Interval change analysis to improve computer aided detection in mammography.

Sheila Timp, Nico Karssemeijer

Department of Radiology
University Medical Center Nijmegen
The Netherlands

Corresponding author:

Sheila Timp
Department of Radiology
Radboud University Hospital
Geert Grooteplein Zuid 18
6525 GA Nijmegen
The Netherlands
Tel: +31-24-3619811
Fax: +31-24-3540866

Abstract

We are developing computer aided diagnosis (CAD) techniques to study interval changes between two consecutive mammographic screening rounds. We have previously developed methods for the detection of malignant masses based on features extracted from single mammographic views. The goal of the present work was to improve our detection method by including temporal information in the CAD program. Toward this goal, we have developed a regional registration technique. This technique links a suspicious location on the current mammogram with a corresponding location on the prior mammogram. The novelty of our method is that the search for correspondence is done in feature space. This has the advantage that very small lesions and architectural distortions may be found as well. Following the linking process several features are calculated for the current and prior region. Temporal features are obtained by combining the feature values from both regions. We evaluated the detection performance with and without the use of temporal features on a data set containing 2873 temporal film pairs from 938 patients. There were 589 cases in which the current mammogram contained exactly one malignant mass. Cross validation was used to partition the data set into a train set and a test set. The train set was used for feature selection and classifier training, the test set for classifier evaluation. FROC (free response operating characteristic) analysis showed an improvement in detection performance with the use of temporal features.

1 Introduction

Currently most CAD (computer aided diagnosis) programs in mammography use a single view to detect abnormalities. However, when mammograms from multiple examinations are available, and CAD makes use of correlations between exams, a higher accuracy may be achieved in detecting malignancies. There have been some studies to the effect of using multiple views in CAD programs. These studies combine information from medio lateral oblique and cranio caudal views [1, 2], from left and right views [3, 4, 5], or from previous and current views [6, 7, 8, 9]. In this study we concentrate on using information from previous and current views.

There are some known advantages of using previous mammograms. First, comparing the current mammogram with mammograms from previous screening rounds may bring attention to subtle signs of malignancy such as a small mass or new or increasing calcifications [10]. These changes might have been overlooked if the previous mammogram was not available for comparison. Radiologists often use this technique to detect developing abnormalities. This technique can also be used for CAD programs to increase the number of true positive detections. Second, suspicious regions on the current mammogram can be evaluated more precisely when the region is compared with the corresponding region on the previous mammogram. For example, if a mass is detected in the current mammogram, a radiologist can use the previous mammogram to determine whether this is a new or existing density. If the mass was already visible on the prior, the radiologist can compare the size and the contrast of both lesions. He will pay attention to signs of malignancy such as an increase in size or contrast [11]. In this study we incorporated this knowledge in our CAD system to improve the classification of regions as normal or abnormal.

A third advantage of using prior mammograms for CAD systems is that additional clues can be found to remove false positive detections. Many false positive detections are caused by mammographic structures that are present in both current and prior mammograms. These structures will have a similar appearance in both mammograms. Examples are crossing vessels and benign lymph nodes. Temporal change analysis can be used to measure the similarity between the region on the current mammogram and the corresponding region on the prior mammogram. If both regions are similar, it is likely that the region represents a false positive or a slowly growing benign mass.

Despite the above mentioned advantages of using previous mammograms the development of CAD systems that include temporal information has not yet received much attention. Previous work can be divided into two main categories: (1) methods that compare current images with priors to detect subtle changes in the breast and (2) methods that compare suspicious regions in current mammograms with corresponding regions in priors. Vujovic et al. used the first method to detect abnormalities. They first divided the current and the prior mammogram into several regions using internal control points. Based on the location of the control points both mammograms were partitioned into statistically homogeneous regions. Circular regions whose centers were determined by distances to the control points were compared using texture and contrast measures. They found that the intensity histogram carried useful information in separating the normal from the abnormal tissue [7]. The second method has been used in [8] and [6]. Kok-Wiles et al. [8] represented the breast as a nested structure of salient regions and used this representation to compare prior and current regions. In a previous study [6] we compared the contrast and size of regions on prior and current mammograms. We found that the difference of these feature values could be used to improve the detection performance of a CAD system.

