Ultrasonics 79 (2017) 68–80

Contents lists available at ScienceDirect

Ultrasonics

journal homepage: www.elsevier.com/locate/ultras

Breast cancer detection in automated 3D breast ultrasound using iso-contours and cascaded RUSBoosts

Ehsan Kozegar [a], Mohsen Soryani [a], Hamid Behnam [b], Masoumeh Salamati [c], Tao Tan [d,*]

[a] School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran
[b] School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran
[c] Department of Reproductive Imaging, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran
[d] Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen 6525 GA, The Netherlands

ARTICLE INFO

Article history:
Received 25 August 2016
Received in revised form 21 March 2017
Accepted 18 April 2017
Available online 20 April 2017

Keywords:
Automated breast ultrasound
Mass
Computer aided detection
Isocontours
Cascade classification

ABSTRACT

Automated 3D breast ultrasound (ABUS) is a new popular modality as an adjunct to mammography for detecting cancers in women with dense breasts. In this paper, a multi-stage computer aided detection system is proposed to detect cancers in ABUS images. In the first step, an efficient despeckling method called OBNLM is applied on the images to reduce speckle noise. Afterwards, a new algorithm based on isocontours is applied to detect initial candidates as the boundary of masses is hypo echoic. To reduce false generated isocontours, features such as hypoechoicity, roundness, area and contour strength are used. Consequently, the resulted candidates are further processed by a cascade classifier whose base classifiers are Random Under-Sampling Boosting (RUSBoost) that are introduced to deal with imbalanced datasets. Each base classifier is trained on a group of features like Gabor, LBP, GLCM and other features. Performance of the proposed system was evaluated using 104 volumes from 74 patients, including 112 malignant lesions. According to Free Response Operating Characteristic (FROC) analysis, the proposed system achieved the region-based sensitivity and case-based sensitivity of 68% and 76% at one false positive per image.

1. Introduction

Breast cancer is the most common cancer and the second leading cause of cancer death after lung cancer among women worldwide. Early detection is the most effective way to reduce the mortality. The best practical approach to detect breast cancer is breast imaging by which radiologists can detect cancer's symptoms in early stages [1].

Nowadays, many imaging modalities are being developed to find breast cancer among which mammography is the most common. However, it has some limitations such as having low sensitivity for women with dense breasts and exposure to X-ray. X-ray is an ionizing radiation which is not applied for women under 30 years old because exposure to X-ray radiation may increase the risk of incidence of breast cancer [2]. In addition, mammography may cause pain for women with sensitive breasts due to pressure on breast during imaging process. Moreover, mammography may result in many false positives such that 65–85 percent of cases are wrongly recalled to undergo biopsy surgery [3]. Biopsies are usually expensive and stressful. Therefore, using mammography by itself is not the best approach for young women. Hence, it is recommended to use an alternative modality besides mammography to increase the detection rate of radiologists. A common approach is to use handheld ultrasound imaging by which small tumors that are missed in mammograms due to existence of dense tissues can be detected. Also, handheld ultrasound is a modality for high risk women who are not supposed to be exposed to X-ray radiation. Although handheld sonography has some advantages compared to mammography, it has some limitations. For example, an expert radiologist is needed for image acquisition and it is operator dependent. That is, handheld ultrasound images cannot be reproduced and even expert radiologists produce different images for same cases [4]. Reproducibility is an important feature when radiologists need to track mass changes in certain time intervals. In addition, this type of screening is expensive because a radiologist instead of a technician needs to gather images. Moreover, handheld ultrasound is time consuming since whole region of breast should be scanned by radiologists. Simultaneous gathering and interpreting images make radiologists exhausted and may increase the missing rate of cancers [5].

* Corresponding author.
E-mail address: tao.tan911@gmail.com (T. Tan).

http://dx.doi.org/10.1016/j.ultras.2017.04.008
0041-624X/

The problems of handheld ultrasound imaging can be solved using automated breast ultrasound systems (ABUS) in which image acquisition is handled by a radiographer who can be a trained technician not an expert radiologist. During scanning, transducer automatically moves across or around the breast region and gathers 2D image slices that are stacked together to form a 3D volumetric image. Kelly et al. have shown that ABUS as an adjunct to mammography for dense breasts not only increases sensitivity but also decreases false positive rate [6]. The ABUS imaging modality is faster, reproducible and it is not expensive. However, the workload of radiologists dramatically increases because they need to interpret many slices for each of the volumes obtained from different views. Therefore, volume screening process will be time consuming and exhaustive for radiologists and may cause them to miss small lesions. When image resolution and number of slices in each volume are increased it is inevitable to use computer aided systems to help radiologists in finding masses. A computerized method can be designed to show suspicious regions to radiologists enabling them to perform the screening more precisely by concentrating on these regions.

Computer aided systems assist radiologists to find and diagnose abnormalities in images. In fact, these systems act as a second interpreter to detect disorders and they are not designed to replace radiologists by computers. Computer aided systems are generally categorized into Computer Aided Detection systems (CADe) for detecting disorders and Computer Aided Diagnosis systems (CADx) for diagnosing them. Contrary to mammography, development of computer aided systems for 3D ABUS images is in preliminary stages and designing a CADe system for this type of images is more challenging and more required. We propose a CADe system for detecting cancers in 3D ABUS images.

Lo et al. used sigma filter to reduce speckle noise of images and segment them by a special version of watershed transform called toboggan watershed [7]. Watershed transform and its variants suffer from a well-known phenomenon called over-segmentation especially for images with many local minima just like ABUS images. Moon et al. proposed a blob detection algorithm to find the location of masses which usually look like dark blobs in ABUS images [8]. They enhanced the blob like structures in images based on eigenvalues of the Hessian matrix. Based on their method, blob structures that exist in one slice are enhanced but most of these structures are false positives because there are many dark blob regions like fat tissues that emerge in one slice and disappear in subsequent slices. In fact, masses are structures that appear in at least two subsequent slices. The same enhancement is occurred for line structures that should not be enhanced. Moreover, some malignant masses have hyper-echoic properties that means inside of malignant masses can be bright not dark. Hence, their algorithm fails to enhance such malignant masses. Ikedo et al. considered a new hypothesis that a normal case consists of semi horizontal edges but an abnormal case consists of some vertical edges on the boundary of the mass. Canny edge detector followed by a thinning method on each slice of the volume was applied to detect edges [9]. All edge based methods like Canny suffer from low resolution and speckle noise of ultrasonic images which result in discontinuous edges. This discontinuity can undermine their strong presumptions about edge characteristics of normal and abnormal cases. Moreover, the angle of semi-horizontal and semi-vertical edges seems to be very tight in their study. Their database includes 36 masses (16 malignant masses, 5 fibro adenomas, and 15 cysts). Drukker et al. studied women whose cancer was not observable in mammograms, those with BIRADS category 3 or 4 for whom sonography had been prescribed after mammography [4]. They combined a k-means clustering followed by an adaptive thresholding. In their method, the whole image is clustered into 3 region categories: bright regions (Parenchyma), dark regions (nipple) and candidate regions. The most drawback of this method is the use of k-means itself because it considers only gray level of the pixels not the spatial information. Moreover, k-means is sensitive to centroid initialization such that results may vary for every run. Tan et al. applied a two-stage detection scheme to detect breast cancers in ABUS. In the first stage a voxel classification is implemented to determine the cancer candidates and in the second stage, a region classification is used to reduce false positives [10]. The first stage was later improved by incorporating Haar-like features [11]. The disadvantage of these two methods is that the initial voxel candidate detection stage tends to miss over 10% cancer cases.

In this paper a computer aided detection system for detecting cancers in 3D ABUS images is proposed. In this system, preprocessing operations including speckle noise reduction and unnecessary slide removal are applied at first to the images. Then, suspicious candidate regions are extracted using a novel algorithm based on isocontours. Domain specific knowledge is used in the next module to discard false positives. Afterwards, the centers of the remaining regions are calculated as markers (seeds). The sensitivity of this stage is high and those seeds which are considered as "hit" are close to the center of masses. Because of this closeness, it is promising to use a voxel classifier to reduce too many seed points extracted in previous step. In this work, 432 voxel features are fed to a new cascade learning module based on RUSBoost classifier to filter out false positives even more.

