three orthogonal planes of the ABUS volumes. In the scheme, texture features were expected to capture the internal inhomogeneities of breast cancers in ABUS. To cope with the high dimensionality of the original texture features, features were grouped by type and a combination of FOCs was made. Hundreds of texture features were reduced to three features, each representing a set of texture features. A classification experiment was conducted on a dataset of 185 cancers and 239 benign lesions. SVM was adopted to train the classifiers with 10-fold cross validation. The AUC values were generated for each of the different texture feature sets and for the combination of the proposed features. The performance of the classification improved when texture features were used (AUC was increased from 0.90 to 0.91, $p < 0.001$).

There are some cases which were misclassified by the previous CAD system or the current CAD system. Figure 8(a) shows a cancer which was misclassified by the previous CAD but correctly classified by the current CAD. The echogenic texture pattern inside the cancer complements the new CAD system. Figure 8(b) shows a cancer which was correctly classified by the previous CAD but misclassified by the current CAD system. The reason might be that the homogeneous and hypoechonic texture confuses the new CAD system. It should be noted that the comparison between the two CAD systems should be determined by the AUC computed from overall cases in the dataset.

Our study highlighted the benefits gained by using texture features such as LBP, GLCM, and Gabor features. After integrating texture features into the existing CAD system appropriately, the classification performance was significantly improved. The results indicated that the benign and malignant breast masses have different texture structures. It was also proven that the scheme of FOCs can help to achieve feature dimension reduction.

Some studies[12,38] have shown that using CAD can improve the radiologist performance of distinguishing malignant from benign breast lesions in 3-D US images. The observer study[20] showed that the previous system performs as good as the best reader. As this system improves the previous one, we expect that this system can play a role in the clinical use.

There are limitations in our work. For example, we only extracted LBP features inside the lesion. In the future, we will investigate the value of extracting LBP features from the surroundings of the segmented lesion. Apart from the six commonly used Haralick texture features, we can try more features, such as the sum of squares and autocorrelation. The results from Chen et al.[39] showed that the classification performance of texture features computed from 3-D GLCM is better than that from 2-D GLCM when analyzing breast lesions in magnetic resonance images (MRIs). In future work, we will investigate whether we can benefit more from 3-D GLCM-based texture features for breast lesion classification in ABUS. There are also improvements that can be made for the Gabor features by selecting different parameters to construct the kernels. In this work, the features based on coronal plane were only extracted from the slice at the lesion center, which may not be an optimal way to utilize coronal information, as coronal planes have unique characteristics due to the ABUS imaging principle. One improvement in the future might be to extract features from multiple coronal planes to enrich diagnostic information. From the clinical point of view, we will further investigate to what extent our method can improve the radiologist's capability of differentiating between benign and malignant breast lesions.

Acknowledgments

The authors would like to thank André Grivegnée from Cancer Prevention and Screening Clinic, Jules Bordet Institute, Brussels, Belgium; László Tabár from the Department of Mammography, Falun Central Hospital, Falun, Sweden; and Matthieu Rutten and Mathijn de Jong from Department of Radiology, Jeroen Bosch Ziekenhuis, 's-Hertogenbosch, The Netherlands, for providing data for this study.

