# Joint two-view information for computerized detection of microcalcifications on mammograms

Berkman Sahiner,[a)] Heang-Ping Chan, Lubomir M. Hadjiiski, Mark A. Helvie,
Chinatana Paramagul, Jun Ge, Jun Wei, and Chuan Zhou
*Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109-0904*

(Received 10 January 2006; revised 3 May 2006; accepted for publication 4 May 2006; published 26 June 2006)

We are developing new techniques to improve the accuracy of computerized microcalcification detection by using the joint two-view information on craniocaudal (CC) and mediolateral-oblique (MLO) views. After cluster candidates were detected using a single-view detection technique, candidates on CC and MLO views were paired using their radial distances from the nipple. Candidate pairs were classified with a similarity classifier that used the joint information from both views. Each cluster candidate was also characterized by its single-view features. The outputs of the similarity classifier and the single-view classifier were fused and the cluster candidate was classified as a true microcalcification cluster or a false-positive (FP) using the fused two-view information. A data set of 116 pairs of mammograms containing microcalcification clusters and 203 pairs of normal images from the University of South Florida (USF) public database was used for training the two-view detection algorithm. The trained method was tested on an independent test set of 167 pairs of mammograms, which contained 71 normal pairs and 96 pairs with microcalcification clusters collected at the University of Michigan (UM). The similarity classifier had a very low FP rate for the test set at low and medium levels of sensitivity. However, the highest mammogram-based sensitivity that could be reached by the similarity classifier was 69%. The single-view classifier had a higher FP rate compared to the similarity classifier, but it could reach a maximum mammogram-based sensitivity of 93%. The fusion method combined the scores of these two classifiers so that the number of FPs was substantially reduced at relatively low and medium sensitivities, and a relatively high maximum sensitivity was maintained. For the malignant microcalcification clusters, at a mammogram-based sensitivity of 80%, the FP rates were 0.18 and 0.35 with the two-view fusion and single-view detection methods, respectively. When the training and test sets were switched, a similar improvement was obtained, except that both the fusion and single-view detection methods had superior test performances on the USF data set than those on the UM data set. Our results indicate that correspondence of cluster candidates on two different views provides valuable additional information for distinguishing FPs from true microcalcification clusters. © *2006 American Association of Physicists in Medicine*. [DOI: 10.1118/1.2208919]

Key words: computer-aided diagnosis, microcalcification clusters, segmentation

## I. INTRODUCTION

There is strong evidence that imaging the breast in two views—mediolateral oblique (MLO) and craniocaudal (CC) views—increases the cancer detection sensitivity while decreasing the recall rate.[1,2] The radiologist combines the information from the two views to confirm true positives (TPs) and to reduce false positives (FPs). It is expected that computerized detection could also benefit from the joint two-view information available in a screening study. Our laboratory has been developing image analysis methods to exploit the joint two-view information for FP reduction in computerized detection of masses[3] and microcalcifications[4] on mammograms.

In recent years, a number of research groups have investigated the use of two-view mammograms of the same breast in a given examination to improve computerized lesion analysis. Our group investigated the fusion of information from two mammographic views to improve the performance a CAD system for breast mass detection.[3,5,6] The distance between the nipple and computer-detected objects on the two views was used to geometrically pair the objects, which were then classified using a correspondence classifier. To establish the geometric relationship of the locations of the same object seen in two mammographic views, we used a data set of 116 two-view cases containing masses, microcalcification clusters, and large benign calcifications. The absolute value of the nipple-to-object distance (NOD) difference on the two views was found to be less than 16 mm for 83% of the lesions. Yam *et al.*[7] and Kita *et al.*[8] developed a method to extract three-dimensional (3D) information about breast lesions from two mammographic views. Their technique was based on a breast model to estimate the deformation of the canonical breast representation under compression from that without compression. The method was applied to 3D reconstruction of microcalcifications, as well as to the prediction of the lesion location on one view from the location on the other view. For a data set of 37 lesions, their method could predict the location in the second view within a band of pixels ±27 mm from an epipolar line.[8] The average minimum distance from the epipolar line was 6.8 mm, while the average distance using the NOD difference was 8.6 mm. In a different publication,[7] on a data set of 35 lesions, they reported average distances of 6.5 and 6.9 mm using the epipolar lines and the NOD difference, respectively. Chang *et al.*[9] compared two methods for predicting a search region on the MLO view (or the CC view) for a lesion detected on the CC view (or the MLO view). The first method was based on the ratio of the NOD on the two views, and the second method was on a Cartesian straight-line distance. They found that the two methods had essentially similar performance in predicting the lesion location. Despite the efforts by many investigators in studying the geometric correspondence between the lesion locations on two mammographic views, to our knowledge the study by Paquerault *et al.*[3] was the only journal publications to date that used two-view information to improve the single-view detection of masses. Sahiner *et al.*[4] performed a preliminary study to investigate the use of joint two-view information to improve computerized microcalcification detection. The current study further improved the two-view fusion scheme and evaluated its performance with an independent data set.

## II. METHODS

The joint two-view detection method used in this study is based on the assumption that if a single-view detection algorithm detects the corresponding true cluster on the CC and MLO views of the same breast, the TP clusters on the two views will exhibit similarities in their geometric, morphological, and textural features. A FP cluster detected on the CC view is expected to exhibit a lesser degree of similarity with the true cluster on the MLO view, and vice-versa. Similarly, the degree of similarity exhibited by two FP clusters on two different views is expected to be lesser than that between two TPs. In this study, we made use of this assumption by performing similarity analysis between cluster candidates detected on the two views and distinguishing true pairs (TP-TP pairs) from false pairs (FP-TP, TP-FP, and FP-FP pairs). We used the NOD difference to define a limited number of object pairs. The scores resulting from the similarity classifier may not provide adequate sensitivity if used alone. The reasons are twofold. First, some lesions may not be visible on both views. Second, even if a lesion is visible on both views, it may have been missed by the computer on one view. However, we found that by designing a proper strategy in which the two-view pair classification scores are fused with the single-view scores, the overall accuracy of the detection system can be significantly improved. The block diagram of the two-view fusion method is shown in Fig. 1.

### A. Data sets

Two independent data sets collected at different institutions were used for training and testing the two-view detection algorithm. Each data set consisted of a group of two-view mammograms that contained at least one microcalcification cluster (the positive group) and a different group of two-view mammograms that were free of microcalcification clusters (the normal group).

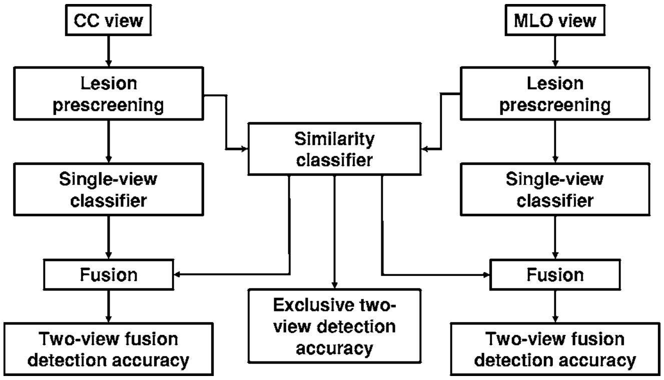

FIG. 1. The block diagram of the relationship between the single-view detection, exclusive two-view detection, and two-view detection methods. The lesion prescreening, single-view classifier, and fusion blocks for the CC and MLO views are identical.