2. Materials and methods

2.1. Database

The automated 3-D breast ultrasound images used in this paper are cases obtained in routine clinical care or screening from the Radboud University Nijmegen Medical Centre (Nijmegen, The Netherlands), the Jeroen Bosch Ziekenhuis (Den Bosch, The Netherlands), the Falun Central Hospital (Falun, Sweden), and the Jules Bordet Institute (Brussels, Belgium). The used images were generated by two types of ABUS systems: the SomoVu automated 3-D breast ultrasound system developed by U-systems (Sunnyvale, CA, USA) and the ACUSON S2000 automated breast volume scanning system developed by Siemens (Erlangen, Germany). In our dataset, images from the device by U-systems have a maximum size of 14.6 cm $*$ 16.8 cm on the coronal plane and a maximum depth of 4.86 cm while images from the device by Siemens have a maximum size of 15.4 cm $*$ 16.8 cm on the coronal plane and a maximum depth of 6 cm. The transducer of the U-systems device model used in this study had a fixed frequency of 8.0 or 10.0 MHz while the frequency of the transducer by Siemens is variable between 5.0 and 14.0 MHz, and can be adjusted according to the breast size. Each 3-D volumetric view by the device from U-systems was generated with a minimal voxel size of 0.29 mm (along the transducer) by 0.13 mm (in depth direction) by 0.6 mm (along the sweeping direction) while images by the device from Siemens have a minimum voxel size of 0.21 mm $*$ 0.07 mm $*$ 0.52 mm. The bit depth of ABUS images is 8 and to make processing easier and faster, the images were downsampled to volume with an isotropic voxel size of 0.6 mm using bilinear interpolation. The dataset used in this study consists of 104 images (views) from 74 patients; including 112 cancerous and 33 benign lesions (10 cysts and 23 solid lesions). For patients with lesions, only those views in which the lesions were visible were included in this study.

2.2. Proposed computer aided detection system

The proposed method for cancer detection in 3D ABUS images follows the block diagram shown in Fig. 1. This algorithm was

implemented on a laptop with Intel (R) Core i7 CPU, 2.6 GHz, 8 GB RAM, with Matlab R2015a on Windows 8.1. The modules of Fig. 1 are explained as follows.

2.2.1. Preprocessing

The main goal of the preprocessing step is speckle noise reduction of images. Various state of the art noise reduction methods such as Total Variation [12], K-SVD [13], Block Matching and 3D Filtering (BM3D) [14] and Bayes-Shrink [15] were assessed but the results were not satisfying. All mentioned filters obtained less sensitivity and more false positives at their best parameter fitting. The reason could be that, these methods are not designed specifically for despeckling. It is notable that even special despeckling methods like Speckle Reducing Anisotropic Diffusion (SRAD) [16] and Lee filter [17] are not applicable for ABUS images because they need a predefined window to be initialized by users. This window should be placed on a region whose variance corresponds to noise variance (i.e. a constant region in reality). These methods highly depend on the window while it is not known for the users where the constant region is. This crucial limitation makes these methods inappropriate in this application. In this paper, an effective approach called Optimized Bayesian Non Local Mean (OBNLM) [18] is used which can significantly remove the speckle noise without blurring edges as shown in Fig. 2. In addition to despeckling, the first seven and the last seven coronal slices are ignored because they are related to the skin and beyond the ribs in which no masses would be emerged.

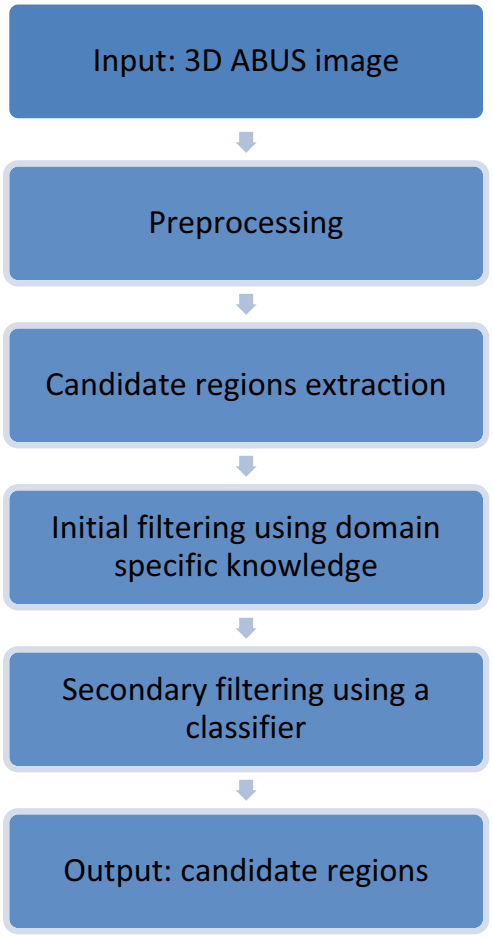

Fig. 1. Block diagram of the proposed mass detection algorithm.

2.2.2. Initial candidate region extraction

In this section, the main finding of this paper is described in details. We found that although inside of some masses such as malignant masses are neither hypo echoic nor homogenous, both of these properties increase near the mass boundary. That is, a contour called isocontour usually exists around some subsequent slices of the mass on which pixels have almost similar intensity values (Fig. 3). Isocontours have a parameter called isovalue which determines the intensity of pixels on the contours to be extracted on each slice. To extract the isocontour geometry Marching Squares and linear interpolation are used and followed by sorting the contour geometry into separate sorted contours [19].

After isocontours extraction, a number of contours are obtained which can be emerged anywhere with any shape. An example of applying this algorithm with different isovalues is shown in Fig. 4. As indicated in this figure, if isovalues are too low or too high then the resulted contours are not satisfying. Moreover, morphology of many regions is far different from the morphology of masses in terms of size and circularity. Accordingly, those regions are false positives which can be removed. As mentioned above, we know that there is an isocontour near the mass boundary but the question is what the isovalue should be. To answer this, we have to try different iso-values because an isocontour that emerges around a mass is specific to that mass not all masses. However, we need to decrease the number of isovalues to reduce false positives without

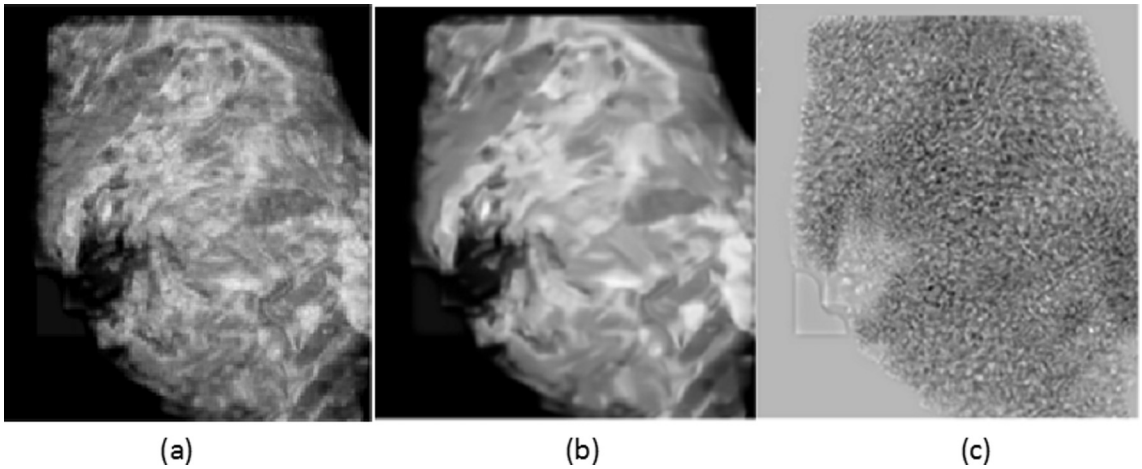

Fig. 2. (a) Original noisy image. (b) Despckled image after applying OBNLM filtering. (c) Speckle pattern.