Both methods depend more or less on the accuracy of the temporal registration. Temporal registration includes global registration and regional registration. Global techniques register current and prior mammograms. In the literature some approaches have been described for global mammogram registration, cf. [12, 13, 14]. A comparative study for global registration

methods in mammography has been done by Van Engeland et al [15]. They compared four methods for mammogram registration: alignment based on nipple position, alignment based on the center of mass of the breast tissue, warping, and registration based on mutual information. The performance of all methods was measured by comparing the distance between the center of the manual segmentation of abnormalities on the previous and the current view before and after registration. They found that the method based on mutual information worked best. The method based on center of mass alignment worked reasonably well, in particular if the pectoral muscle was excluded for center of mass calculation. The method based on nipple alignment only worked if the nipple was visible in profile. The method based on warping performed worst and could cause unrealistic deformations inside the breast area.

Regional registration techniques link regions in current mammograms with corresponding regions in priors. Sanjay-Gopal et al. [16, 17] developed a regional registration technique for the semi-automatic registration of lesions on temporal pairs of mammograms. For each suspicious region on the current they defined a warp-shaped search area on the prior mammogram. For each location inside this search area a correlation measure was calculated. The location with the highest correlation was selected as final estimate for the location on the prior.

In this study we investigated the effect of temporal features on the detection performance. For this purpose we developed a complete method that includes temporal information in a CAD system. This method is summarized in Figure 1. The method starts with mammograms of two consecutive screening rounds: the prior and the current mammogram. On both mammograms the breast area and the pectoral muscle are segmented. A global registration method based on center of mass alignment registers the current and the prior mammogram. Then a pixel level mass detection algorithm assigns each pixel inside the breast area a measure of suspiciousness. This measure represents the likelihood for the presence of a malignant mass. The most suspicious locations on the current image are selected and linked to a corresponding location on the prior. After linking both locations are segmented into regions and features are calculated for each region. The combination of features from both regions results in the so-called temporal features. FROC analysis was done to evaluate the detection performance with and without the use of temporal features.

The remainder of this paper is organized as follows. In section 2 we first briefly discuss the single view CAD program and then explain the proposed temporal CAD program in more detail. Section 3 describes the experiments to evaluate the regional registration technique and the temporal CAD program. In section 4 we present the results of our experiments. Section 5 includes some discussion and comparison with a conclusion in the last section.

2 Methods

Our CAD technique consists of a single view and a temporal part. Sub-section 2.1 describes the single view part that starts with preprocessing an image and application of a pixel level mass detection algorithm to detect suspicious locations, followed by segmentation of the suspicious locations and feature extraction. Sub-section 2.2 explains the three steps of the temporal CAD program: global registration, regional registration and combination of features. Figure 1 gives the outline of the complete method. The last sub-section describes feature selection and classification for both single view and temporal CAD methods.