The training data set included mammograms from the publicly available University of South Florida (USF) digitized mammogram database.[10] The positive training group consisted of malignant microcalcification cases in the USF database digitized with a Lumisys 200 laser scanner (volumes: cancer_01, cancer_02, cancer_05, cancer_09, and cancer_15). This group initially contained 124 cases (124 CC and 124 MLO view mammograms). Eight cases (16 mammograms) were excluded from the positive training group because these cases contained diffuse microcalcifications scattered over a large breast area, and the correspondence of the microcalcification locations on two views cannot be established. The positive training group therefore consisted of 232 mammograms. On these positive mammograms, 254 microcalcification locations were identified, of which 235 were proven to be malignant by biopsy. The remaining 19 locations, which were detected on breasts that had undergone biopsy for a different suspicious cluster, did not have biopsy proof. The normal training group consisted of the contralateral mammograms of the patients included in the positive training group, as well as mammograms of patients with a detected breast mass in the contralateral breast. Initially, the pool of normal cases included 494 mammograms. Two MQSA radiologists at our institution examined the cases to confirm that they are free of microcalcification clusters and calcified vessels. Upon this inspection, 44 pairs of mammograms were excluded because at least one view contained a calcified vessel or a microcalcification cluster that was not marked in the USF database. Our normal training group therefore consisted of 406 mammograms (203 cases). The nipple location for each mammogram was manually identified at our institution.

The lesions in the training database were rated for their subtlety by experienced radiologists and provided with the USF database. The distribution of the subtlety ratings for the training data set is shown in Fig. 2, where 1 indicates the most obvious clusters, and 5 indicates the most subtle. Note

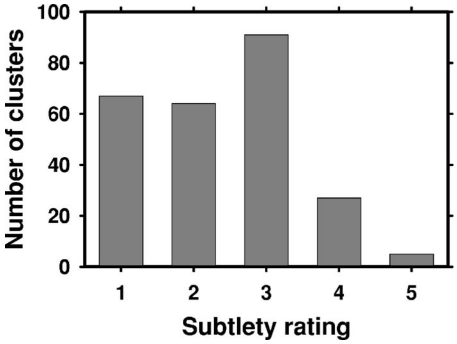

FIG. 2. The distribution of the subtlety ratings (1: most obvious, 5: most subtle) for the microcalcification clusters in the training data set. The ratings were provided with the USF database. Note that in order to be consistent with our rating scale in which a subtle lesion has a higher rating (e.g., Fig. 4), we reversed the original ratings (and their interpretation) in the USF database for this figure.

TABLE I. The positive and normal data groups for the training and test data sets.

| | Training (USF data set) | | Test (UM data set) | |
|---|---|---|---|---|
| | Positive | Normal | Positive | Normal |
| Number of two-view cases | 116 | 203 | 96 | 71 |
| Number of cluster locations | 254 | | 218 | |
| Number of malignant cluster locations | 235 | | 66 | |
| Number of benign cluster | | | 148 | |
| Number of locations with unknown status | 19 | | 4 | |

that in order to be consistent with our rating scale in which a subtle lesion has a higher rating (e.g., Fig. 4), we reversed the original ratings in the USF database for this figure. An assessment that follows the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) categories was also provided for the clusters. The distribution of the assessment ratings for the training data set is shown in Fig. 3. Most clusters had an assessment rating of 4 (suspicious abnormality, biopsy should be considered), or 5 (highly suggestive of malignancy, appropriate action should be taken). This is consistent with the fact that all mammograms contained at least one biopsy-proven malignant microcalcification cluster.

The test data set consisted of mammograms collected with Institutional Review Board approval at the University of Michigan (UM). The positive test group consisted of 96 pairs of mammograms, each of which contained at least one biopsy-proven microcalcification cluster. The cases were collected consecutively from our biopsy-proven mammogram database, with the exception that any case containing diffuse microcalcifications or calcified vessels was excluded. There were 218 microcalcification cluster locations marked by a MQSA radiologist on 192 mammograms. The same radiologist also established the correspondence of the clusters on two views using all the clinical information related to the case. Ten of the clusters were visible only on one mammographic view, and the remaining clusters were seen on both views. We thus had 104 clusters seen on both views, accounting for 208 marked locations. Sixty-six of the marked locations corresponded to biopsy-proven malignant clusters, and 148 locations corresponded to clusters that were benign either by biopsy or follow-up. The remaining four locations (corresponding to two clusters seen on both views) were on biopsy-proven mammograms containing malignant clusters, but their pathology could not be ascertained because these clusters did not undergo biopsy. The normal test group initially contained 100 pairs of mammograms from patients with a detected breast mass in the contralateral breast. Two MQSA radiologists examined these mammograms to confirm that they are free of microcalcification clusters and calcified vessels. The inspection revealed that 29 pairs of mammograms contained a calcified vessel or a microcalcification cluster on at least one view. Our normal test group therefore consisted of 142 mammograms (71 cases) after exclusion of the 29 pairs. The nipple location for each case was manually identified. Table I summarizes the training and test data sets.

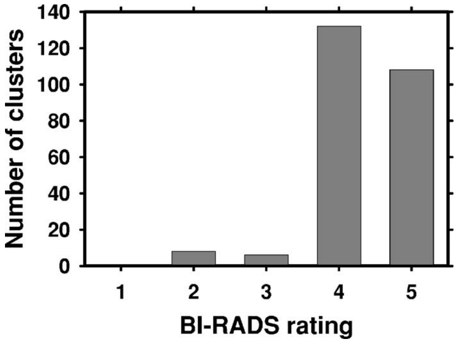

FIG. 3. The distribution of the assessment ratings for the microcalcification clusters in the training (USF) data set. The assessment follows the American College of Radiology BI-RADS lexicon and was provided with the USF database. Since all cases had biopsy-proven malignant clusters, there were very few ratings of 2 (benign finding) and 3 (probably benign finding, short-interval follow-up suggested). A majority of the ratings were 4 (suspicious abnormality, biopsy should be considered), or 5 (highly suggestive of malignancy, appropriate action should be taken).

The subtlety of the microcalcification clusters in the UM data set was rated by an experienced MQSA radiologist on a scale of 1 (obvious) to 10 (subtle) relative to the visibility range of microcalcifications encountered in clinical practice. The distribution of the subtlety ratings for benign and malignant clusters is shown in Fig. 4. It can be seen that the malignant and benign clusters had similar subtlety ratings, with the malignant clusters slightly more subtle than benign clusters. Since there are no standards or methods for calibration of the subtlety ratings across different institutions, it is not possible to compare the subtlety ratings of the UM cases with those of the USF cases. The same experienced MQSA radiologist also provided a likelihood of malignancy rating

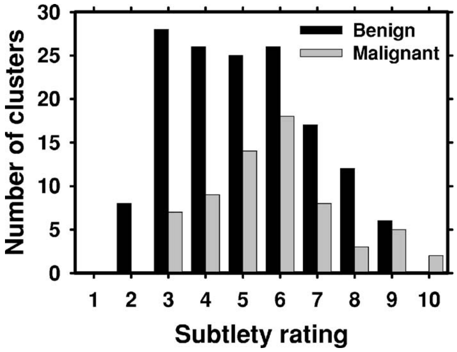

FIG. 4. The distribution of the subtlety ratings (1: most obvious, 10: most subtle) for the malignant and benign microcalcification clusters in the test (UM) data set. The ratings were provided by MQSA radiologists at UM.

for each cluster in the UM data set. The distribution of the likelihood of malignancy ratings for benign and malignant clusters is shown in Fig. 5.