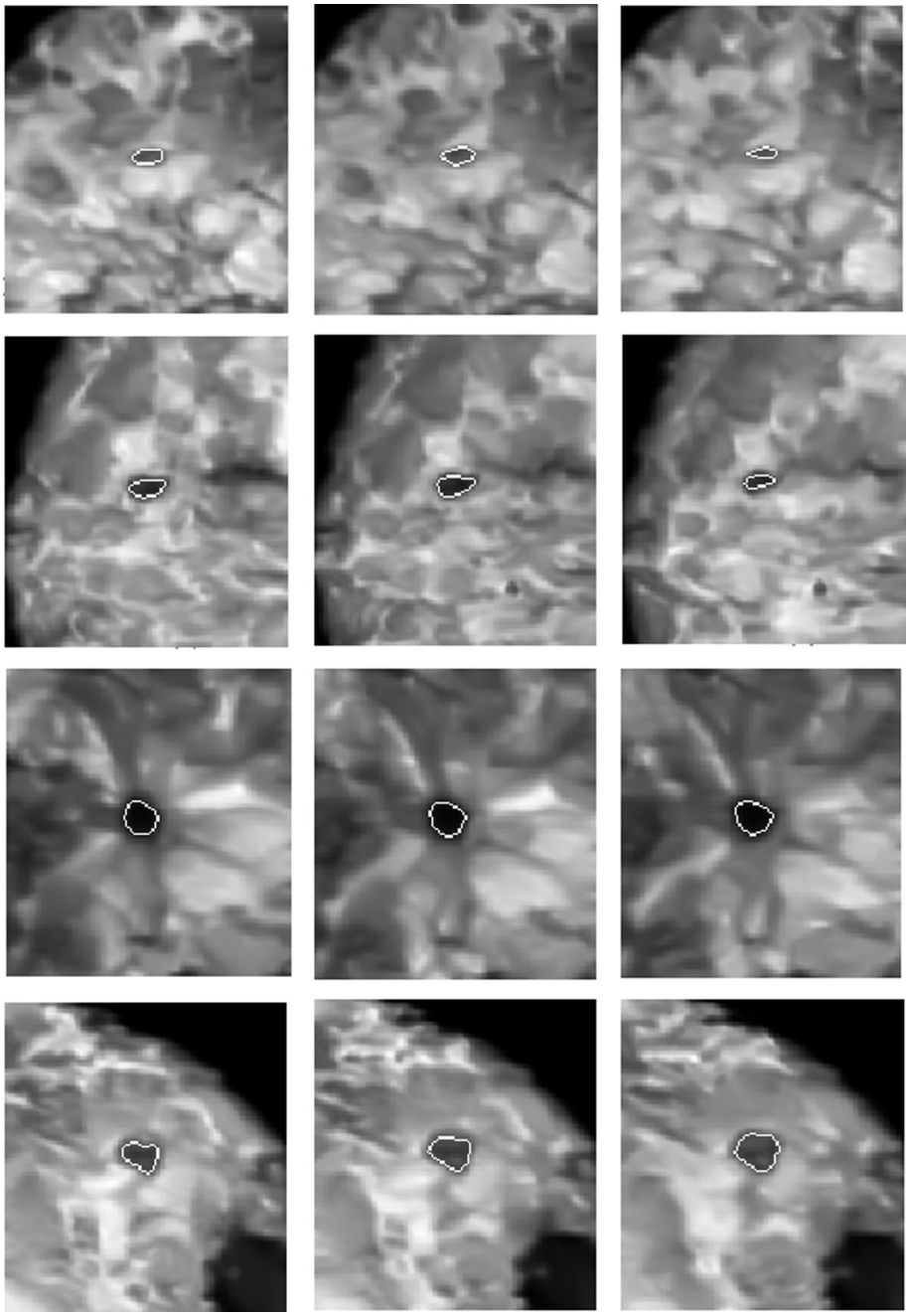

Fig. 3. Isocontours around three subsequent slices. Each row corresponds to a different cancerous mass.

loss of sensitivity. Since close isovalues lead to similar results (as indicated in Fig. 5), a good solution to reduce the number of testing isovalues can be quantization of gray levels by which close intensity values are set to the same value. Therefore, mass detection performance can be improved by quantizing ABUS images. We decreased the number of quantization levels from 256 to 128, 64, 32 and 16. While decreasing the quantization levels down to 32 levels, the false contours were reduced without loss of sensitivity. For 16 levels the sensitivity was degraded. Subsequently, we tested the median of 32 and 16 to see the impact of 24 quantization levels. For 24 levels, false positive rate was decreased and the true positive rate was preserved. Therefore, in the presented work, the images are quantized to 24 levels with values 1/24 ... 23/24, 1. Consequently, only contours with isovalues between *LowerIsoThreshold* and *UpperIsoThreshold* are included and the others are excluded because closed regions that are near the boundary of masses are not very hypo-echoic or hyper-echoic. These two thresholds are set to 5/24 and 20/24, respectively.

After quantization and isocontour extraction, suspicious markers are derived according to the algorithm shown in Fig. 6.

The algorithm consists of two nested loops. The outer loop depends on the number of isovalues and the inner loop is related to the number of slices. Let isovalue be *LowerIsoThreshold* at first. The inner loop starts from slice 1 and then all isocontours with isovalue equals to *LowerIsoThreshold* are obtained on this slice. Afterwards, many irrelevant contours can be discarded using simple domain specific knowledge. For example, contours whose roundness is far from a circle can be ignored because the shape of a mass on each slice is almost circular. In addition, a size filter can be used to remove small false positive regions whose diameter is less than

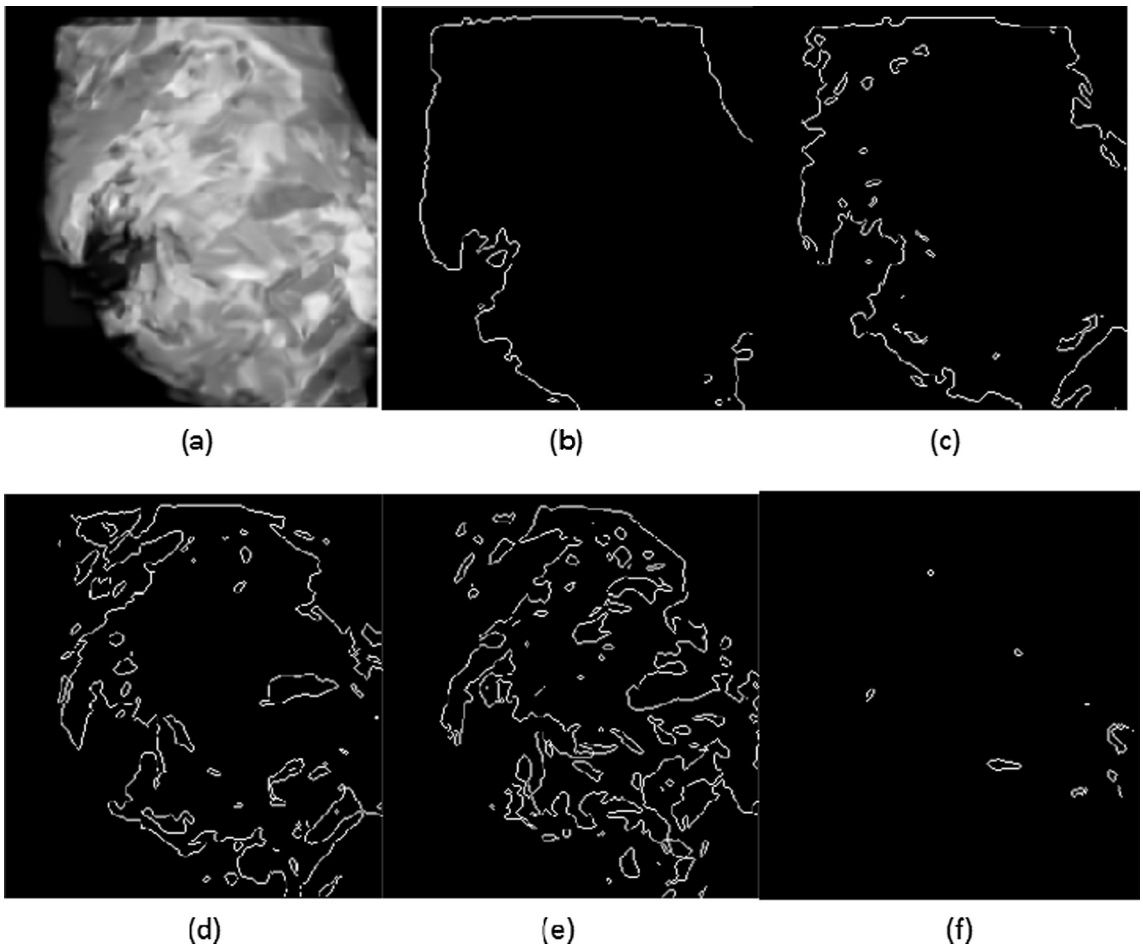

Fig. 4. (a) Original image after despeckling (b–f) Contours obtained by applying isovalues 0.1, 0.4, 0.5, 0.6 and 0.9, respectively.

