Front. Electr. Electron. Eng. China 2008, 3(3): 267–273
DOI 10.1007/s11460-008-0059-6

RESEARCH ARTICLE

Ying WANG, Xinbo GAO

Mass detection algorithm based on support vector machine and relevance feedback

Abstract To improve the detection of mass with appearance that borders on the similarity between mass and density tissues in the breast, an support vector machine classifier based on typical features is designed to classify the region of interest (ROI). Furthermore, relevance feedback is introduced to improve the performance of support vector machines. A new mass detection scheme based on the support vector machine and the relevance feedback is proposed. Simulation experiments on mammograms illustrate that the novel support vector machine classifier based on typical features can improve the detection performance of the featureless classifier by 5%, while the introduction of relevance feedback can further improve the detection performance to about 90%.

Keywords support vector machine, relevance feedback, mass detection, feature extraction

1 Introduction

Breast cancer is well-known as one of the primary causes of cancer deaths among women. Early diagnosis and treatment can effectively increase the survival chance of patients. Mass is the major indication of early breast cancer on mammograms. However, it is difficult to distinguish masses from normal tissues because of various appearances and their ambiguous edge. Most mass detection algorithms consist of two stages: detection of suspicious regions via mammography and classification of suspicious regions as mass or normal tissue. The second stage, which is to reduce as many false positives as possible, is also the decision component of detection methods. Sahiner et al. proposed a texture feature-based convolution neural network for this task [1]. Wei et al. investigated the use of global and local multi-resolution texture features to reduce the number of false positive detections on a set of manually extracted regions of interest (ROI) [2]. Tourassi et al. employed template matching and then detected abnormalities by mutual information [3]. However, these methods seldom employ complete feature data sets, and the classification approaches are generally classical machine learning algorithms such as the decision tree method and neural networks. Therefore, their detection performances need improvement.

The support vector machine (SVM) is a statistical learning method based on structural risk minimization. The advantage of SVM over other classifiers is that its setting is easier. It generally performs better on novel data and is able to compress useful information of high-dimensional space into a small number of elements named support vectors. SVM is therefore capable of learning in a sparse, high-dimensional space by using very few training examples. SVM classifier has already been applied to calcification detection and yielded very good results [4]. A featureless approach based on SVM for the detection of masses has been proposed in Ref. [5], whose sensitivity is not very satisfactory. To improve detection performance, a new SVM detection method based on typical features is developed in this paper. On the basis of coarse detection steps, large numbers of suspicious areas are extracted. The feature data of suspicious areas are extracted and classified by means of the SVM classifier. To improve the performance of the SVM classifier, a relevance feedback method is introduced. The proposed feedback learning model can achieve better detection results, as it successfully removes more false positives from suspicious areas.

2 SVM Classifier

The simplest case of the SVM classifier has training patterns that are linearly separable [6,7]. Using the training data set $\{(\boldsymbol{x}_i, y_i), i = 1,2,\ldots,l\}$, a linear SVM classifier in its original form is formulated as a minimization of the following cost function:

Translated from *Journal of Xidian University*, 2007, 34(2): 239–245
[译自: 西安电子科技大学学报(自然科学版)]

Ying WANG (✉), Xinbo GAO
School of Electronic Engineering, Xidian University, Xi'an 710071, China
E-mail: wying@lab202.xidian.edu.cn

$$J(\boldsymbol{w},\xi)=\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+C\sum_{i=1}^{l}\xi_i, \tag{1}$$

subject to

$$y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\right)\geqslant 1-\xi_i,\quad \xi_i\geqslant 0,\quad i=1,2,\ldots,l,$$

where C is a user-specified, positive parameter, and ξ_i are slack variables. When the two classes are separable, the SVM classifier can maximize the separating margin between the two classes. The solution of cost function can be obtained through a Wolfe dual problem with Lagrangian multiplier α_i :

$$Q(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j\left(\boldsymbol{x}_i\cdot\boldsymbol{x}_j\right). \tag{2}$$

Subject to the following constraints:

$$\sum_{i=1}^{l}\alpha_i y_i=0,\quad 0\leqslant \alpha_i\leqslant C, \tag{3}$$

where α_i are the Lagrange multipliers that correspond to each of the inequality constraints in Eq.(1). In practice, the maximization in Eq. (2) is solved numerically through quadratic programming. It can be easily identified that only a small number of α_i is nonzero, which corresponds to the support vector among samples.