The USF data set was digitized using a Lumisys 200 laser scanner with an optical density (OD) range of 0–3.6, and the UM data set was digitized using a LUMISCAN 85 laser scanner with an OD range of 0–4.0. Both digitizers were calibrated so that the gray values were linearly and inversely proportional to the OD, with a slope of −0.001 OD unit/pixel value. All mammograms were digitized at a pixel resolution of $0.05\times0.05$ mm with 4096 gray levels. The image matrix size was reduced by averaging every $2\times2$ adjacent pixels and down-sampling by a factor of 2, resulting in images with a pixel size of $0.1\times0.1$ mm for further analysis.

## B. Cluster prescreening

The purpose of the lesion prescreening stage in the CAD system is to identify areas containing microcalcification cluster candidates so that these areas can be further analyzed in subsequent stages to determine whether they contain a true cluster or a FP cluster. First, the image is processed using a difference-image technique to enhance the signal-to-noise ratio (SNR) of the microcalcifications.[11] Second, potential signals are segmented from the image background using global and locally adaptive segmentation techniques. Rule-based classification is applied to the signal size, contrast and SNR to identify suspected individual microcalcifications.[11,12] A convolution neural network (CNN)[12] is trained to further exclude FPs. Finally, a regional clustering procedure is used to identify clustered microcalcifications. Isolated signals, considered to be either noise points or isolated calcifications are excluded, while signals that are within a neighborhood of other signals are retained as potential microcalcifications within a cluster.

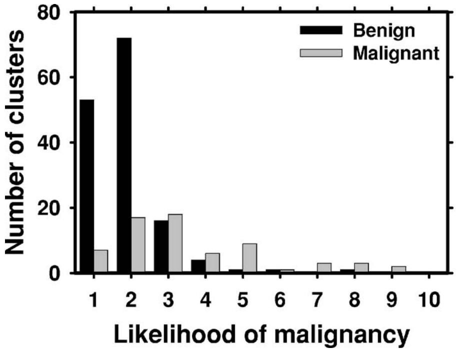

FIG. 5. The distribution of the likelihood of malignancy ratings (1: least likely to be malignant, 10: most likely to be malignant) for the malignant and benign microcalcification clusters in the test data set. The ratings were provided by MQSA radiologists at UM.

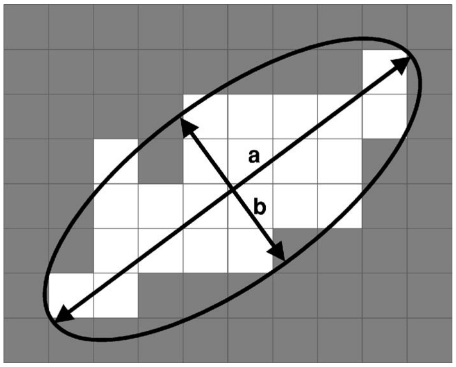

FIG. 6. The extraction of the shape features. An ellipse is fitted to the segmented microcalcification using a moment method,[13] and the lengths of the major axis (a) and the minor axis (b) are determined. The eccentricity feature was defined as $\sqrt{a^2-b^2}/a$, and the axis ratio feature was defined as $a/b$. The moment ratio feature was defined as the ratio of the smaller second moment of the shape to the larger second moment.

## C. Feature extraction

Three types of features were extracted from each cluster or the region enclosing the cluster: Morphological features, texture features, and features derived from the CNN scores of the microcalcifications.

A number of morphological features were extracted for a cluster. First, 11 morphological features related to the size, mean density, shape, and contrast were extracted from each individual microcalcification. The size of a microcalcification was estimated as the number of pixels in the segmented microcalcification region. The mean density was found by averaging the pixel values within the segmented microcalcification region. Three shape features were extracted based on an ellipse fitted to each segmented microcalcification. Figure 6 depicts these features, which were explained in more detail previously.[13] Six features related to the contrast of the microcalcification were extracted based on the statistics of the gray level values within the segmented microcalcification area and the background surrounding the segmented microcalcification. The extraction of these contrast features is described in Fig. 7. The background surrounding the segmented lesion was obtained by dilating the segmented lesion with a circular structuring element. The radius of this structuring element was defined as $R_s=\max\{2.0, 0.6R_{\text{eq}}\}$ pixels, where $R_{\text{eq}}$ is the radius of a circle with the same area as the segmented microcalcification. Let $C$ and $S$ denote the segmented microcalcification and its background region, respectively. Let the

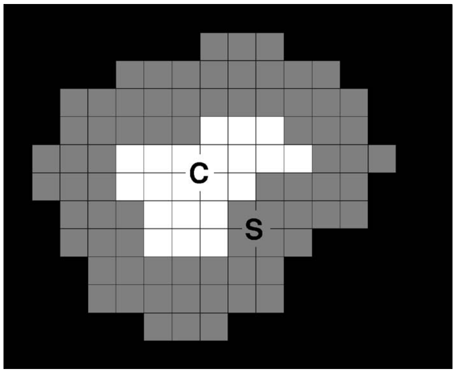

FIG. 7. The extraction of the features related to the microcalcification contrast. The white region C at the center represents the segmented microcalcification, and the surrounding gray region S represents the background, obtained by using a dilation operator as described in the text.

average, variance, minimum, and maximum of gray levels within $C$ be denoted by $\mathrm{av}_C$, $\mathrm{var}_C$, $\mathrm{min}_C$, and $\mathrm{max}_C$, and the corresponding quantities within $S$ be denoted by $\mathrm{av}_S$, $\mathrm{var}_S$, $\mathrm{min}_S$, and $\mathrm{max}_S$. The six features related to the contrast of the microcalcification are defined as: $\mathrm{fc1}=\mathrm{av}_C/\mathrm{av}_S$, $\mathrm{fc2}=\mathrm{av}_C-\mathrm{av}_S$, $\mathrm{fc3}=(\mathrm{av}_C+\mathrm{av}_S)/(\mathrm{av}_C-\mathrm{av}_S)$, $\mathrm{fc4}=(\mathrm{av}_C-\mathrm{av}_S)^2/(\mathrm{var}_C+\mathrm{var}_S)$, $\mathrm{fc5}=(\mathrm{av}_C-\mathrm{min}_C)/(\mathrm{av}_S-\mathrm{min}_S)$, and $\mathrm{fc6}=(\mathrm{max}_C-\mathrm{min}_C)/(\mathrm{max}_S-\mathrm{min}_S)$.

After the features were extracted from each microcalcification in a cluster, the mean, standard deviation, maximum, and coefficient of variation of each feature were calculated over each cluster to define cluster features. The number of microcalcifications in a cluster was also defined as a cluster feature. We had 45 morphological features for each cluster.

The texture feature were extracted from regions of interest (ROIs) containing the cluster using the second-order statistics provided by the spatial gray-level dependence (SGLD) matrix.[14,15] The distribution of the SGLD matrix elements reflects the average spatial relationship of pairs of gray-level values with respect to the distance $d$ and direction $\theta$ used in SGLD matrix construction. To define the ROI, the bounding box of a detected cluster was enlarged by 5 mm (50 pixels) in each direction so that the background of the surrounding tissue could be included in the analysis. The SGLD matrix was computed in four directions ($\theta=0°$, 45°, 90°, and 135°) and three distances ($d=1,2,4$) as described in the literature.[15] From each SGLD matrix, 13 texture features related to the distribution of the matrix elements were extracted, namely, energy, correlation, entropy, inertia, inverse difference moment, sum average, sum variance, sum entropy, difference average, difference variance, difference entropy, information measure of correlation 1, and information measure of correlation 2. The formulation of these texture measures has been described in the literature.[16] Two corresponding features in the diagonal direction ($\theta=45°$ and 135°) were averaged to yield a single feature. Similarly, two corresponding features in the horizontal and vertical directions ($\theta=0°$ and 90°) were also averaged. We thus had a total of $2\times3\times13=78$ texture features for each cluster.