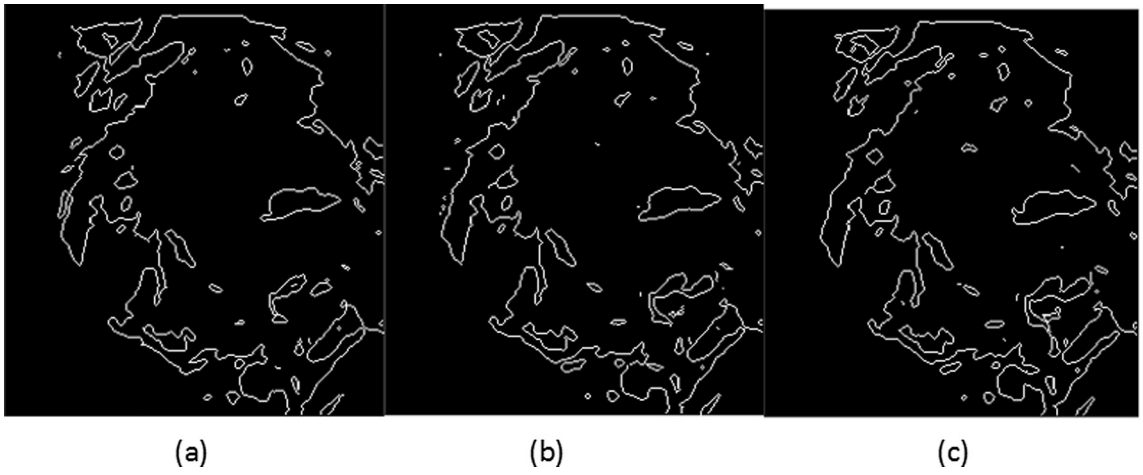

Fig. 5. (a–c) Contours obtained by applying isovalues 0.5, 0.51 and 0.52, respective.

LowerDiameterThreshold and large false positives whose diameter is bigger than an *UpperDiameterThreshold*. These two thresholds are set to 2 mm and 50 mm, respectively. It is notable that the thresholds of the algorithm are tuned by evaluating just a few images so that the system's generalization is not degraded. The algorithm uses another domain specific knowledge to further remove false positives. Accordingly, the algorithm discards contours whose inside is brighter than its outside; a contour is excluded if the average intensity of its interior pixels is bigger than isovalue. In the following, one of the most important modules called weak contour filtering is described in details.

The main drawback of using isocontours is that some isocontours may emerge in smooth areas where there is no structure and no masses. Moreover, we observed that the isocontours derived for the masses are not exactly located and must be tuned to match the mass boundaries. In order to push the isocontours

```
input: 3D ABUS image
output: centers of mass candidates
P:= 3D ABUS image which is preprocessed
Q:= Preprocessed image which is quantized
For iso-value= LowerIsoThreshold to UpperIsoThreshold
 begin
   For i:= 1 to Number_Of_Slices
        T(i)←extract isocontours with value equals to iso-value on slice i^th of Q
        T(i)← apply circularity filter on T(i) to remove contours that are less circular
        T(i)← apply area filter on T(i) to discard regions that are too small or too big
        T(i)← apply brightness filter on T(i) to remove bright regions
        T(i)← apply strength filter on T(i) to discard contours on which average gradient of pixels is too small
    End
    BW←Apply hole filling followed by 3D connected component labeling to integrate contours in volume T
    BW2←Discard regions which are emerged in only one slice
    centers← get centroids of regions in volume BW2 and append them to the previous centers
 End
Output=centers;
```

Fig. 6. Pseudo code of the initial mass candidate selection.

toward the boundaries where the gradients are high, an active contour model called Distance Regularized Level Set Evolution (DRLSE) [20] with a little change is applied. The regularization term of DRLSE model is ignored in our implementation to make the contours to approach the boundaries as close as possible. After fine tuning of isocontours, to remove false positives, mean gradient of all pixels on each isocontour is considered as a measure of contour strength. Then weak contour filtering can drastically decrease the number of false positives without loss of sensitivity. The threshold value to discard weak contours is set to 0.1. Therefore, if mean gradient magnitude of pixels of the contours is below 0.1 then those contours are removed. As shown in Fig. 7, filtering out false positives using domain specific knowledge can significantly improve the effectiveness of the mass detection algorithm. For the whole images of our database, the number of false positives is substantially reduced (more than 15 times) without loss of sensitivity.

By now, the sequence of modules in the inner loop has been described. Once the inner loop finishes, the binary slices are integrated by connected component labeling operation to form a 3D binary volume. Since a mass should be observed at least in two consequent slices, contours that are emerged in only one slice are excluded. This sequence is iterated for all isovalues.

After applying all isovalues and finishing the outer loop, the centers of the remaining regions are marked by a '+' sign. The number of markers in each mass varies regarding to the mass size and its observability such that for some big and observable masses the

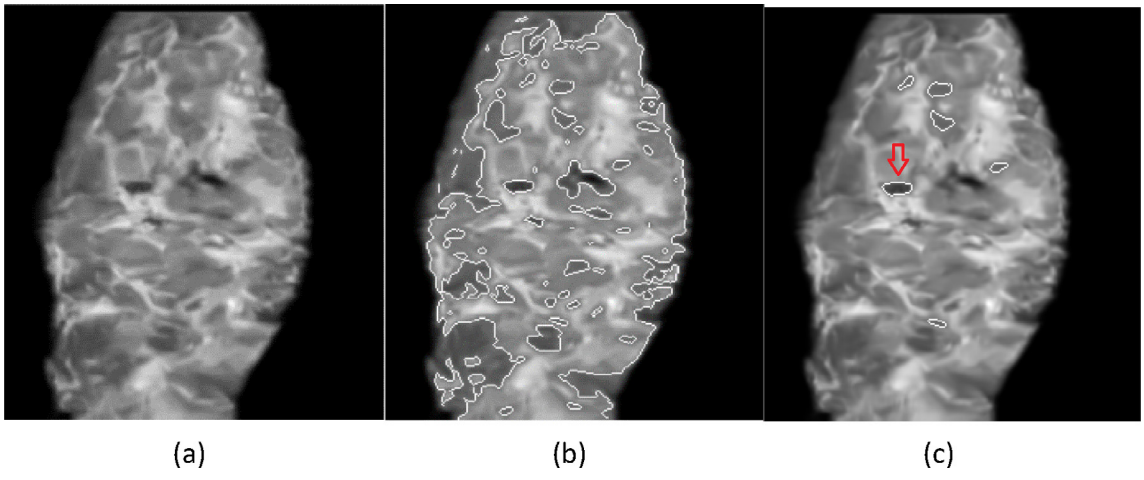

Fig. 7. (a) Original image. (b) Isocontours (c) the remaining isocontours after applying morphology, brightness and weak contour filtering, red arrow indicates the true mass. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

algorithm may find 15 markers. However, small and not much observable masses may contain only one marker. Thus, all markers have the same level of importance and should be considered by radiologists equivalently. An example of running the initial marker detection is shown in Fig. 8. The markers pointed by arrows indicate true positives and the others are false positives. It is observed that a lot of difficult false positives are still left and it is necessary to refine them. The most promising approach to this aim is to use a classifier as an intelligent filter.

2.2.3. Filtering markers using classification

As indicated in Fig. 1, the last module to filter false positives is a classifier that needs a training set to learn patterns of classes. To construct a training set for the classifier a feature vector is extracted corresponding to each marker and a class label is assigned to that vector. Based on our detection criteria [10], a marker is counted as a mass (label 1) if its distance from the ground truth is less than 1 cm otherwise it is a false positive (label 0). In this paper, 432 features are extracted for each marker, including