In a more general case where the data points are not linearly separable in the input space, a nonlinear transformation is used to map the data vector into a high dimensional space (called feature space) prior to applying the linear maximum-margin classifier. To avoid the potential pitfall of overfitting in the higher dimensional space, the SVM uses a kernel function, in which nonlinear mapping is implicitly embedded. With the use of a kernel, the discriminant function in an SVM classifier has the following form:

$$f(\boldsymbol{x})=\sum_{i=1}^{L_{\mathrm{s}}}\alpha_i y_i K(\boldsymbol{x}_i,\boldsymbol{x})+b, \tag{4}$$

where $K(\cdot,\cdot)$ is the kernel function, $\boldsymbol{x}_i$ are the so-called support vectors determined from training data, L_{s} is the number of support vectors, y_i is the class indicator (e.g., +1 for class 1 and -1 for class 2) associated with each $\boldsymbol{x}_i$, and α_i are constants also determined from training data.

3 SVM classifier based on relevance feedback

3.1 Relevance feedback

The relevance feedback (RF) is a supervised learning method for improving the efficiency of information systems. The method employs a post-query process to refine the search by using positive and/or negative indications from the user of the relevance of retrieved images. During the general feedback procedure, users always label samples randomly, which can contribute slightly to refining the retrieval result, if the useless samples are indicated. Therefore, if the system can actively provide valuable samples to users, the efficiency and accuracy of the retrieval system will be significantly improved.

3.2 Relevance feedback of SVM classifier

Since the information of original training samples is limited, using the SVM classifier only once cannot achieve satisfactory detection results. To obtain more feature information from the feedback samples and improve the performance of the SVM classifier, the RF method is introduced to allow the SVM classifier to learn more from the feedback procedure. We combine RF with our SVM detection system hoping that the feedback information can improve detection performance.

As we know, Eq. (4) is the classification function of the SVM, which represents the distance from the classification hyperplane to each sample and also denotes whether the sample is correctly classified. Its value approaches zero when the samples are close to the hyperplane, which indicates greater uncertainty of classified results, and vice versa. If the uncertainty of each misclassified sample can be calculated, and the samples with larger uncertainty scores can be provided to the user or fed back directly for the next training, the final classifier will possess better performance since more class information of samples is obtained and the certainty of samples is increased. For adaptively providing the most uncertain samples to experts or systems, a weight parameter is introduced to represent the uncertainty of classified results. Samples with high weights will then be selected for reducing the negative impact of useless feedback samples.

During the feedback procedure, history information also plays an important role. The system will rapidly achieve good results if one adds the information into the current feedback process. The history information will blend into the feedback procedure by assigning weights to each unlabelled image:

$$w(i)=(1-\beta)w(i)+f(\boldsymbol{x}_i), \tag{5}$$

where β is an attenuation coefficient used to control the effect of history information on the feedback procedure.

To control the current procedure simultaneously, Eq. (5) is rewritten as

$$w(i)=(1-\beta)w(i)+\alpha f(\boldsymbol{x}_i), \tag{6}$$

where α is also an attenuation coefficient. The coefficient not only keeps the effect of history information, but also emphasizes particularly on the requirement of the current detection procedure. The method effectively avoids the result of running into a local solution, and significantly improves the detection performance.

We have studied the active relevance feedback (ARF) and developed an efficient false positive removal method based on ARF for detecting any abnormality in mammograms [8]. However, when the detection is implemented on the whole image, a satisfying performance can be obtained only by applying excellent classifiers, since similar objects cannot be distinguished just through a similarity computation of features. Therefore, the SVM classifier that possesses superior generalization ability is selected and an improved SVM learning method named RF-SVM is proposed by combining our ARF method with the RF method in Refs. [9,10]. The brief procedure of the algorithm is given in Fig. 1.