The CNN, based on the neocognitron structure of Fukushima,[17] is a backpropagation neural network that operates on images. The input image is filtered by successive nodes in the hidden layers, where each node consists of a group of trainable weights. The output score is a scalar, ideally indicating the likelihood of a true microcalcification in our application. The CNN has been used extensively for the detection of microcalcifications,[12] as well as for detection of lung nodules[18] and mammographic masses.[19] The network architecture and weights were trained previously using a training set that was independent of the data set used in this study.[20] In our previous work, the CNN was used for each individual microcalcification before the clustering stage. In this study, we used a low CNN threshold before the clustering stage to exclude only very obvious FPs. After clustering, four CNN score features were extracted for each cluster, namely, the average, standard deviation, maximum, and minimum of the individual microcalcification scores within the cluster. These features were used to define cluster CNN scores to be used for FP reduction.

## D. Single-view and similarity classifiers

The above-described features were used in two different classifiers to differentiate true microcalcification clusters from FPs. The first classifier was a single-view classifier that used the features extracted from a cluster on a particular view (CC or MLO). The second classifier was a similarity classifier that jointly used the features of two cluster candidates on two views.

The single-view classifier was trained using stepwise feature selection[21] and linear discriminant analysis (LDA)[22] on the training set. LDA with stepwise feature selection has been previously used in CAD for several applications, including FP reduction for mass detection on mammograms, classification of masses as malignant or benign, classification microcalcification clusters as malignant or benign, and FP reduction for lung nodule detection on CT scans. Stepwise feature selection involves the selection of three parameters, namely, $F_{\mathrm{in}}$, $F_{\mathrm{out}}$, and tolerance (tol). A discussion of how these parameters are related to the feature selection process and how they affect the performance can be found in the literature.[21] In this study, we used a leave-one-case-out method and the area under the receiver operating characteristic curve ($A_z$) as a figure-of-merit within the training set to determine the best values of these parameters that could provide high classification accuracy with a relatively small number of features. Once they were determined, we used the chosen set of parameters to select a final set of features and LDA coefficients using the entire training set. The feature space available for stepwise selection included 127 features, of which 45 were morphological, 78 were texture, and 4 were CNN score features. Note that only a small subset of the available features will be selected during the classifier design process using the training set.

Our two-view classification algorithm is designed to distinguish between true (TP-TP) pairs and false (FP-TP, TP-FP or FP-FP) pairs by using the similarities between the two

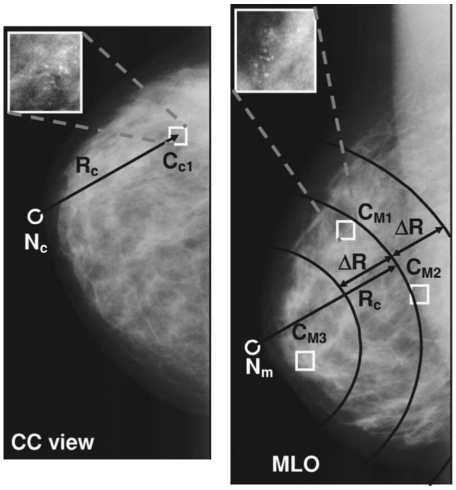

FIG. 8. Geometric pairing of the clusters detected on the CC and MLO views. For a cluster $C_{C1}$ on the CC view, the nipple-to-object distance $R_c$ is computed. On the MLO view, any object that falls within the annular region delineated by the two concentric arcs $R_c+\Delta R$ and $R_c-\Delta R$ centered on the nipple location is paired with the cluster candidate $C_{C1}$ on the CC view. In this example, two pairs are defined: $C_{C1}$-$C_{M1}$ pair and $C_{C1}$-$C_{M2}$ pair. Although a third cluster candidate exists on the MLO view, it is not paired with $C_{C1}$ because it falls outside the defined annular region. The half-width $\Delta R$ of the annular region is determined using the training data set.

objects on different views that constitute the pair. The initial step in this task is to define the object pairs. For a deformable object like the breast under compression, the corresponding locations in the two views cannot be determined exactly based on the two projection mammograms. From the geometry of the mammographic image acquisition, it is known that an object seen on the CC view can appear only in a limited subregion in the MLO view, and vice-versa. Radiologists at our institution routinely use the NOD on the two views to estimate the correspondence between objects seen on different views of the same breast. Based on our previous studies,[3,6] in this work we searched for the member clusters of a pair based on the difference between the NODs on the CC and MLO views. Figure 8 illustrates this geometric pairing procedure. First, the nipple locations $N_c$ and $N_m$ are determined on the CC and MLO views, respectively. Next, the nipple-to-object distance $R_c$ is computed for the cluster candidate $C_{C1}$ on the CC view. To find objects on the MLO view to be paired with $C_{C1}$, an arc of radius $R_c$ centered at $N_m$ is defined on the MLO view. Next, two concentric arcs, with radii $R_c+\Delta R$ and $R_c-\Delta R$ are also drawn on the MLO view. Any object that falls within the annular region delineated by these two concentric arcs is paired with the cluster candidate $C_{C1}$ on the CC view. In this example, two pairs are defined: $C_{C1}$-$C_{M1}$ pair and $C_{C1}$-$C_{M2}$ pair. Although a third cluster candidate $C_{M3}$ exists on the MLO view, it is not paired with $C_{C1}$ because it falls outside the defined annular region. The half-width of the annular region $\Delta R$ is determined using the training data.

A similarity classifier was designed to score the defined pairs as to their likelihood of being a true pair. For the similarity classifier, two sets of features were generated from the feature set that was used in the single-view classifier. The first set of features consisted of the squared difference of the corresponding features of the two objects in a pair. This set is thus designed to extract the similarity (or dissimilarity) of the features between the detected objects on the two views. For example, one would expect that for a true pair, the average eccentricity of the microcalcifications in the same cluster on the CC and MLO views would be close to each other. For a microcalcification cluster paired with a FP cluster, the difference may be large. The second set of features consisted of the average of the corresponding features. For example, for a true pair, the average CNN score of the individual microcalcifications in the cluster is expected to be large on both views, therefore the corresponding average feature for the pair would also be large. If the microcalcification cluster is paired with a FP cluster, or if two FP clusters are paired, then the CNN score is expected to be small for at least one of the clusters, and therefore the corresponding average feature for the pair would be relatively small. In addition to these two feature sets, we included the NOD difference between the paired clusters as another feature measure. The feature pool for the similarity classifier thus contained a total of $2\times 127+1=255$ features. The classifier was trained using stepwise feature selection and LDA on the training set, following the same steps described earlier for the single-view classifier.