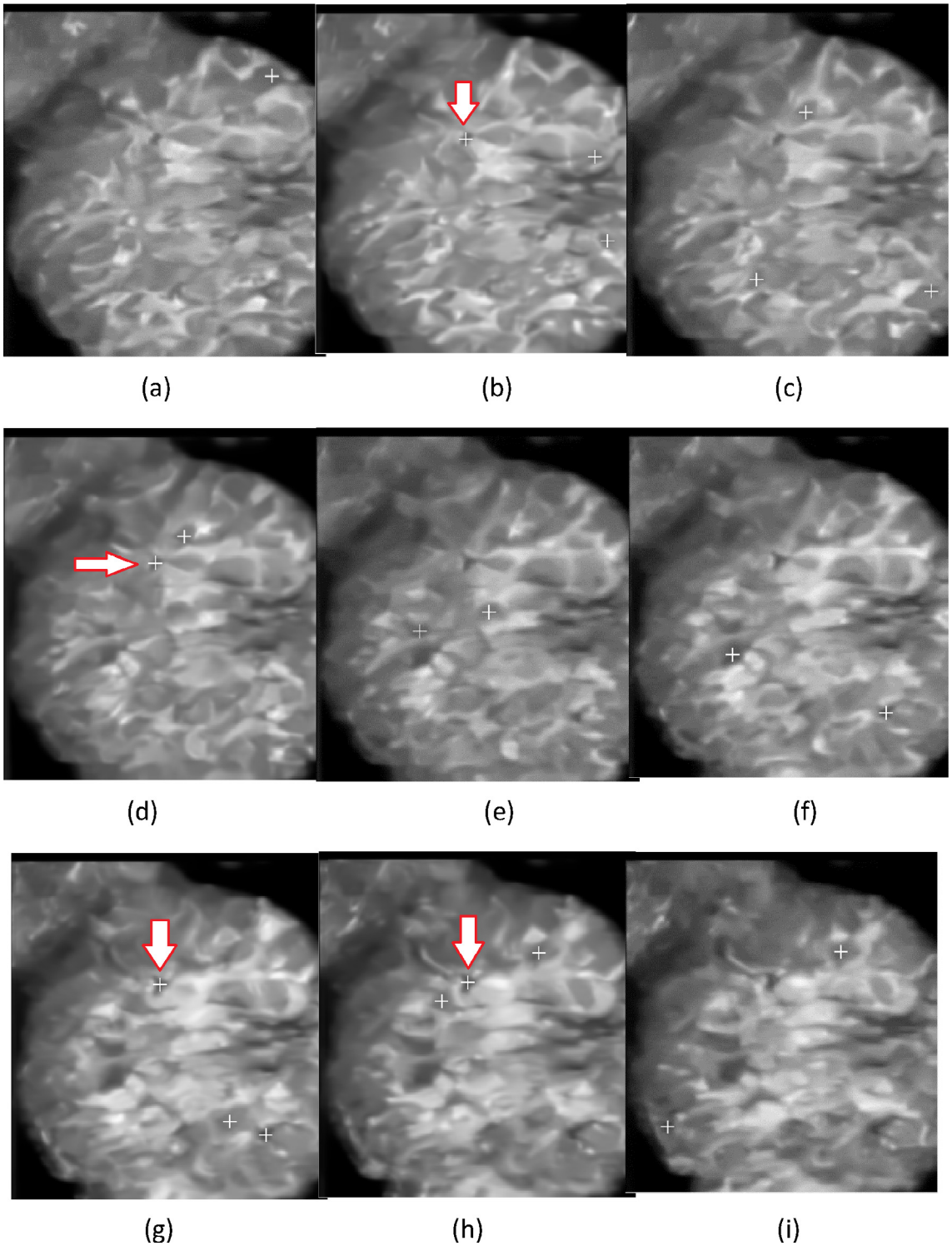

Fig. 8. Markers extracted for slices 34–43 of a sample volume. The arrows represent true positives.

Gabor (240 features), Gray Level Co-occurrence Matrix (GLCM) (72 features), Local Binary Patterns (LBP) (108 features) and basic features (12 features), all of them have been introduced in [25].

After feature extraction, a classifier should be trained to discriminate true positive and false positive markers. In the proposed method, a cascade classifier is used (Fig. 9) for reducing false positives.

After initial mass candidate detection, we faced with an imbalanced dataset because the proportion of true positive samples against false positives is 1 to 66. Therefore the class distributions are very different and this may degrade the classification performance. A method which can handle the problem of imbalanced training sets with numerous features is boosting algorithm. After assessment of various boosting classifiers such as AdaBoost [21], GentleBoost [22], RobustBoost [23] and Random Under-Sampling Boosting (RUSBoost) [24], we found that RUSBoost outperforms other boosting algorithms because it is inherently designed for imbalanced datasets. RUSBoost prevents the classifier to get biased toward the major class by using undersampling technique during the training phase. The training and testing algorithms of the cascade classifier is shown in Fig. 10.

2.2.3.1. Training. The main process of the proposed training phase is to tune a specific threshold for determining the class of samples in the first three RUSBoosts. In our method, RUSBoost returns two prediction scores for each instance that imply the support of belonging to each class. Higher prediction score means higher probability of belonging to the corresponding class. Let p0 be the prediction score of the false positive class and p1 the prediction score of the true positive class. Both prediction scores are positive values which can be scaled between 0 and 1 via division by sum of them.

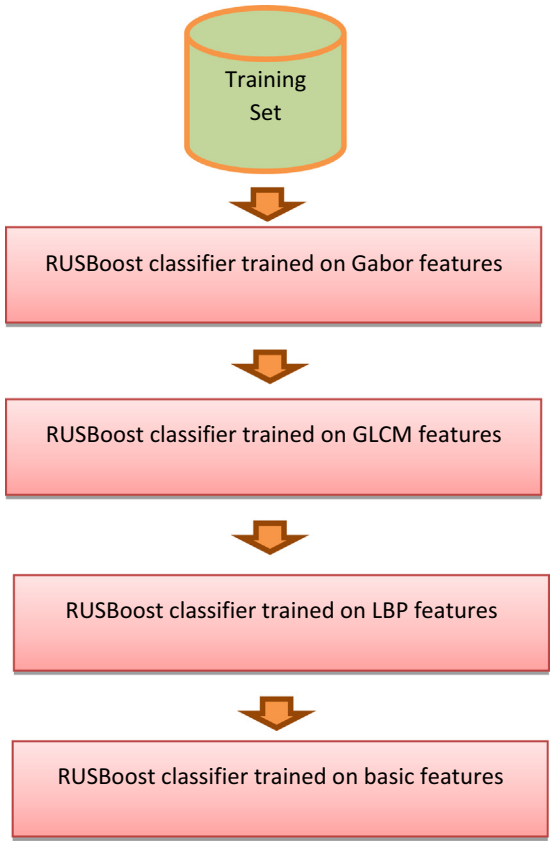

Fig. 9. The proposed cascade classifier block diagram.

Let th1 be the threshold value of the first RUSBoost. The thresholding operation is as follows. If p1 is greater than th1 then the related sample is classified as mass, otherwise it is a false positive. However, the main question is how to set the threshold value. To answer the question, leave one case out strategy is used in which the threshold is selected such that no true positive marker is wrongly classified as false positive (i.e. the false negative rate should be zero). At the beginning, the threshold is set to zero and after each iteration it is increased by 0.01. The loop iterates as long as false negative rate remains zero. The threshold value is the latest value before breaking the loop. Afterwards, all samples that are correctly classified as false positives are removed from the training set. This can reduce the number of false positives and moderate the imbalancity of the training set. The modified training set is then fed to the second classifier to further reduce the imbalancity of the training set. The procedure for threshold selection in the second and third classifiers is exactly similar to that of the first classifier but their input feature vectors are different. The first, second, and third classifiers use Gabor, GLCM and LBP features, respectively. The fourth RUSBoost classifier utilizes the most discriminative features known as basic features.

2.2.3.2. Testing. To determine the class label of a test marker, if it is classified as false positive by any of the first three classifiers by considering their thresholds then its label is false positive, otherwise the output of the forth classifier determines the final class label.

3. Results

The dataset used in this study consists of 104 images from 74 patients; including 112 malignant masses. In this paper we used patient based and region based analysis. With patient-based analysis the true positive fraction is computed as the fraction of cancerous patients in which at least one cancerous region is detected out of all cancerous regions. Using region-based analysis, all regions are treated independently because sometimes it could be that there are multiple cancers in one patient. Detecting each of them is also count as a true positive. To obtain unbiased detection results we used leave one patient out evaluation to measure the sensitivity and false positive rate of the proposed system and for statistical analysis we used bootstrapping. Each bootstrap contains 1000 instances that are obtained by sampling dataset with replacement.

The impact of speckle noise removal using OBNLM on the detection rate is shown in Tables 1 and 2. As shown in the tables, after this enhancement not only the case based sensitivity in the candidate detection stage increases but also the number of false positives per image decreases. To evaluate the significance of OBNLM the p-value is calculated using a two-sided Wilcoxon rank sum test because the standard deviation of OBNLM is zero (as shown in Table 2) and it does not follow a normal distribution. The resulted p-value for the significance test is less than 0.0001.