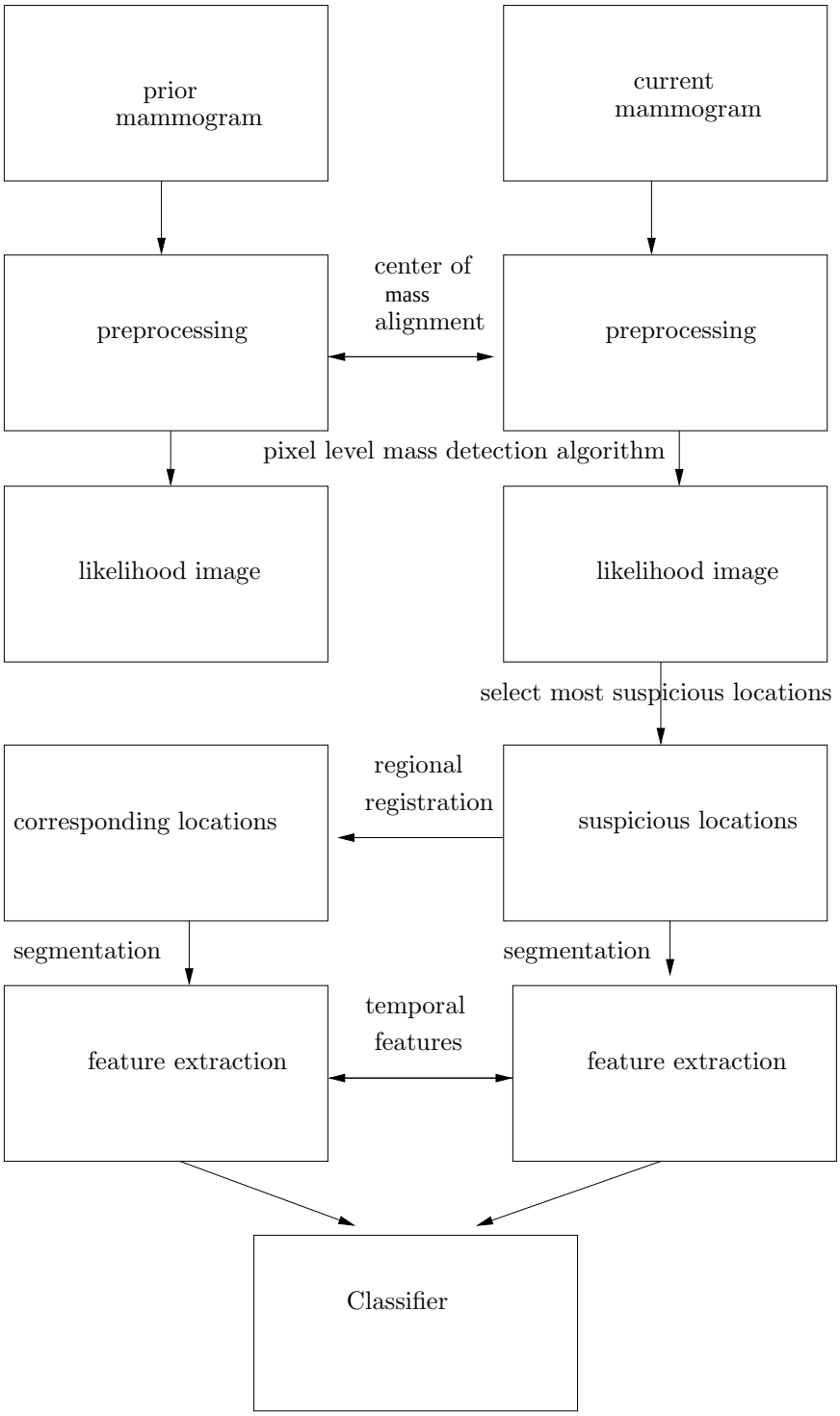


Figure 1: Outline of the temporal CAD method. First the prior and current image or globally aligned. A pixel level mass detection algorithm assign each pixel in the breast area a measure of suspiciousness. The most suspicious locations on the current mammogram are selected and linked to a corresponding location on the prior. Both locations are segmented and features are calculated. Temporal features are obtained by subtracting prior from current feature values. A classifier then assigns each current region a measure of malignancy based on both the original and temporal features.

2.1 Single view CAD method

2.1.1 Preprocessing

Our single view CAD program starts with segmentation the image into breast area, background tissue and pectoral muscle on both current and prior mammograms. For this purpose we use a breast boundary segmentation algorithm developed previously in our group [18]. This algorithm first uses a global thresholding technique to segment the background from mammographic tissue. Then we mark a region of interest (ROI) inside the image in which the pectoral boundary is searched for. Inside this ROI we calculate the gradient magnitude and direction of the image by applying the 3x3 Sobel operator. We transform this gradient image to Hough space in which each peak represents a straight line [19]. The best peak in Hough space is selected as the pectoral boundary. In the next pre-processing step we adjust the gray values of the pectoral region to the gray values of the breast area to make the boundary region more homogeneous. Without this adjustment problems can arise when calculating contrast measurements for tumors that are partly inside and partly outside the pectoral region. The algorithm first calculates the mean gray value of all pixels inside the pectoral region with equal distance to the pectoral edge. Then the pixels inside the pectoral region are normalized as follows:

$$y' = y + \overline{I_0} - \overline{I_d} \tag{1}$$

where y' is the normalized gray value, y the original gray value, $\overline{I_d}$ the mean gray value of all pixels with distance d to the pectoral edge and $\overline{I_0}$ the mean gray value of all pixels that are exactly on the pectoral edge. Finally we apply a peripheral enhancement algorithm to the breast area to correct for differences in tissue thickness. This algorithm first calculates for each pixel the distance to the breast boundary. The maximum distance is d_{max}. We calculate the mean ($\text{mean}_{0.4}$) and minimum ($\text{min}_{0.4}$) gray value of all pixels with a distance $d > 0.4d_{max}$. We then define a threshold $T = (\text{mean}_{0.4} + \text{min}_{0.4})/2$ and adjust all pixels inside the mammogram for which the smoothed gray value $y_s < T$ as follows:

$$y'' = y' + (T - y_s),$$

where y_s is obtained by smoothing the original image with a Gaussian filter with a sigma of 5 mm.

2.1.2 Pixel level mass detection algorithm

After the preprocessing steps a pixel level mass detection algorithm was applied to all pixels inside the breast area. This algorithm selects pixels that display tumor characteristics. To this end two features were calculated for each pixel to detect stellate lesions and two features to detect lesions with a focal mass. A neural network classifier trained on these characteristics assigned each pixel a measure of suspiciousness. We call the classifier output image the *likelihood image*, as it assigns a high value to suspicious locations and vice versa. For details about the algorithm, see [20] and [21]. The most suspicious sites were selected for further processing. If a selected site was found closer than 1 cm to another selected site it was considered to belong to the same suspicious region and the least suspicious site was removed. The method was applied at a high sensitivity level to detect most of the masses in the set. The average number of selected sites per image was 10. Figure 2 and 3 show temporal pairs of images and the corresponding likelihood images. Figure 2 shows a newly developed mass, whereas the mass on Figure 3 was already visible on the previous mammogram. The three most suspicious sites are indicated with their corresponding numbers.