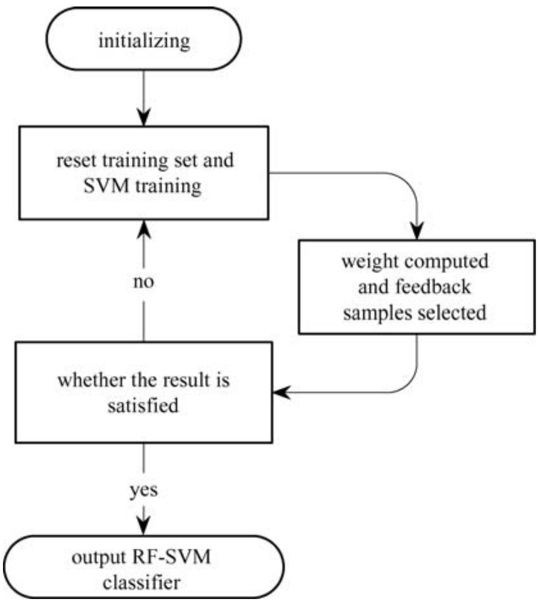

Fig. 1 RF-SVM learning framework

Let T_{p}, T_{n} denote the positive and negative training set, F_{p}, F_{n} denote the corresponding feedback sample set, and the new mass detection method based on RF-SVM can be described in detail as follows:

Step 1 Initialize T_{p}, T_{n}, F_{p}, F_{n}, and validation set V, set $w(i)=0$.

Step 2 Reset the training set as in Eqs. (7) and (8) before training, and train the SVM classifier that contains more sample information as the current classifier.

$$T_{\mathrm{p}}=T_{\mathrm{p}}\cup F_{\mathrm{p}}, \tag{7}$$

$$T_{\mathrm{n}}=T_{\mathrm{n}}\cup F_{\mathrm{n}}. \tag{8}$$

Step 3 Classify the set V, then save misclassified samples to the preliminary feedback set $F'_{W_{\mathrm{p}}}(k)$ and $F'_{W_{\mathrm{n}}}(k)$. The weight of the feedback sets $F_{W_{\mathrm{p}}}(k)$ and $F_{W_{\mathrm{n}}}(k)$ that needs computation shall be determined as follows:

$$F_{W_{\mathrm{p}}}(k)=\left(F'_{W_{\mathrm{p}}}(k)\cup F_{\mathrm{p}}\right)-F_{\mathrm{n}}, \tag{9}$$

$$F_{W_{\mathrm{n}}}(k)=\left(F'_{W_{\mathrm{n}}}(k)\cup F_{\mathrm{n}}\right)-F_{\mathrm{p}}. \tag{10}$$

Step 4 Compute the weights of misclassified samples of $F_{W_{\mathrm{p}}}(k)$ and $F_{W_{\mathrm{n}}}(k)$ respectively on the basis of Eqs. (4) and (6). Arrange the absolute values of these weights in ascending order. The samples with small scores are selected to experts or fed back directly for the next learning and detection step.

Step 5 The procedure will stop when the detection rate is more than 90%, or the amount of samples in F_{p} and F_{n} has been less than 1. This means that the current classifier has reached the system requirement, and the current classifier will be saved. Otherwise, go to Step 2 for further feedback and learning.

After learning, final detection performance of the SVM and RF-SVM methods will be achieved by classifying the test set (which is different from the validation set) using SVM and RF-SVM respectively.

4 Mass detection based on typical features and RF-SVM

For effectively improving the mass detection accuracy, a method based on typical features and RF-SVM is developed in this paper. On the basis of a coarse detection step, large numbers of ROIs are extracted. The typical features of suspicious areas are then extracted as input of SVM. To further improve the performance of the SVM classifier, an RF-SVM classifier is introduced for feedback learning. Figure 2 shows the whole detection procedure, and some steps will be depicted as follows.