## E. Fusion

The similarity classifier produced a score for each cluster pair. These scores were converted into scores for each individual cluster before being combined with the single-view classifier scores. A cluster on the CC view can be a member of several cluster pairs (paired with more than one cluster on the MLO view), and vice-versa, as illustrated in Fig. 8. The two-view cluster score of a cluster $C_i$ was defined as the maximum of all similarity scores for pairs in which $C_i$ is a member, if that maximum value exceeded a paired-cluster threshold, $\mathrm{th}_p$, determined using the training set. If a cluster $C_i$ was not paired with any cluster on the other view, or if the maximum value was below $\mathrm{th}_p$, then it was assigned a large negative two-view cluster score, chosen arbitrarily as $-100$ in this study. The idea behind using the paired-cluster threshold $\mathrm{th}_p$ is that if a cluster is geometrically paired with a cluster on the other view, but the evidence for the similarity of the two clusters is weak, then it is likely a false pair and may be eliminated.

By using a constant paired-cluster threshold, and varying the decision threshold on the similarity scores, one can obtain a FROC curve for the detection of microcalcification clusters. Since this detection method uses only the paired information from the two mammographic views, it is termed the exclusive two-view detection method in the following discussion. To utilize both the one-view and two-view information, an effective fusion method has to be designed. In this study, we found that good fusion performance could be achieved for a cluster by averaging its single-view score and exclusive two-view detection score.

### F. Evaluation

The accuracy of the single-view and similarity classifiers was evaluated using ROC analysis.[23] A cluster was considered to be a TP if its bounding-box overlapped the radiologist-defined microcalcification location by more than 40%. Other detected clusters were considered as FPs. The overall detection accuracy of the two systems was compared using free response ROC (FROC) analysis. The sensitivity axis in FROC analysis was based on the positive data set. The number of FPs per mammogram was estimated using the normal data set. For the test set, we plotted two types of FROC curves, using mammogram-based and cased-based analyses. In the mammogram-based analysis, the same cluster seen on two views are counted independently. A TP was defined as a true microcalcification cluster detected by the computer. Since there were 218 microcalcification cluster locations marked by the radiologist on the test set, the denominator for sensitivity in mammogram-based analysis was 218. If a radiologist-marked cluster was detected as more than one cluster by the computer, they would be counted only as one TP. In the case-based analysis, a TP was defined as a positive case for which a cluster was correctly detected on one or both views. Since we had 96 mammogram pairs in our positive test set, the denominator for sensitivity in case-based analysis was 96. If more than one TP cluster was detected for a case, this was counted as only one TP.

## III. RESULTS

The prescreening algorithm detected an average of 3.06 clusters per mammogram in the normal training (USF) group, and 4.11 clusters in the normal test (UM) group. On the positive training (USF) group, 89% (226/254) of the true cluster locations were detected by prescreening, while the corresponding sensitivity for the positive test (UM) group was 93% (202/218).

To determine the parameters to be used in feature selection for the single-view classifier, we used a leave-one-case-out method within the training set. Mammograms corresponding to each case were left out once as a validation sample, feature selection and LDA design were performed on the rest of the training set, and the LDA scores were obtained for the clusters in the validation samples. After the validation scores were obtained for each case, these scores were pooled and an $A_z$ value was derived from ROC analysis. Table II shows the average number of selected features and the validation $A_z$ values obtained using the leave-one-case-out method on the training set for different values of the parameters used in feature selection. It can be seen that the $A_z$ value was not very sensitive to these parameters in the range that we studied. The parameters for feature selection were therefore selected as $F_{\text{in}}=9.4$, $F_{\text{out}}=9.2$, and tol=0.01, which provided the highest $A_z$ with the smallest number of features. When these parameters were applied to the entire training set, a total of 8 features were selected. Four of these were CNN features, two were texture features, two were contrast features, and one was the number of microcalcifications.

TABLE II. The area $A_z$ under the ROC curve and the number of selected features for different values of stepwise feature selection parameters used in the design of the one-view classifier. The area $A_z$ was obtained from the left-out validation samples in a leave-one-case-out resampling within the training data set. The number of selected features represents the average number of selected features in each cycle of the leave-one-out process.

| $F_{\text{in}}$ | $F_{\text{out}}$ | Tol | $A_z$ | Number of selected features |
|---|---|---|---|---|
| 9.4 | 9.2 | 0.0001 | 0.89±0.01 | 8 |
| 9.4 | 9.2 | 0.01 | 0.89±0.01 | 8 |
| 11.4 | 11.2 | 0.0001 | 0.88±0.01 | 7 |
| 11.4 | 11.2 | 0.01 | 0.88±0.01 | 7 |
| 7.4 | 7.2 | 0.0001 | 0.89±0.01 | 9 |
| 7.4 | 7.2 | 0.01 | 0.89±0.01 | 9 |
| 5.4 | 5.2 | 0.0001 | 0.87±0.01 | 12 |
| 5.4 | 5.2 | 0.01 | 0.87±0.01 | 11 |

Figure 9 shows the NOD differences for the CC and MLO view mammograms for the positive training and test data sets. Based on the histogram of the training data set, we selected $\Delta R=26$ mm for the geometric pairing of the data. As can be observed from Fig. 9, this choice for $\Delta R$ resulted in 4 true missed pairs on the training set, and 5 true missed pairs on the test set. On the other hand, a large fraction of false pairs were eliminated. For the training set, there were a total of 2843 pairs that could be defined on the normal mammograms without the geometric constraint. The geometric pairing using $\Delta R=26$ mm eliminated 62% (1750/2843) of these pairs, resulting in 1093 false pairs. Similarly, for the normal test set, the geometric pairing eliminated 61% (1183/1943) of the possible false pairs, resulting in 760 pairs.

The feature selection parameters for the similarity classifier were also determined using a leave-one-case-out method within the training set. Table III shows the average number of selected features and the validation $A_z$ values obtained using the leave-one-case-out method on the training set for different values of the parameters used in feature selection. Similar to the results for the single-view classifier, the $A_z$ value was not very sensitive to the feature selection param-

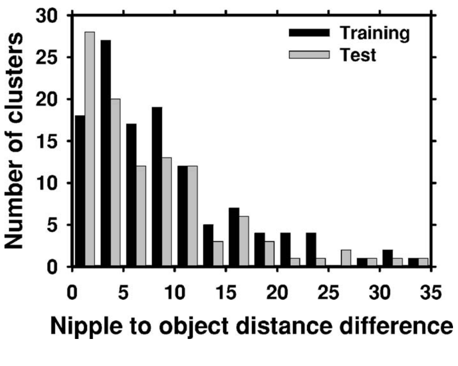

FIG. 9. The distribution of the NOD differences for the true cluster centroids on the CC and MLO views. Based on the distribution of the NOD differences for the training data set, the half-width $\Delta R$ of the annular region was selected as $\Delta R=26$ mm.

TABLE III. The area $A_z$ under the ROC curve and the number of selected features for different values of stepwise feature selection parameters used in the design of the two-view classifier. The area $A_z$ was obtained from the left-out validation samples in a leave-one-case-out resampling within the training data set. The number of selected features represents the average number of selected features in each cycle of the leave-one-out process.

| $F_{in}$ | $F_{out}$ | tol | $A_z$ | Number of selected features |
|---|---|---|---|---|
| 35.4 | 35.2 | 0.0001 | 0.96±0.003 | 13 |
| 35.4 | 35.2 | 0.01 | 0.96±0.003 | 13 |
| 45.4 | 45.2 | 0.0001 | 0.95±0.003 | 10 |
| 45.4 | 45.2 | 0.01 | 0.95±0.003 | 9 |
| 25.4 | 25.2 | 0.0001 | 0.96±0.003 | 16 |
| 25.4 | 25.2 | 0.01 | 0.96±0.003 | 16 |
| 15.4 | 15.2 | 0.0001 | 0.96±0.003 | 18 |
| 15.4 | 15.2 | 0.01 | 0.96±0.003 | 18 |
| 5.4 | 5.2 | 0.0001 | 0.95±0.003 | 47 |
| 5.4 | 5.2 | 0.01 | 0.95±0.004 | 37 |

eters in the range that we studied. Based on these results, the parameters for feature selection were chosen as $F_{\text{in}}=35.4$, $F_{\text{out}}=35.2$, and tol=0.01. When these parameters were applied to the entire training set, a total of 13 features were selected, of which 8 were squared difference features and 5 were average features. Of the squared difference features, three were from contrast features, two from CNN features, one from a texture feature, one from a shape feature, and one from the number of microcalcifications. Of the average features, two were contrast features, one was a CNN feature, one was a texture feature, and one was the number of microcalcifications.