We have used free-response receiver operating characteristic (FROC) analysis to determine the performance of the system based on the suspiciousness level obtained for each voxel candidate (seed point). This suspiciousness is equal to the classifier's normalized output. As mentioned in the previous section RUSBoost can produce a probabilistic output which can be used for drawing the FROC curve. The performance of the proposed system is shown in Fig. 11 according to both patient based and region based analysis. As expected, the region based sensitivity is less than patient based sensitivity for a given false positive rate. Some operating points (i.e. sensitivity and False positive pairs) of the proposed classifier are (76%, 1), (80%, 2), and (94%, 9) for patient based analysis. For region based, the operating point are (68%, 1), (73%, 2) and (86%, 16). The

```
Training Phase:

Input1: TS= Training set including feature vectors of the cases
Input2: Index_Gbor= Position of Gabor features in TS
Input3: Index_GLCM= Position of GLCM features in TS
Input4: Index_LBP= Position of LBP features in TS
Input5: Index_Basic= Position of basic features in TS

Output1: Classifiers // 4 RUSBoost trained classifiers
Output2: Thresholds // An array containing 3 elements, each of them is related to the first three trained RUSBoost

Index=[index_Gabor, Index_GLCM, Index_LBP, Index_Basic] ;
// Index is an array containing 4 elements, each of them is a vector corresponding to the related feature //positions

N= number of training cases;
T=3; // number of RUSBoost classifiers in the proposed cascade classifier
Thresholds=[0,0,0]; // initialization of thresholds variable

For i=1 to T
  Begin
    [ P0 , P1, classifier(i) ] = predict("RUSBoost" , TS, Index(i)) ;
    // Train a RUSBoost classifier on TS by considering only feature positions of index(i) and evaluate
    // prediction scores for each feature vector in TS using leave-one-case-out strategy.

    Normalize(P0 , P1);
    // Converting prediction scores to likelihood scores. P0 and P1
    th=0;
    While (th<=1)
       Begin
          Label= (P1>=th)   // Label is a binary vector, each element corresponds to a feature vector
          If (there is any false negative in Label vector) then
             Begin
                Th=th-0.01;
                break the loop;
             End
          Else
             th=th+0.01;
       End // End of while loop

     Thresholds(i)=th;
     Update(TS , classifier(i) , Thresholds(i));
     //Remove correctly classified false positives from TS and then reassign the remaining to TS
  End // End of For loop
[ P0 , P1, classifier(i) ] = predict("RUSBoost" , TS, Index(4)) ;  // training the fourth RUSBoost based on basic features
////////////////////////////////////////////////////////////////////////////////////////////////
Testing Phase:
Input1: Classifiers  // 4 RUSBoost trained classifiers
Input2: Thresholds // An array containing 3 elements, each of them is related to a trained RUSBoost
Input3: Sample      // A feature vector to be tested
Output: Label       // A Boolean value ( false positive or true positive)
For i=1 to 3
 Begin
   [p0,p1]= test(classifier(i) , Sample)  // prediction scores of input sample is calculated using the i^th RusBoost
   If P1 < Thresholds(i) then
     Begin
       Label= False Positive;
       Exit; // exit the function
     End
 End  // End of the for loop

[p0,p1]= test(classifier(i) , Sample);
Label= (P1>P0);
```

Fig. 10. The proposed algorithm for training and testing phases.

Table 1
The impact of OBNLM filter on the detection rate of the preliminary stage based on leave one out evaluation.

	Sensitivity (%)	FP/Image
Initial stage (without despeckling)	85	80
Initial stage (after despeckling)	100	66

Table 2
The impact of OBNLM filter on the detection rate of the preliminary stage based on bootstrapping.

	Sensitivity (%)	FP/Image
Initial stage (without despeckling)	87 ± 3	77 ± 9
Initial stage (after despeckling)	100 ± 0	64 ± 7

impact of cascading classifiers is illustrated in Figs. 12 and 13. It improves both region based and patient based performance of the detection system. The performance comparison of RUSBoost against other boosting algorithms is shown in Table 3. As indicated in the table, RUSBoost reaches higher sensitivity at given false positives per image for both case-based and region-based analysis.

For statistical comparison of the results of the proposed cascade classifier with non-cascading we used bootstrapping. Each bootstrap contains 1000 instances that are obtained by sampling dataset with replacement. We used the difference of the mean sensitivity of both classifiers in the range between 0.1 FP/image and 10 FPs/image on a logarithmic scale. By using two-sided *t*-test, it has been shown that mean sensitivity of the cascade classifier is significantly greater than that of the classifier without cascading. P-value for this statistical comparison is <0.001.

To evaluate the computation time of the method we obtained the average time for different phases. The image processing part of the algorithm (despeckling and candidate region extraction) takes 18 min per image and the training and testing phases (for each round of leave one out) take 5 min and 1 s, respectively.

4. Conclusion and discussion

In this paper, an automatic computer aided detection system (CADe) has been developed to assist radiologists for finding cancers in automated 3D breast ultrasound (ABUS) images. The proposed system consists of two main steps: (1) mass candidate region extraction (2) false positive reduction using a classifier. Based on our experiments, it was found that at one false positive per image the region-based sensitivity and case-based sensitivity were 68% and 76%, respectively. Moreover, this system achieved case-based sensitivity of 90% at 9 false positive detected regions per image that proves this system is promising for radiologists who intend to have a practical system with high sensitivity.

In the initial detection stage of our method, we proposed a new hypothesis based on isocontours to find candidate regions. Based on this hypothesis, an isocontour usually exists around some subsequent slices of the mass in which contour's pixels have similar intensity values. According to this, we developed a new algorithm to detect mass like structures in ABUS images.

The sensitivity of the initial stage determines the maximum sensitivity of the detection system. If a cancer is missed in this step, it cannot be retrieved by the next step. Therefore, the aim of the initial stage is to reach as high sensitivity as possible at a manageable false positive rate. In our experiments it is found that four percent of cancerous masses were missed in the initial stage. Fig. 14 illustrates two examples of such missed masses. These two types of masses are problematic for our detection method. Fig. 14(a) shows a big mass connected to the breast border. In this case, our system cannot find any closed iso-contour near the boundary of that mass because the mass boundary in not closed. Although this type of connectivity is rarely happened, it is not impossible and a solution for finding this type of masses is left for the future work. In addition to this type of masses, very small masses that are not circular and homogeneous, as shown in Fig. 14(b), may be missed by the initial filtering using domain specific knowledge. It is notable that there is a tradeoff between sensitivity and false positive rate even in the initial stage because if false positive rate

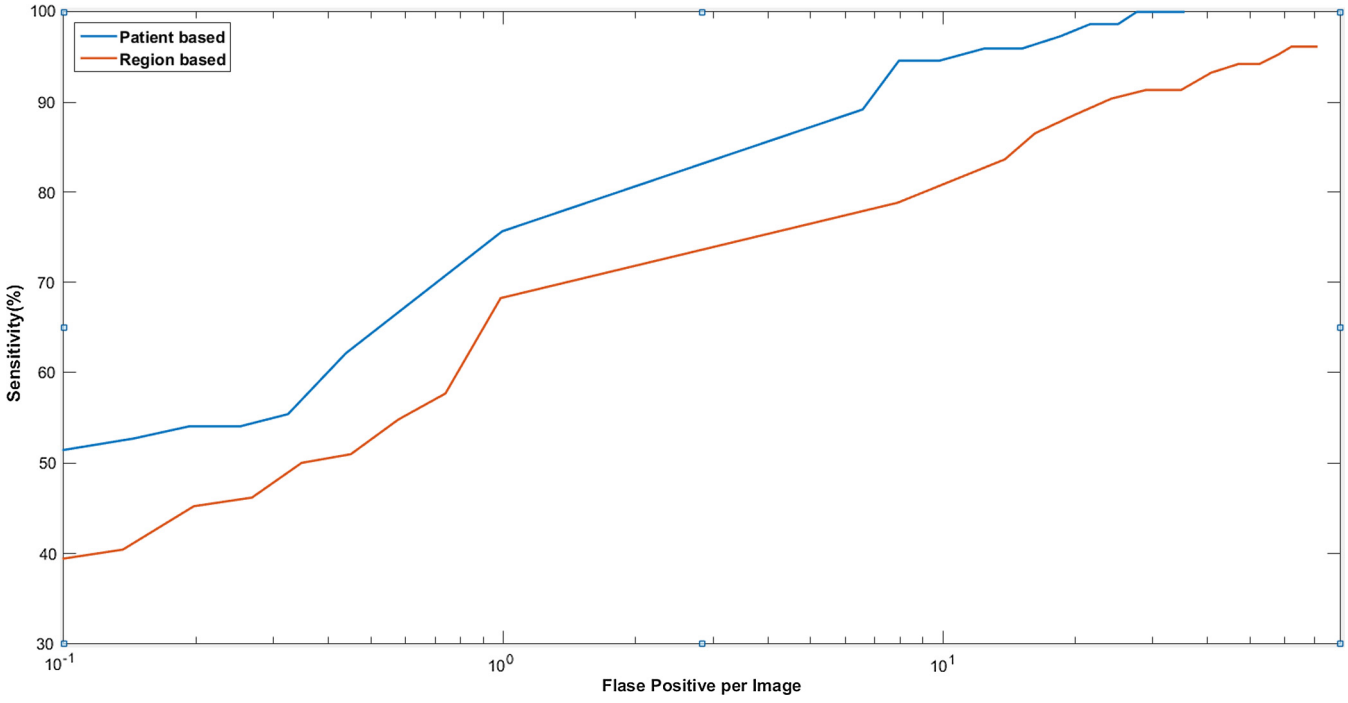

Fig. 11. FROC curves of the proposed system for patient and region analysis.