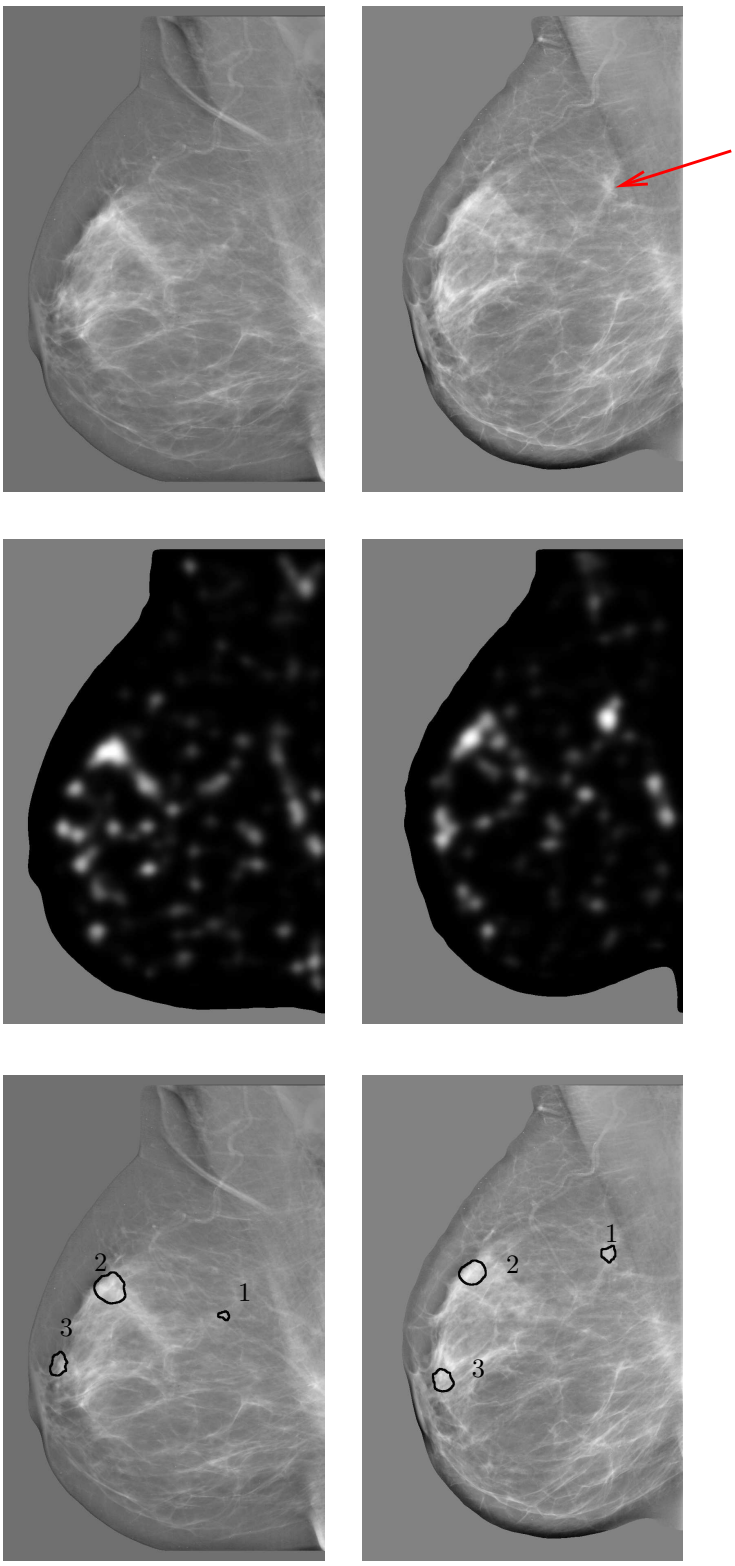
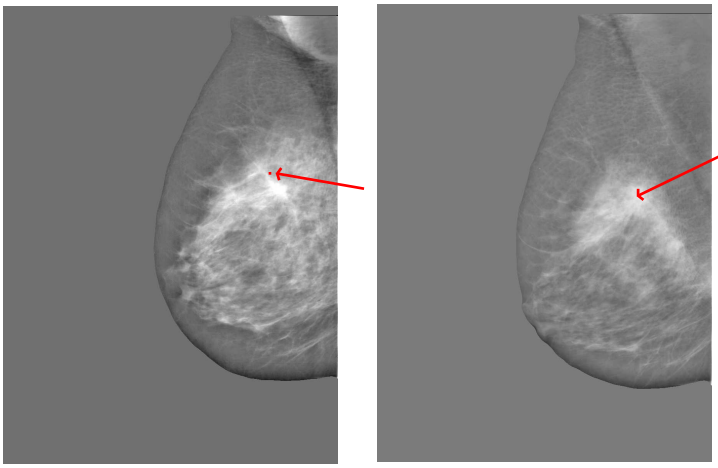
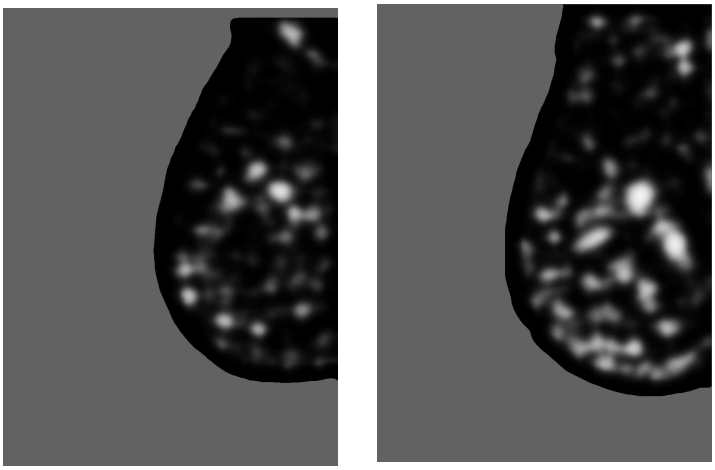