To select the paired-cluster threshold, the resubstitution FROC curves for the training set were plotted for different selected values of the threshold. Figure 10 shows these FROC curves for five of the selected thresholds. Based on these plots, the paired-cluster threshold was selected as $\text{th}_p=0.0$.

The designed classifiers were applied to the test set. Figure 11 presents the mammogram-based test FROC curves for the single-view and exclusive two-view detection methods. The comparison shows that the exclusive two-view detection method had a very low FP rate for the test set at low sensitivity. However, the highest per-film sensitivity that could be reached by exclusive two-view detection was 69%. In contrast, single-view detection could reach a sensitivity of 93% (not shown in the figure), but at a high FP rate. Figure 12(a) shows the FROC curve when the scores of these two classifiers were fused. The result indicates that at relatively high FP rates, the two-view fusion classifier behaved essentially as the single-view classifier. However, at lower FP rates, the sensitivity of the two-view fusion classifier was much higher. To study whether the improvement was significant, we applied JAFROC analysis.[24] The figure-of-merit (FOM) from the output of the JAFROC software was 0.85 and 0.81, respectively, for two-view fusion and single-view detection. The difference between the FOM was statistically significant ($p=0.0002$). Figure 12(b) compares the case-based FROC curves for the single-view detection and two-view fusion ap-

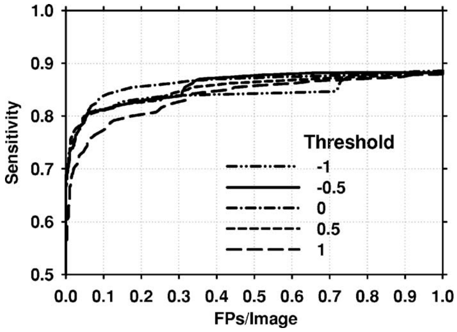

FIG. 10. Mammogram-based resubstitution FROC curves for different values of the paired-object threshold $th_p$. Based on these FROC curves, the paired-cluster threshold was selected as $th_p=0$.

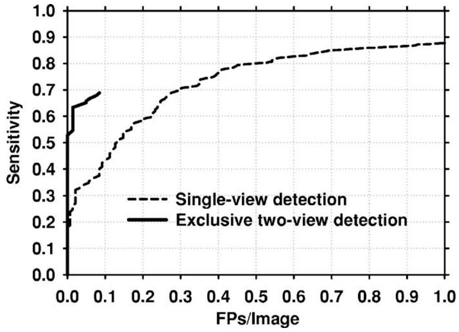

FIG. 11. Mammogram-based test FROC curves for the single-view and exclusive two-view detection methods. The exclusive two-view detection method had a very low FP rate, but could only reach a maximum sensitivity of 69%. The single-view detection method had a higher maximum sensitivity, but had a higher FP rate than the exclusive two-view detection method at low sensitivity.

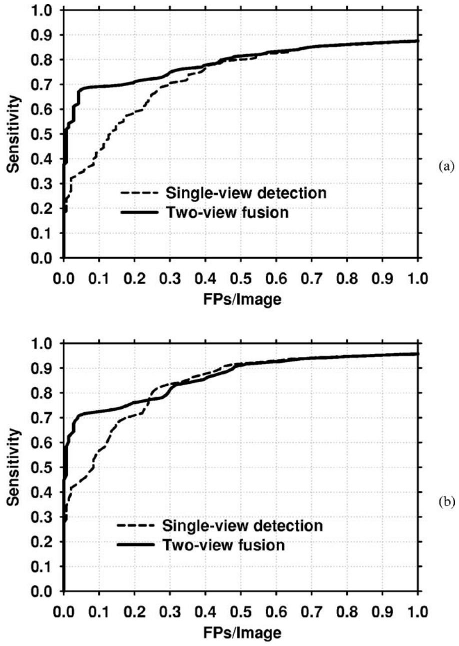

FIG. 12. FROC curves for the single-view and two-view fusion methods for the entire test data set. (a) Mammogram-based, (b) Case-based.

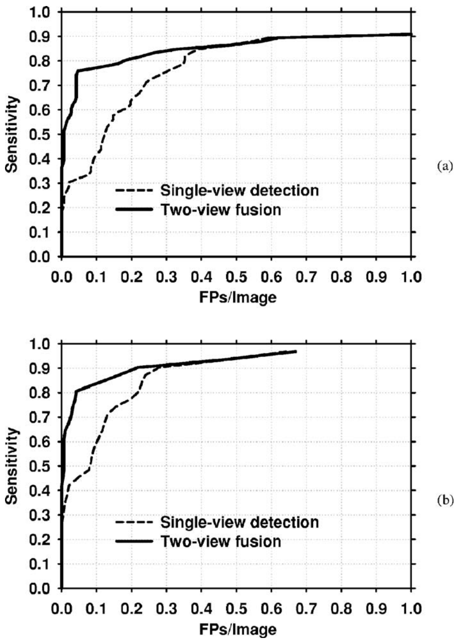

FIG. 13. FROC curves for the malignant test set. (a) Mammogram-based, (b) Case-based.

proach. It can be seen again that two-view fusion resulted in a higher FROC curve.

The improvement with two-view fusion was also analyzed for the subsets of malignant and benign test cases. Figures 13(a) and 13(b) compare the single-view and two-view fusion methods for the malignant cases using mammogram-based and case-based analysis, respectively. Figures 14(a) and 14(b) present the corresponding curves for the benign cases. The detection performance with both single-view detection and two-view fusion is higher for malignant clusters than for benign clusters. Two-view fusion results in a greater improvement in performance for malignant clusters than for benign clusters. The improvement with two-view fusion was found to be significant for both malignant ($p=0.0007$) and benign ($p=0.0013$) clusters using JAFROC analysis. Table IV compares the FP rates for single-view detection and two-view fusion at selected sensitivities for the entire data set, the subset of malignant clusters, and the subset of benign clusters.

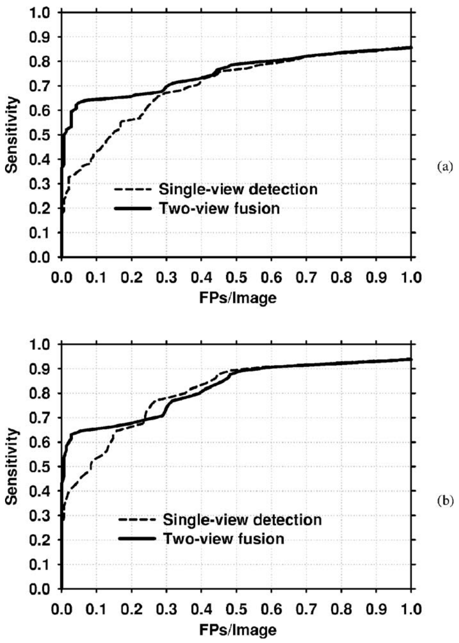

FIG. 14. FROC curves for the benign test set. (a) Mammogram-based, (b) Case-based.