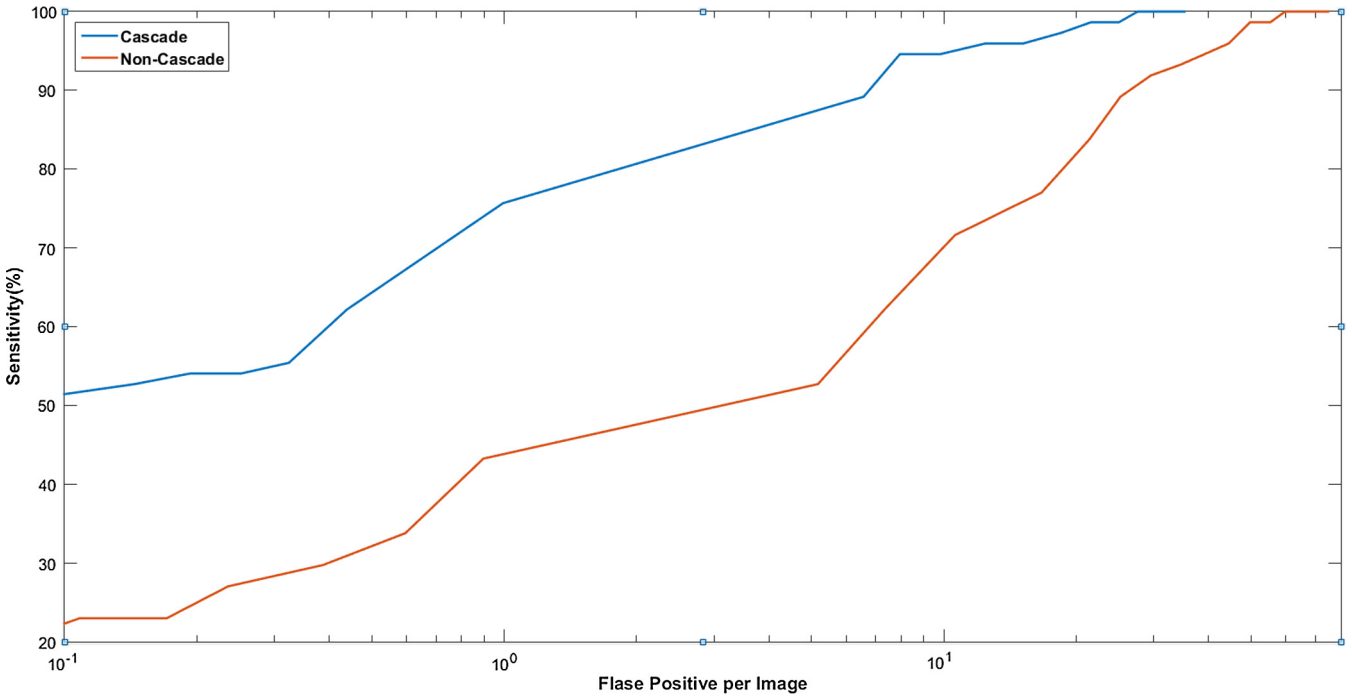

Fig. 12. Patient based performance of the proposed cascade classifier against classification without cascading.

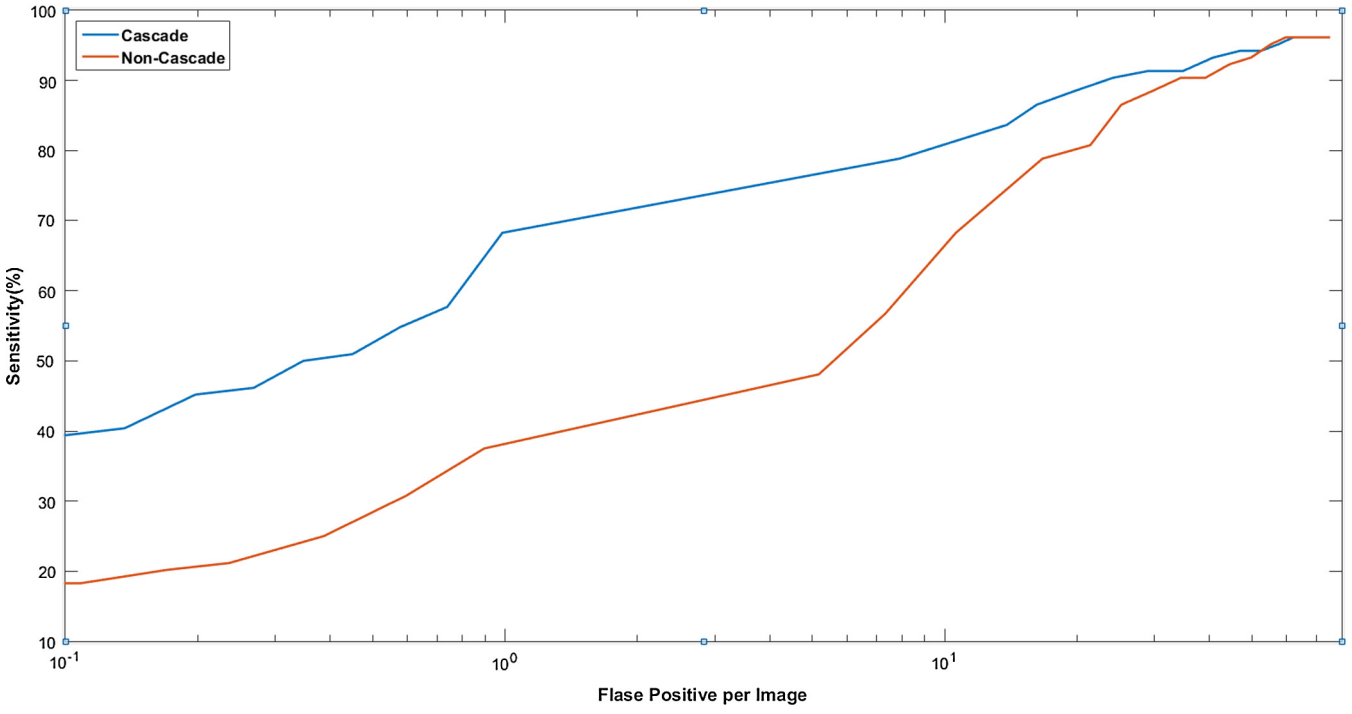

Fig. 13. Region based performance of the proposed cascade classifier against classification without cascading.

is very high then in the next step our classifier cannot manage the imbalancity of the constructed dataset. To manage this tradeoff, it is inevitable to miss such small masses. In our future works, we will extend our system to detect such small and inhomogeneous lesions by combing it with another detection method that is dedicated to find only small masses.

After initial mass candidate detection, the proportion of true positive samples against false positives is 1–66. Hence, we used Random Under-Sampling Boosting (RUSBoost) to handle the imbalancity of training sets. For classification, four groups of features are extracted for each candidate. To increase the performance of the detector, we organized four RUSBoost classifiers in a cascade structure, each trained on a group of features.

It is notable that we assign likelihoods for all samples rather than make an absolute binary classification for the CADe system. Based on thresholding on these likelihoods, we obtain FROC curve and the related operating points. Therefore, we can report classification performance measures for operating points only and not for

Table 3
comparison of different boosting algorithms in terms of sensitivity and FP pairs.

	Boosting algorithm	Sensitivity (%)	False positives per image
Case-based analysis	RUSBoost	76	1
	GentleBoost	60	1
	RobustBoost	45	1
	AdaBoost	37	1
	RUSBoost	80	2
	GentleBoost	68	2
	RobustBoost	55	2
	AdaBoost	49	2
	RUSBoost	94	9
	GentleBoost	78	9
	RobustBoost	67	9
	AdaBoost	61	9
Region-based analysis	RUSBoost	68	1
	GentleBoost	53	1
	RobustBoost	37	1
	AdaBoost	31	1
	RUSBoost	73	2
	GentleBoost	60	2
	RobustBoost	48	2
	AdaBoost	41	2
	RUSBoost	86	16
	GentleBoost	70	16
	RobustBoost	58	16
	AdaBoost	52	16

the whole CADe system. The classifier performance is summarized for some operating points in Table 4. As shown in the table, when the sensitivity (recall) is increased the precision is dramatically degraded. For example, when the recall of the system grows 5% (from 0.68 to 0.73) the number of false positives is doubled. This is a challenge for all CADe systems for detecting cancers in medical images because we deal with imbalanced datasets which bias the classifier in favor of the major class. The major class samples dominate the feature space such that increasing the recall drastically increases the number of false positives. But it is noteworthy to put into account that recall is more important than precision since cancer detection is naturally a cost sensitive problem (i.e. miss classification costs are not the same in this type of problems). That is, if a real cancer is classified as a FP, the cost is much higher than when a normal region is classified as a cancer. In the first condition, the patient would not undergo treatment and the lesion will gradually grow and spread. In the second one, the patient would be referred to pathology and the suspicious region would be diagnosed as a FP after biopsy. For this reason, FROC curves and operating points are used for the final evaluation of CADe systems. Moreover, we cannot deterministically decide which sensitivity-FP pair is better. Some radiologists may prefer CADe systems with high sensitivity (which has high FP as a disadvantage) and the others may prefer a system with lower false positive rate (which has lower sensitivity as a disadvantage). The selection of the threshold to reach the desired pair depends on the radiologist's point of view.