4.1 Coarse detection

The ROIs in mammograms need to be extracted before the classification procedure. Most ROIs are obtained by cropping images in square areas with different sizes. Since the location and size of masses are uncertain, it is difficult

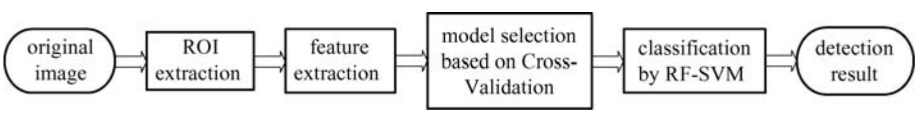

Fig. 2 Mass detection framework

to correctly crop the image. The feature extraction step may be affected as the crops also include the background area. To accurately extract the suspicious areas, a simple but effective method is employed. It begins with morphological enhancement that can remove the background noise and the structure noise inside the suspected mass patterns. The regions in enhanced images, which take on a certain intensity and contrast values, are then extracted and selected as seed regions. The ROIs will be obtained later using a fuzzy region growing algorithm [11]. Since the limitations of intensity and contrast are not strict, suspicious regions in mammograms are extracted entirely after coarse detection. Thus, there still exists a large number of false positives that need further processing.

4.2 Feature extraction

The general guidelines of feature selection are:

1) Features of patterns in different classes should have significantly different values.

2) Features should have similar values for the patterns of the same class.

3) These features should not be strongly correlated with each other.

4) Some redundant features should be deleted, and a small number of features is preferred for reducing the complexity of the classifier.

Similarity estimation of object areas is based on the typical features that characterize the areas well. Since the feature extraction of mass is based on region, it is important to select the features that can properly describe region characteristics. Many useful image features have been proposed by the academic community. These features can be divided into three categories: intensity, geometric, and texture features. To describe the features of mass as well as possible, we summarize 42 typical features in Table 1.

Table 1 Features of mass

feature sub-space	features
intensity features	contrast; invariant moment; mean gray and gradient of ROIs; standard derivation inside ROIs; higher order moments of ROIs; mean gradient of ROIs boundary
geometric features	circularity; compactness; sphericity; Fourier descriptor
texture features	laws texture; wavelet texture; co-occurrence matrix texture

The selected intensity features are always widely used in image processing [12], which can describe the average intensity, homogeneity, contrast and the margin density variance of the region. The efficient geometric features are also important for detection, since the masses always have certain features that normal areas do not have. Therefore, most selected geometric features are invariant to shifts, rotation and scaling, and they can properly express the margin roughness and shape of the region [12]. Since mass areas are usually homogeneous compared to normal ones, texture features are also significant in classification. They have been successfully applied to medical image analysis, as they can depict the texture of images such as uniformity, smoothness and difference among adjacent pixels. Mean value and standard derivation of a texture energy map are the Laws texture features we employ, which describe the characteristics of images filtered using the Laws template [13]. Features based on a co-occurrence matrix are measures related to specific textural characteristics of the image, including homogeneity, contrast, entropy and energy [12]. Since Daubechies wavelets DB_6 and DB_{20} provide a good combination of regular prototype wavelets with varying sizes to extract texture information with varying spatial frequency [13], the energy and entropy of the decomposed wavelet coefficients are computed as wavelet features.

4.3 SVM model selection

Selection of kernels is the major problem during SVM model design, since different functions will form different algorithms. To select models with good performance, we adopt a widely used statistical method called m-fold cross-validation. The training set is divided into m subsets, and the holdout method is repeated m times. Each time, one of the m subsets is used as the test set and the other $m-1$ subsets are put together to form a training set. The average error across all m trials is then computed, and the model with the smallest generalization error is adopted. Two commonly used kernel functions are given as follows:

Polynomial kernel:

$$K(\boldsymbol{x}_i,\boldsymbol{x}) = [(\boldsymbol{x}_i,\boldsymbol{x}) + 1]^d. \tag{11}$$

Gaussian radial basis (RBF) kernel:

$$K(\boldsymbol{x}_i,\boldsymbol{x}) = \exp\left[-\frac{(\boldsymbol{x}-\boldsymbol{x}_i)^2}{2\sigma^2}\right]. \tag{12}$$

In the experiment, the SVM classifier is trained using a 10-fold cross-validation procedure to confirm the best model and parametric setting. The generalization error of two kernels can be found in Table 2. Finally, the Gaussian RBF kernel function with $\sigma = 3$ and $C = 100$ is employed in SVM training and learning, since the generalization error is smallest under this setting.