References

1. M. Grayson, "Breast cancer," *Nature* **485**(7400), S49 (2012).
2. L. Tabar et al., "Reduction in mortality from breast cancer after mass screening with mammography: randomised trial from the breast cancer screening working group of the swedish national board of health and welfare," *The Lancet* **325**(8433), 829–832 (1985).
3. L. Tabar et al., "Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening," *The Lancet* **361**(9367), 1405–1410 (2003).
4. Early Breast Cancer Trialists'Collaborative Group, "Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomised trials involving 31 000 recurrences and 24 000 deaths among 75 000 women," *The Lancet* **339**(8784), 1–15 (1992).
5. R. McWhirter, "V. the value of simple mastectomy and radiotherapy in the treatment of cancer of the breast" *Br. J. Radiol.* **21**(252), 599–610 (1948).
6. J. Mandelblatt et al., "The cost-effectiveness of screening mammography beyond age 65 yearsa systematic review for the us preventive services task force," *Ann. Intern. Med.* **139**(10), 835–842 (2003).
7. C. H. Ahern and Y. Shen, "Cost-effectiveness analysis of mammography and clinical breast examination strategies: a comparison with current guidelines," *Cancer Epidemiol. Biomarkers Prev.* **18**(3), 718–725 (2009).
8. L. J. Martin and N. Boyd, "Mammographic density. Potential mechanisms of breast cancer risk associated with mammographic density: hypotheses based on epidemiological evidence," *Breast Cancer Res* **10**(1), 1–14 (2008).
9. N. F. Boyd et al., "Heritability of mammographic density, a risk factor for breast cancer," *N. Engl. J. Med.* **347**(12), 886–894 (2002).
10. W. A. Berg et al., "Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer," *Jama* **299**(18), 2151–2163 (2008).
11. M. Nothacker et al., "Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. a systematic review," *Bmc Cancer* **9**(1), 335 (2009).

12. B. Sahiner et al., "Malignant and benign breast masses on 3d us volumetric images: effect of computer-aided diagnosis on radiologist accuracy," *Radiology* **242**(3), 716–724 (2007).
13. W. K. Moon et al., "Computer-aided classification of breast masses using speckle features of automated breast ultrasound images," *Med. Phys.* **39**(10), 6465–6473 (2012).
14. W. K. Moon et al., "Quantitative ultrasound analysis for classification of bi-rads category 3 breast masses," *J. Digital Imaging* **26**(6), 1091–1098 (2013).
15. M. B. Nagarajan et al., "Classification of small lesions in dynamic breast mri: eliminating the need for precise lesion segmentation through spatio-temporal analysis of contrast enhancement," *Mach. Vis. Appl.* **24** (7), 1371–1381 (2013).
16. M. Hussain, "Ensemble classifier for benign-malignant mass classification," *Int. J. Comput. Vis. Image Process.* **3**(1), 66–77 (2013).
17. L. Nanni, S. Brahnam, and A. Lumini, "A very high performing system to discriminate tissues in mammograms as benign and malignant," *Expert Syst. Appl.* **39**(2), 1968–1971 (2012).
18. W. Gómez, W. Pereira, and A. F. C. Infantosi, "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound," *IEEE Trans. Med. Imaging* **31**(10), 1889–1899 (2012).
19. M. M. Matsumoto, C. M. Sehgal, and J. K. Udupa, "Local binary pattern texture-based classification of solid masses in ultrasound breast images," *Proc. SPIE* **8320**, 83201H (2012).
20. T. Tan et al., "Computer-aided lesion diagnosis in automated 3-d breast ultrasound using coronal spiculation," *IEEE Trans. Med. Imaging* **31**(5), 1034–1042 (2012).
21. C.-H. Chen, L.-F. Pau, and P. S.-P. Wang, *Handbook of Pattern Recognition and Computer Vision*, World Scientific, Singapore (2010).
22. T. Tan et al., "Computer-aided detection of cancer in automated 3d breast ultrasound" *IEEE Trans. Med. Imag.* **32**(9), 1698–1706 (2013).
23. J. Wang, R. Engelmann, and Q. Li, "Segmentation of pulmonary nodules in three-dimensional ct images by use of a spiral-scanning technique," *Med. Phys.* **34**(12), 4678–4689 (2007).
24. T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.* **29**(1), 51–59 (1996).
25. R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE* **67**(5), 786–804 (1979).
26. J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *JOSA A* **2**(7), 1160–1169 (1985).
27. A. V. Alvarenga et al., "Complexity curve and grey level co-occurrence matrix in the texture evaluation of breast tumor on ultrasound images," *Med. Phys.* **34**(2), 379–387 (2007).
28. R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man Cybernet.* **3**(6), 610–621 (1973).
29. J. Zhang, G.-L. Li, and S.-W. He, "Texture-based image retrieval by edge detection matching glcm," in *Proc. of 10th IEEE International Conference on High Performance Computing and Communications*, Dalian, China, pp. 782–786 (2008).
30. J. Y. Tou, Y. H. Tay, and P. Y. Lau, "One-dimensional grey-level co-occurrence matrices for texture classification," in *IEEE Int. Symp. on Information Technology*, Kuala Lumpur, Malaysia, Vol. 3, pp. 1–6 (2008).
31. S. C. Liew et al., "Texture analysis of sar images," in *IEEE 'Quantitative Remote Sensing for Science and Applications', Int. Geoscience and Remote Sensing Symp*, Firenze, Italy, Vol. 2, pp. 1412–1414 (1995).
32. D. Gabor, "Theory of communication. Part 1: The analysis of information," *Electr. Eng.-Part III: J. Inst. Radio Commun. Eng.* **93**(26), 429–441 (1946).
33. M. Mohamed et al., "Prostate cancer diagnosis based on gabor filter texture segmentation of ultrasound image," in *IEEE Canadian Conf. on Electrical and Computer Engineering*, Montreal, Quebec, Canada, Vol. 3, pp. 1485–1488 (2003).
34. J. Stoitsis et al., "Texture characterization of carotid atherosclerotic plaque from b-mode ultrasound using gabor filters," in *IEEE Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, Minnesota, pp. 455–458 (2009).
35. W.-L. Chang, C.-C. Chang, and W.-M. Chen, "Computer-aided diagnosis system for variance estimation of 3d ultrasonography based on gabor filter," in *2010 International Conference on Biomedical Engineering and Computer Science (ICBECS)*, pp. 1–4 (2010).
36. B. Efron and B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, Vol. 38, SIAM, Philadelphia, PA (1982).
37. H. Abdi, "Bonferroni and Šidák corrections for multiple comparisons," in *Encyclopedia of Measurement and Statistics*, N. J. Salkind, Ed., pp. 103–107, Sage, Thousand Oaks, CA (2007).
38. T. Tan et al., "Evaluation of the effect of computer-aided classification ofbenign and malignant lesions onreader performance in automated three-dimensional breast ultrasound," *Acad. Radiol.* **20**(11), 1381–1388 (2013).
39. W. Chen et al., "Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images," *Magn. Reson. Med.* **58**(3), 562–571 (2007).