We also studied whether two-view fusion would be effective when the training and test sets were switched. For this purpose, we designed the single-view and similarity classifiers and the paired-cluster threshold using the UM data set, and applied the designed classifiers to the USF data set. The resulting mammogram-based FROC curves are shown in Fig. 15. The FROC curve for the two-view fusion classifier is again higher than that of the single-view classifier. At 80% sensitivity, the single-view and the two-view fusion classifiers had an average FP rate of 0.18 and 0.06 FPs/mammogram, respectively. It is also observed that these FP rates are lower than the corresponding mammogram-based rates obtained using the UM data set as the test set (Table IV).

## IV. DISCUSSION

Our results indicate that the two-view classifier and the fusion of the two-view scores with single-view scores improved the microcalcification detection accuracy of our CAD system. A significant improvement in the mammogram-based test FROC curves was achieved using two-view fusion compared to single-view detection. Much of this improvement resulted from the lower FP rates of our similarity classifier at lower sensitivities. At high sensitivity (above 80%), the mammogram-based FROC curves of the two-view fusion and single-view detection methods were almost identical. The improvement was also significant for the subsets of malignant and benign clusters of our test data set.

There were 104 microcalcification clusters seen on both views in the abnormal test group. Of these, 90 (87%) were matched with the corresponding cluster after geometric pairing. Of the 14 missed matches, five were caused by the absolute value of the NOD being larger than $\Delta R=26$ mm, and the remaining nine were caused by either one or both of the clusters being missed by the prescreening algorithm. The paired-cluster threshold, $\mathrm{th}_p$, eliminated an additional 16 matches, resulting in a total of 74 correctly matched microcalcification clusters for the exclusive two-view detection. The highest sensitivity of exclusive two-view detection was 69% (150/218), because the scores of two of the true clusters that were matched with false clusters in the corresponding view were above $\mathrm{th}_p$. In return for missing 16 out of a total of 104 true cluster pairs (15% miss), a substantial number of FP pairs were eliminated with the use of the paired-cluster threshold. After geometric pairing, we had a total of 760 false pairs in the normal test group. The use of the

TABLE IV. The average number of FPs at selected sensitivities for the single-view and two-view fusion detection methods. The average number of FPs is compared for the entire test data set, as well as the malignant and benign subsets. For computing the sensitivity in mammogram-based scoring, the same cluster seen on two views was counted independently. Since there were 218 microcalcification cluster locations marked by the radiologist on the test set, the denominator for sensitivity in mammogram-based analysis was 218. In case-based analysis, a TP was defined as a positive case for which a cluster was correctly detected on one or both views. Since we had 96 mammogram pairs in our data positive test set, denominator for sensitivity in case-based analysis was 96.

| | | Average number of FPs per image | | | |
|---|---|---|---|---|---|
| | | 80% sensitivity | | 70% sensitivity | |
| Scoring method | Data subset | Single-view | Fusion | Single-view | Fusion |
| Mammogram-based | Entire test set | 0.53 | 0.46 | 0.30 | 0.18 |
| | Malignant | 0.35 | 0.16 | 0.24 | 0.04 |
| | Benign | 0.63 | 0.56 | 0.37 | 0.30 |
| Case-based | Entire test set | 0.26 | 0.30 | 0.17 | 0.04 |
| | Malignant | 0.22 | 0.04 | 0.13 | 0.03 |
| | Benign | 0.35 | 0.39 | 0.24 | 0.25 |

paired-cluster threshold eliminated 754 of these, resulting in a total of only 6 false pairs, or an average of 0.08 (12/142) FP clusters/normal image.

The paired-cluster threshold in this study was determined using the training set as $\text{th}_p=0$. To evaluate the effect of this threshold on the test results, we also obtained test FROC curves for different values of $\text{th}_p$. The curves for $\text{th}_p=0.5$, 0, −0.5, −1, and −2 are shown in Fig. 16. It is observed that the FROC curve may be higher if thresholds lower than that obtained by training were used. However, as seen in Fig. 10, a lower threshold than $\text{th}_p=0$ resulted in worse performance for the training set. Since the parameters of a CAD system ought to be selected using the training set, and not by comparing the test performance for different parameters, we maintain that unbiased test results are those found by using $\text{th}_p=0$ (Fig. 12).

Comparing mammogram-based test FROC curves with case-based curves, we find that the improvement with two-view fusion is more limited for case-based detection. Figures 12(a) and 14(a) demonstrate that single-view detection may provide slightly lower FP rate for some sensitivity values when case-based scoring is employed. However, at low sensitivity, case-based two-view fusion FROC curves are substantially higher. Since there is currently no accepted method to compare case-based FROC curves, we did not evaluate the statistical significance in the difference between these curves.

Comparing the test FROC curves in Figs. 12(a) and 15, we found that the detection performance for the USF and UM sets are different. For example, at a sensitivity of 85%, the average number of FPs for the USF and UM data sets are 0.17 and 0.71, respectively. Part of this difference may be attributed to the fact that the USF data set contained biopsy-proven malignant cases, whereas the UM data set contained a mixture of malignant and benign clusters. A comparison of the FROC curves for malignant and benign subsets of the UM data set (Figs. 13 and 14) indicates that the CAD system has a higher detection accuracy for malignant microcalcification clusters. However, this does not completely explain the difference between Figs. 12(a) and 15. Figures 13(a) and 15 show that the detection accuracy for the USF data set is higher than that of the malignant UM subset. For example, at a sensitivity of 85%, the average number of FPs for the USF and malignant UM data sets are 0.17 and 0.33, respectively. This indicates that the UM data set may contain more difficult cases than the USF data set. The subtlety ratings of the USF (Fig. 2) and UM (Fig. 4) cases also indicate that the UM cases may be more subtle, although the subtlety rating scales are subjective and may not be easily compared among radiologists. These differences also underscore the difficulty of comparing the performances for algorithms evaluated with data sets collected at different institutions.

Our study had a number of limitations. The nipple locations in our study were hand-extracted. Therefore, the two-view detection process was not entirely automated. We have been developing automated nipple detection methods on mammograms.[25] Joint two-view detection with automatically identified nipple locations will be studied in the future. Cases containing diffuse microcalcifications and calcified vessels

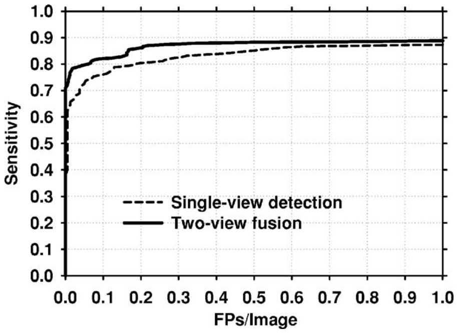

FIG. 15. Single-view and two-view fusion mammogram-based test FROC curves when the training and test sets are switched. To obtain these curves, classifiers were trained on the UM data set. The trained classifiers were then applied to the USF data set.