5 Experimental results

In this study, the mammograms are selected from the USF DDSM database (http://marathon.csee.usf.edu/

Table 2 Generalization error of 10-fold cross-validation

model parameters	RBF kernel (Eq. (12))					polynomial kernel (Eq. (11))				
	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$	$\sigma=6$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$
$C=1$	0.0425	0.045	0.0575	0.0675	0.075	0.035	0.04	0.045	0.05	0.0525
$C=10$	0.035	0.035	0.035	0.04	0.045	0.04	0.04	0.045	0.05	0.0525
$C=100$	0.0325	0.025	0.03	0.03	0.0325	0.04	0.04	0.045	0.05	0.0525
$C=1000$	0.05	0.0375	0.03	0.0275	0.03	0.04	0.04	0.045	0.05	0.0525
$C=\text{Inf}$	0.05	0.0375	0.0325	0.0275	0.035	0.04	0.04	0.045	0.05	0.0525

Mammography/Database.html); all the lesions in images have been marked by experts. These mammograms have a dimension of 5000 × 3000 pixels, with a spatial resolution of 0.05 mm/pixel and 12-bit/16-bit gray scale. For the training purpose, the marked masses are extracted as positive samples, and the negative samples are selected representatively from numerous negative samples extracted by the coarse detection method. In our experiment, the training set includes 192 images, which contains 200 mass regions and 200 negative samples extracted from these images. Another data set of 150 mammograms is used to test and evaluate the performance of proposed algorithms. Of the 150 images, 100 images are validated data in feedback and the others are test data. According to the coarse detection result, the former contains 132 mass areas with 1084 false-positive areas, and the latter contains 64 mass areas with 599 false-positive areas.

The detection performance was evaluated quantitatively using the free-response receiver operating characteristic (FROC) curves. An FROC curve plots the true positive fraction versus the average number of false positives per image for the continuum of the decision threshold. A larger area under the curve indicates better performance.

Figure 3 shows the detection results of the BP neural network (BPNN) and SVM classifier under the same conditions. As can be seen, the SVM classifier achieves better performance. Since the testing set has 50 images, the sensitivity (true-positive (TP) rate) of the SVM classifier is 85.9%, with a false-positive (FP) fraction of 4.0 marks per image.

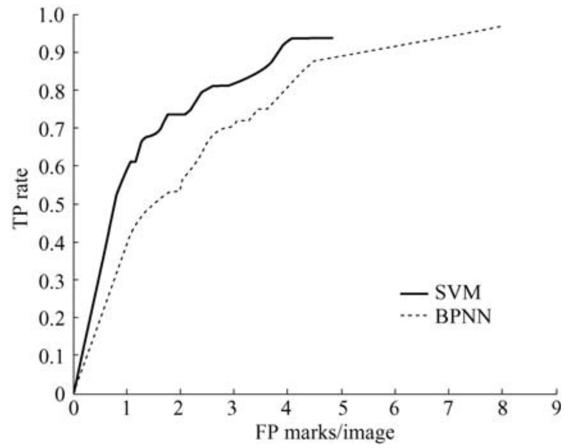

Fig. 3 Detection results of BPNN and SVM

Figure 4 presents some detection results of two mammograms. Figures 4(a) and 4(d) are two mammograms marked by experts. Figures 4(b) and 4(e) give the ROIs extracted by coarse detection. These ROIs have mass-like appearance, which is very similar to the masses drawn by experts. The test results of about 300 mammograms demonstrate that all the masses can be detected after coarse detection. We also show detection results using the SVM classifier in Figs. 4(c) and 4(f). As we can see, good performance was obtained and most false positives were removed.

However, some mammograms cannot obtain good detection results by using the SVM classifier only once, due to the complexity of glandular structures in the