Haixia Liu received her master's degree in computer science from Uppsala University, Sweden, in 2013. She has been working as a visiting researcher at the Department of Radiology and Nuclear Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands, focusing on texture feature analysis of breast lesions in automated 3-D breast ultrasound. She is currently a PhD candidate in the Department of Computer Science, Nottingham University.

Tao Tan received his PhD degree at the Radboud University, the Netherlands in Feb 2014. Since Feb 2014, he has been working at the Radboud University Medical Center as a postdoc researcher focusing on computer-aided detection of breast cancers in medical images.

Jan van Zelst is a PhD candidate at the Department of Radiology at the Radboud University Nijmegen Medical Centre. He graduated in medicine at the Radboud University Nijmegen and he started his PhD project focusing on 3D automated breast ultrasound screening in the EU-funded ASSURE-project.

Ritse Mann is a breast radiologist at the Department of Radiology of the Radboud UMC. He is a researcher in breast imaging with a continuous focus on improving patient care. His main focus is on the evaluation, validation, and implementation of new technical and radiological techniques in clinical practice. He is a member of the scientific committee of the European Society of Breast Imaging and member of the executive board of the Society for Breast MRI.

Nico Karssemeijer joined the Department of Radiology of the Radboud University Nijmegen Medical Center in 1989, where he formed a research group in computer-aided diagnosis. He was involved in the development of the R2 ImageChecker, the most widely used CAD system to date, and is co-founder of Matakina, Ltd. (Wellington, New Zealand), a company that develops technology for quantitative mammography. He also has an appointment at Fraunhofer MEVIS in Bremen, Germany.

Bram Platel received his PhD degree at the Eindhoven University of Technology, the Netherlands, in 2007. Subsequently he worked at the Maastricht University Medical Center as an assistant professor until 2010, where he led the Image Guided Surgery group. From 2010 until 2012 he worked for Fraunhofer MEVIS as senior researcher. He now works at the Radboud UMC as a CAD researcher and manager of the EU ASSURE project on personalized breast cancer screening.