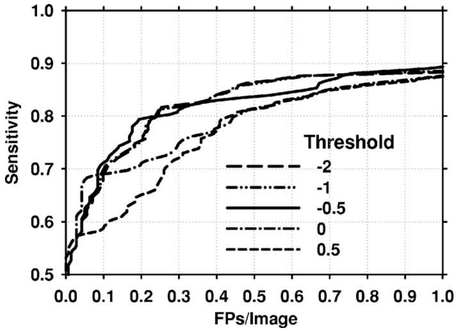

FIG. 16. Mammogram-based FROC curves for different values of the paired-clusters threshold $th_p$. When the value of $th_p$ was set to less than that obtained by training $(th_p<0)$ the FROC curves for the test set were higher than the test FROC curve obtained with $th_p=0$, which was determined by training.

were excluded from both the training and test data sets because of the difficulty of establishing the correspondence of the clusters on the two views. If the two-view detection method were applied to cases containing diffuse microcalcifications, a large number of TP pairs would be identified. Although it may be impossible to judge which pairs truly correspond to each other, it is quite likely that some of the pairs will attain a high two-view score, and thus result in a correct detection. If the two-view detection method were applied to normal cases containing calcified vessels, again a large number of FP pairs would be detected. A reasonable method to handle cases containing calcified vessels may be to automatically identify such vessels after preprocessing, apply only single-view detection, and mark the detected calcified vessels in a special way (e.g., different color than other detections). In this study, we did not attempt any type of automated detection for calcified vessels.

A number of parameters in this study were optimized based on the training set, such as the feature selection thresholds in the LDA design and the paired-cluster threshold. However, we did not perform a joint optimization of many other parameters, such as the width of the annular region used in geometric pairing, the thresholds used in prescreening, or the weights for fusion of the two classifiers. A methodical optimization of the large number of parameters in the single-view and two-view detection algorithms may improve the final classification accuracy. Any improvement in the geometric pair definition will also improve the joint two-view detection.

## V. CONCLUSION

We have developed a joint two-view detection method to improve the computerized detection of microcalcification clusters on two-view mammograms. A geometric method was used to pair the clusters from the two views. A two-view classifier was designed to distinguish between true and false pairs by using the similarities between the two clusters on different views that constitute the pair. The scores of the two-view similarity classifier were fused with the single-view cluster scores obtained from a conventional classifier designed for differentiating true and false clusters on one view. Our results indicate that two-view fusion may significantly reduce FP clusters and improve the FROC curve of microcalcification detection for both malignant and benign clusters. The improvement for malignant clusters was more substantial than that for benign clusters. Directions for future work include automated identification of nipple locations, and optimization of parameters used in single-view and two-view detection.

## ACKNOWLEDGMENTS

This work is supported by USPHS Grant No. CA95153 and U. S. Army Medical Research and Materiel Command Grant No. DAMD17-02-1-0214. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program, and Dev Chakraborty, Ph.D, for the JAFROC program.

[a)]Electronic mail: berki@umich.edu

[1] E. Thurfjell, A. Taube, and L. Tabar, “One-view versus 2-view mammography screening—A prospective population-based study,” Acta Radiol. **35**, 340–344 (1994).

[2] R. Warren, S. Duffy, and S. Bashir, “The value of the second view in screening mammography,” Br. J. Radiol. **69**, 105–108 (1996).

[3] S. Paquerault, N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, “Improvement of computerized mass detection on mammograms: Fusion of two-view information,” Med. Phys. **29**, 238–247 (2002).

[4] B. Sahiner, M. N. Gurcan, H. P. Chan, L. M. Hadjiiski, N. Petrick, and M. A. Helvie, “The use of joint two-view information for computerized lesion detection on mammograms: Improvement of microcalcification detection accuracy,” Med. Phys. **4684**, 754–761 (2002).

[5] S. Paquerault, N. Petrick, H. P. Chan, B. Sahiner, and A. Y. Dolney, “Improvement of mammographic lesion detection by fusion of information from different views,” Proc. SPIE **4322**, 1883–1889 (2001).

[6] S. Paquerault, B. Sahiner, N. Petrick, L. M. Hadjiiski, M. N. Gurcan, C. Zhou, and M. A. Helvie, in *Prediction of Object Location in Different Views Using Geometrical Models*, Toronto, Canada, 2001 (Medical Physics, Madison, WI, 2001), pp. 748–755.

[7] M. Yam, M. Brady, R. Highnam, C. Behrenbruch, R. English, and Y. Kita, “Three-dimensional reconstruction of microcalcification clusters from two mammographic views,” IEEE Trans. Med. Imaging **20**, 479–489 (2001).

[8] Y. Kita, R. P. Highnam, and J. M. Brady, “Correspondence between different view breast X rays using curved epipolar lines,” Comput. Vis. Image Underst. **83**, 38–56 (2001).

[9] Y. H. Chang, W. F. Good, J. H. Sumkin, B. Zheng, and D. Gur, “Computerized localization of breast lesions from two views—An experimental comparison of two methods,” Invest. Radiol. **34**, 585–588 (1999).

[10] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer, R. Moore, K. Chang, and S. Munishkumaran, in *Digital Mammography*, edited by N. Karssemeijer, M. Thijssen, J. Hendriks, and L. van Erning (Kluwer Academic, Dordrecht, 1998), pp. 457–460.

[11] H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, “Improvement in radiologists, detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis,” Invest. Radiol. **25**, 1102–1110 (1990).

[12] H. P. Chan, S. C. B. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, “Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network,” Med. Phys. **22**, 1555–1567 (1995).

[13] H. P. Chan, B. Sahiner, K. L. Lam, N. Petrick, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, “Computerized analysis of mammographic microcalcifications in morphological and texture feature space,” Med. Phys. **25**, 2007–2019 (1998).

[14] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Texture features for

image classification," IEEE Trans. Syst. Man Cybern. **SMC-3**, 610–621 (1973).

[15] H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, K. L. Leung, D. D. Adler, and M. M. Goodsitt, "Computerized classification of malignant and benign microcalcifications on mammograms: Texture analysis using an artificial neural network," Phys. Med. Biol. **42**, 549–567 (1997).

[16] D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," Med. Phys. **24**, 903–914 (1997).

[17] K. Fukushima, S. Miyake, and T. Ito, "Neocognitron: A neural network model for a mechanism of visual pattern recognition," IEEE Trans. Syst. Man Cybern. **SMC-13**, 826–834 (1983).

[18] S. C. Lo, S. L. Lou, J. S. Lin, M. T. Freedman, and S. K. Mun, "Artificial convolution neural network techniques and applications to lung nodule detection," IEEE Trans. Med. Imaging **14**, 711–718 (1995).

[19] B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images," IEEE Trans. Med. Imaging **15**, 598–610 (1996).

[20] M. N. Gurcan, B. Sahiner, H. P. Chan, L. M. Hadjiiski, and N. Petrick, "Selection of an optimal neural network architecture for computer-aided detection of microcalcifications—Comparison of automated optimization techniques," Med. Phys. **28**, 1937–1948 (2001).

[21] N. R. Draper, *Applied Regression Analysis* (Wiley, New York, 1998).

[22] P. A. Lachenbruch, *Discriminant Analysis* (Hafner, New York, 1975).

[23] J. A. Swets, "ROC analysis applied to the evaluation of medical imaging techniques," Invest. Radiol. **14**, 109–121 (1979).

[24] D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modeling, analysis, and validation," Med. Phys. **31**, 2313–2330 (2004).

[25] C. Zhou, H. P. Chan, C. Paramagul, M. A. Roubidoux, B. Sahiner, L. M. Hadjiiski, and N. Petrick, "Computerized nipple identification for multiple image analysis in computer-aided diagnosis," Med. Phys. **31**, 2871–2882 (2004).