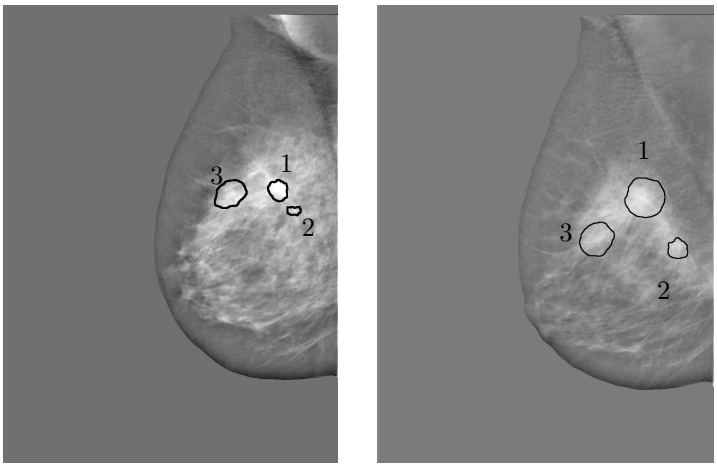
Figure 2: The upper row shows the prior (left) and current image of a newly developing mass. The middle row shows the likelihood images for both prior and current mammograms. The most suspicious locations on the current likelihood image are selected. A regional registration technique links each selected site to a corresponding location on the prior mammogram.



Prior (left) and current image of a growing mass.



Likelihood image: the most suspicious locations are the white spots.



A regional registration techniques links each selected site on the current mammogram to a corresponding location on the prior mammogram.

Figure 3: The upper row shows the prior and the current image of a growing mass. The middle row shows the likelihood images for both prior and current mammograms. The most suspicious locations on the current likelihood image are selected. A regional registration techniques links each selected site to a corresponding location on the prior mammogram.

2.1.3 Segmentation and Feature Extraction

The last step in the CAD program concerns segmentation of the suspicious locations and feature extraction. For segmentation we used an algorithm based on dynamic programming [22]. After segmentation features were calculated for each region. The original feature set consisted of 38 features that are described in more detail in the Appendix. We call these single view features the basic feature set. The features were furthermore grouped into 12 main categories according to the type of characteristic they represent. Table 1 lists the different feature groups.

2.2 Temporal CAD method

The temporal matching step starts with the global registration of previous and current mammograms. Then a regional registration technique is applied to link each suspicious site on the current mammogram with a corresponding site on the prior. After completion of the linking procedure both sites are segmented and features are calculated for each region. Combining features from both views provides temporal information.

2.2.1 Global Registration

The current and the prior image were aligned with a simple procedure based on center of mass alignment. A problem with center of mass alignment is the varying proportion of the pectoral muscle that is visible. Van Engeland et al. [15] found that the registration improved considerably by excluding the pectoral muscle in the center of mass calculation. Therefore we first segmented the pectoral muscle using the Hough transform, see [18]. Then we calculated the center of the breast area for both the prior and the current mammogram with the pectoral muscle excluded. The mammograms were registered using vertical and horizontal translation.

2.2.2 Regional Registration

The next step in the temporal CAD program is regional registration of current and prior locations. Regional registration is a local registration technique that finds for each suspicious site on the current mammogram a corresponding site on the prior mammogram. As both mammograms are globally aligned we can use the coordinates of the lesion in the current mammogram as initial estimate of the location in the prior mammogram. This initial estimate defines the center of a circular search area with radius r as illustrated in Figure 4. For each pixel inside this search area a cost measure is defined. This cost measure is proportional to the square root of the distance from the pixel to the initial estimate and inversely proportional to the value of the likelihood image. The cost for each pixel (i,j) is

$$c(i,j) = -lik(i,j) + \alpha\sqrt{(x^2+y^2)} \tag{2}$$

where $lik(i,j)$ is the value of the likelihood image at pixel (i,j), x and y are the horizontal and vertical distance of (i,j) to the center of the circular search area. The factor α is a weight factor that determines the relative importance of the two terms in the cost function. The pixel with the lowest cost was selected for further processing. In sub-section 4.1 we show the results of our regional registration technique for different values of r and α. Our method was compared with the method from Sanjay et al. [16]. For each suspicious location on the current mammogram they defined a fan-shaped search area in the prior mammogram. A search for a matching structure was performed inside this fan-shaped region. Their search criterion was based on gray scale template matching. We implemented this correlation measure as a search criterion in our registration method and compared it with the search criterion based on the likelihood value.