According to the obtained FROC curve, our proposed system can be promising to assist radiologists as a second interpreter. Future research should improve the performance of this CADe system in terms of increasing sensitivity and reducing false positive rate. Moreover, the detection system can be combined by a CADx to classify malignant masses from benign masses.

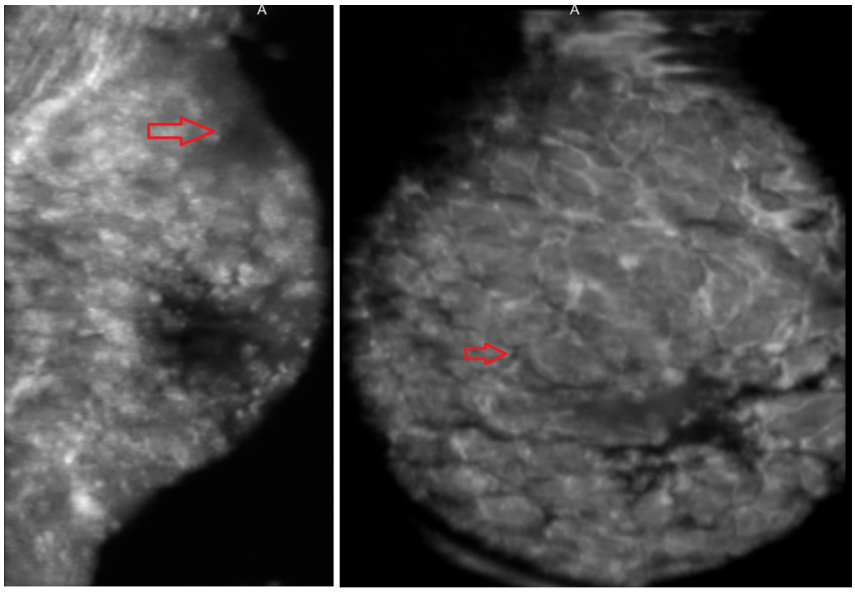

Fig. 14. Two types of masses which were missed by the initial stage; (a) a big mass near the breast border, (b) a very small mass.

Table 4
performance of the proposed classifier in terms of sensitivity, specificity, F-measure and MCC.

Sensitivity (Recall)	Precision	Specificity	f-measure	Matthews Correlation Coefficient (MCC)
0.68	0.41	0.98	0.51	0.52
0.73	0.28	0.97	0.40	0.44
0.86	0.05	0.76	0.09	0.18
0.90	0.03	0.63	0.07	0.14

References

[1] S.V. Sree, E.Y.-K. Ng, R.U. Acharya, O. Faust, Breast imaging: a survey, World J. Clin. Oncol. 2 (4) (2011) 171–178.

[2] A. Pijpe, N. Andrieu, D.F. Easton, A. Kesminiene, E. Cardis, C. Nogus, M. Gauthier-Villars, C. Lasset, J.P. Fricker, S. Peock, D. Frost, D.G. Evans, R.A. Eeles, J. Paterson, P. Manders, C.J. van Asperen, M.G.E.M. Ausems, H. Meijers-Heijboer, I. Thierry-Chef, M. Hauptmann, D. Goldgar, M.A. Rookus, F.E. van Leeuwen, Exposure to diagnostic radiation and risk of breast cancer among carriers of BRCA1/2 mutations: retrospective cohort study (gene-rad-risk), BMJ 345 (2012) e5660.

[3] C.A. Pena-Reyes, M. Sipper, L. Prieto, Sensitive, specific, and interpretable: evolving a fuzzy mammographic-interpretation assessment tool, IEEE World Congress on Computational Intelligence. IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No.02CH37291), vol. 2, 2002, pp. 837–842.

[4] K. Drukker, C.A. Sennett, M.L. Giger, Computerized detection of breast cancer on automated breast ultrasound imaging of women with dense breasts, Med. Phys. 41 (1) (2014) 12901–12909.

[5] H.J. Shin, H.H. Kim, J.H. Cha, Current status of automated breast ultrasonography, Ultrasonography 34 (3) (2015) 165–172.

[6] K.M. Kelly, J. Dean, W.S. Comulada, S.-J. Lee, Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts, Eur. Radiol. 20 (3) (2010) 734–742.

[7] C.-M. Lo, R.-T. Chen, Y.-C. Chang, Y.-W. Yang, M.-J. Hung, C.-S. Huang, R.-F. Chang, Multi-dimensional tumor detection in automated whole breast ultrasound using topographic watershed, IEEE Trans. Med. Imaging 33 (7) (2014) 1503–1511.

[8] W.K. Moon, Y.-W. Shen, M.S. Bae, C.-S. Huang, J.-H. Chen, R.-F. Chang, Computer-aided tumor detection based on multi-scale blob detection algorithm in automated breast ultrasound images, IEEE Trans. Med. Imaging 32 (7) (2013) 1191–1200.

[9] Y. Ikedo, D. Fukuoka, T. Hara, H. Fujita, E. Takada, T. Endo, T. Morita, Development of a fully automatic scheme for detection of masses in whole breast ultrasound images, Med. Phys. 34 (11) (2007) 4378–4388.

[10] T. Tan, B. Platel, R. Mus, L. Tabar, R. Mann, N. Karssemeijer, Computer-aided detection of cancer in automated 3-D breast ultrasound, IEEE Trans. Med. Imaging 32 (2) (2013) 1698–1706.

[11] T. Tan, J.J. Mordang, J. van Zelst, A. Grivegnee, A. Gubern-Merida, J. Melendez, R. Mann, W. Zhang, B. Platel, N. Karssemeijer, Computer-aided detection of breast cancers using Haar-like features in automated 3D breast ultrasound, Med. Phys. 42 (4) (2015) 1498–1504.

[12] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D: Nonlinear Phenomena 60 (1–4) (1992) 259–268.

[13] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, IEEE Trans. Image Process. 15 (12) (2006) 3736–3745.

[14] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3D transform-domain collaborative Filtering, IEEE Trans. Image Process. 16 (8) (2007) 1–16.

[15] S.G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression, IEEE Trans. Image Process. 9 (2007) 1532–1546.

[16] K.Y. Yu, S. Acton, Speckle reducing anisotropic diffusion, IEEE Trans. Image Process. 11 (11) (2002) 1260–1270.

[17] J.S. Lee, Digital image enhancement and noise filtering by using local statistics, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-2 (1980) 165–168.

[18] P. Coupe, P. Hellier, C. Kervrran, C. Barillot, Non-local means-based speckle filtering for ultrasound images, IEEE Trans. Image Process. 10 (18) (2009) 2221–2229.

[19] R. Wenger, A.K. Peters, Marching cubes and variants, in: Isosurfaces: Geometry, Topology and Algorithms, CRC Press, 2013, pp. 17–30.

[20] C. Li, C. Xu, C. Gui, M.D. Fox, Distance regularized level set evolution and its application to image segmentation, IEEE Trans. Image Process. 19 (12) (2010) 3243–3254.

[21] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139.

[22] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, Ann. Stat. 28 (2) (2000) 337–407.

[23] Y. Freund, A More Robust Boosting Algorithm, 2009. Available from: <0905.2138v1>.

[24] C. Ceiffert, T.M. Khoshgoftar, J. Van Hulse, RUSBoost: a hybrid approach to alleviating class imbalance, IEEE Trans. Syst. Man Cybern. 40 (1) (2010) 185–197.

[25] H. Liu, T. Tan, J. van Zelst, R. Mann, N. Karssemeijer, B. Platel, Incorporating texture features in a computer-aided breast lesion diagnosis system for automated three dimensional breast ultrasound, J. Med. Imag. 1 (2) (2014) 1–8.