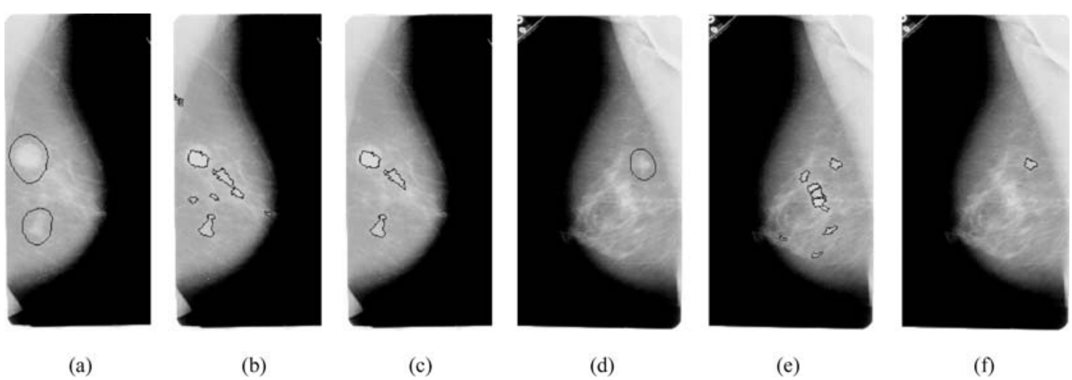

Fig. 4 SVM detection results. (a) and (d) Original image; (b) and (e) ROIs in image; (c) and (f) SVM detection result

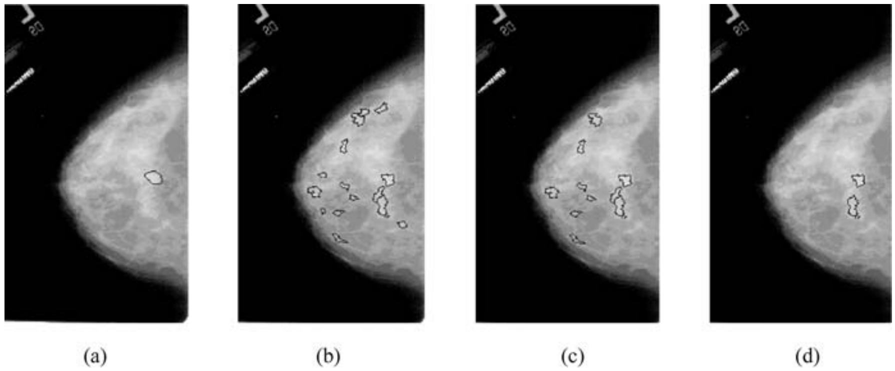

Fig. 5 SVM detection result of density mammogram. (a) Original image; (b) ROIs in image; (c) SVM detection result; (d) RF-SVM detection result

mammary. As shown in Fig. 5(a), the mammary has plenty of glandular tissues and the ROIs extracted are shown in Fig. 5(b). Since many regions possess similar features with masses, only a small fraction of false positives are removed after the detection procedure.

Thus, the relevance feedback method is used to improve detection performance. During the feedback learning procedure, 100 images are used as validation data. Figure 6 shows the FROC curves of detection results using the SVM and different iterations of RF-SVM separately. From the figure we can see that use of the proposed feedback method can improve the detection result. It is also noted that the performance is further improved with more feedback.

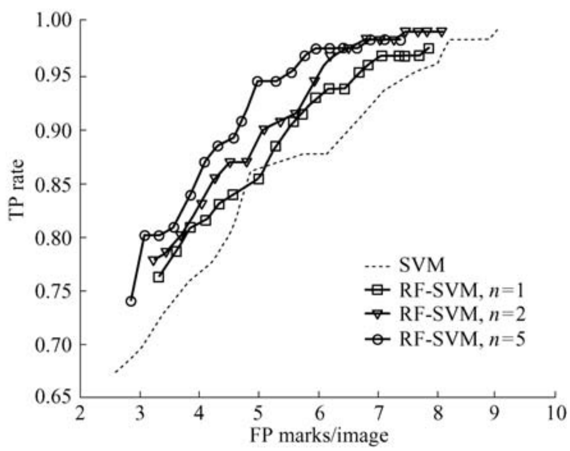

Fig. 6 SVM and RF-SVM detection result on validation set

To verify the efficiency of the feedback learning model, Fig. 7 shows the detection results on the test set using SVM and RF-SVM, which feedbacks and learns from the validation set five times. As we can see from Fig. 7, the feedback learning model can further improve the detection performance of the SVM classifier, while removing more false positives in images. An SVM trained with the feedback learning procedure makes the sensitivity of the SVM classifier rise to 90.6% and the false-positive fraction fall to 3.6 marks per image. Compared with the results reported in Ref. [6], our method can obtain higher sensitivity. Then we use the RF-SVM classifier where the feedback runs five times to classify the ROIs in Fig. 5(b), and the detection result is shown in Fig. 5(d). Obviously, the SVM classifier using feedback learning achieves better detection performance and can remove more false positives in images.