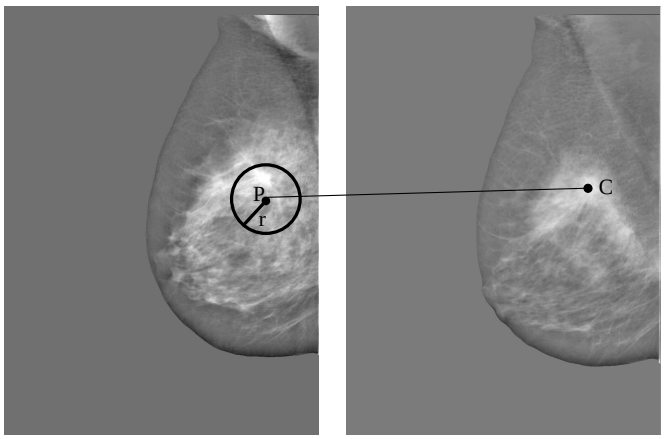


Figure 4: After the global registration each suspicious site C on the current mammogram defines the center P of a circular search area with radius r. The best matching site inside this area is linked to the suspicious location C.

2.2.3 Segmentation and Feature Extraction

The last step in the temporal CAD program concerns segmentation of the linked locations and feature extraction for prior and current regions. The original feature set consisted of 38 features as described in the Appendix. Temporal features were obtained by subtracting prior from current feature values. These features were calculated for all single view features except for location features, dense tissue features and the estimated Wolfe class. This resulted in a total of 29 temporal features. We call the complete set of single view features and temporal features the temporal feature set.

2.3 Classification

Before classification each feature was normalized to zero mean and unit variance, namely each feature f was transformed such that

$$\overline{f} = \frac{f - E(f)}{\sigma(f)},$$

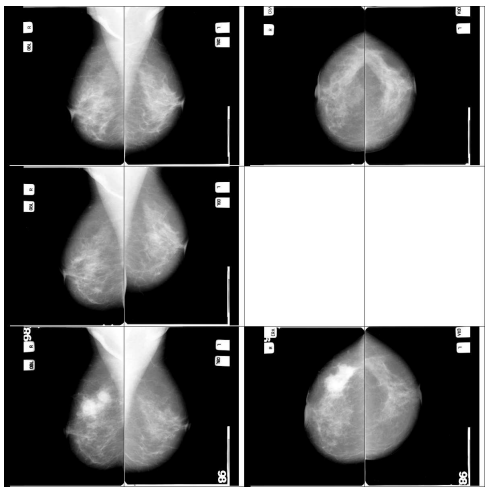
where $E(\cdot)$ and $\sigma(\cdot)$ are the mean and standard deviation of f in the total dataset. The classifier design then consisted of the following two stages: feature selection and classifier training. Both parts were done completely independent from the evaluation of the classifier. A cross validation scheme was used to randomly partition the dataset into a train set and a test set on a 10:1 ratio under the constraint that the images from the same patient were grouped into the same subset.

The first stage of the training involved stepwise feature selection. As feature-selector we used sequential forward floating selection (SFFS) [23]. The train set was used to select the best subset of features.

The second stage involved the training of a simple 3-layer feed-forward neural network classifier based on the train set. By making both feature selection and classifier construction independent of the test set, we aimed at improving the generalizability of our classification results to unknown cases in the patient population.