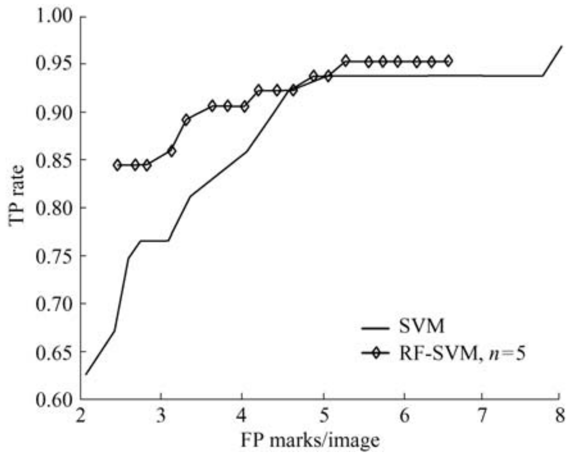

Fig. 7 SVM and RF-SVM detection result on test set

6 Conclusions

In this paper, a mass detection method based on the SVM and the relevance feedback is proposed to separate the masses from normal areas correctly. ROIs are extracted first by the coarse detection method. The SVM is then trained and tested using features extracted from ROIs such as intensity and texture. To remove more false positives in images and improve performance, the RF-SVM method is developed by introducing the relevance feedback procedure. Experimental results demonstrate that the RF-SVM classifier can effectively improve the detection performance of the SVM classifier by removing more false positives. We are also planning to investigate the feedback procedure to improve the generalization capacity of the SVM feedback learning model.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 60771068), the Key Project of Chinese Ministry of Education (No. 104173), and the Program for New Century Excellent Talents in University (No. NCET-04-0948).

References

1. Sahiner B, Chan H P, Petrick N, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. IEEE Transactions on Medical Imaging, 1996, 15(5): 598–610
2. Wei D, Chan H P, Petrick N, et al. False positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis. Medical Physics, 1997, 24(6): 903–914
3. Tourassi G D, Vargas-Voracek R, Catarious D M Jr, et al. Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information. Medical Physics, 2003, 30(8): 2123–2130
4. EI-Naqa I, Yang Y Y, Wernick M N, et al. A support vector machine approach for detection of microcalcifications. IEEE Transactions on Medical Imaging, 2002, 21(12): 1552–1563
5. Campanini R, Dongiovanni D, Iampieri E, et al. A novel featureless approach to mass detection in digital mammograms based on support vector machines. Physics in Medicine & Biology, 2004, 49(6): 961–975
6. Vapnik V N. Statistical Learning Theory. New York: Wiley, 1998
7. Ding A L, Liu F, Yao X. Intelligent target recognition based on the support vector machine. Journal of Xidian University, 2001, 28(6): 743–746 (in Chinese)
8. Wang Y, Gao X B, Dong Y. A partial supervised mammographic masses detection method. Chinese Journal of Medical Imaging Technology, 2005, 21(10): 1572–1575 (in Chinese)
9. EI-Naqa I, Yang Y Y, Galatsanos N P, et al. A similarity learning approach to content-based image retrieval: application to digital mammography. IEEE Transactions on Medical Imaging, 2004, 23(10): 1233–1244
10. Xu Y H, Li J L, Chen E H, et al. A novel SVM based relevance feedback algorithm in image retrieval. Computer Engineering, 2004, 30(24): 116–118 (in Chinese)
11. Guliato D, Rangayyan R M, Carnielli W A, et al. Segmentation of breast tumors in mammograms by fuzzy region growing. In: Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Hong Kong, China. IEEE, 1998, 2: 1002–1005
12. Gonzalez R C, Woods R E. Digital Image Processing, 2nd Edition, New Jersey: Prentice Hall, 2002
13. Miller P, Astley S. Classification of breast tissue by textural analysis. Image and Vision Computing, 1992, 10(5): 277–282