Group name	Number of features in each group	Temp	Description
Dense tissue	5		Features that determine the location of the region with respect to the dense tissue
Spiculation	4	*	Features that detect spiculated lesions
Mass	4	*	Features that detection a focal mass
Initial likelihood	3	*	Measures of suspiciousness from the first detection step
Intensity	1	*	Mean gray value inside the contour
Contrast	5	*	Difference between the gray level histograms of the region inside and outside the contour
Variance	4	*	Variance in the gray level histogram of the region inside and outside the contour
Texture	5	*	Presence of linear texture
Iso denseness	1	*	Iso denseness of the segmented area
Location	3		Features that determine the location of the region with respect to the pectoral muscle and the skin
Size	1	*	Size of the segmented region
Compactness	1	*	Compactness of the segmented region
Wolfe	1		Estimated Wolfe class

Table 1: Summary of the original features. The features are divided into 12 different groups. The first column gives the group name, the second column the number of features in that group and the last column gives a description of the group. The third column indicates the groups for which temporal features are calculated.



prior II

prior I

diagnostic mammogram

Figure 5: Case example: the current mammogram and mammograms of two previous screening rounds

After training the neural network, the classifier assigned all regions in the test set a probability of malignancy.

3 Experiment

3.1 Dataset

The mammograms used in this study all came from the Dutch Breast Cancer screening Program. All women aged 50-70 are invited biannually to participate in this program. Two mammographic views – medio lateral oblique (MLO) and cranio caudal (CC) – are obtained at the initial screening in this program. At subsequent screenings only medio lateral views are obtained, unless there is an indication that additional cranio caudal views would be beneficial. The mammograms were digitized with either a Canon CFS300 or a Lumisys 85 scanner at a pixel resolution of $50\mu m$ x $50\mu m$, and were averaged to a resolution of $200\mu m$ maintaining the original gray value resolution of 12 bits.

The data set for this study consisted of 4871 single view images obtained from 938 women. From these 4871 single view images, 2873 temporal pairs of images were obtained. The number of temporal pairs was larger than half of the number of the images since for some women the mammograms of three consecutive screening rounds were available: a diagnostic mammogram, an immediate prior mammogram (prior I) and the next prior mammogram (prior II) (see Figure 5).

There were 589 mammogram pairs in which the current mammogram contained exactly one malignant mass. In 44% of these cases the mass was visible on the prior mammogram. This results in 262 temporal images pairs with a visible mass on both current and prior mammograms, and 327 image pairs with a newly developing abnormality on the current mammogram. No pathology was present in 2284 images. We call these images normal. All malignant masses have been manually outlined under supervision of an expert radiologist. A dedicated mammographic review station was used for this purpose.

dataset	nr mass pairs	visible on prior	not visible on prior	diagnostic	prior I
total dataset	589	262	327	407	182
subset visible	262	262	262	195	67
subset not visible	327	327	327	212	115

Table 2: Description of different subsets.

For the experiments we made two subdivisions of the total dataset. The first subdivision is between masses that are visible on the prior and masses that are not visible on the prior. We made this subdivision to study whether temporal features are as useful for new lesions as for existing lesions. The second subdivision is between mammogram pairs in which the current mammogram is the diagnostic one and pairs in which the current mammogram is prior I. The different sets are summarized in Table 2.

3.2 FROC Analysis

Free Response Operating Characteristic (FROC) methodology was used to evaluate the detection accuracy of the total dataset and the different subsets for both the basic feature set and the temporal feature set. A tumor was considered detected if the initial detection location was inside the ground truth. If multiple detections were found inside the same ground truth region they were considered as a single hit. Detections outside the ground truth areas were counted as false positive signals. Only film-based FROC curves were used. Case-based analysis was not done because both cranio-caudal and oblique views were present for two consecutive screening rounds in only few cases.

Furthermore we calculated for each partition, obtained by ten fold cross validation of the original dataset, the area under the FROC curve. We are mainly interested in the detection performance obtained for a low number of false positives per image as this corresponds with normal screening situations. Therefore we used a logarithmic scale for the number of false positives per image and calculated the area under the FROC curve from 0.05 FP/image to 1.0 FP/image. We used the two-sided paired Wilcoxon test with 0.95 confidence level to asses the difference in performance between the two feature sets.

4 Results

4.1 Regional Registration

We evaluated the registration performance on a set of malignant lesions with known ground truth. In this set all lesion were visible on the current and prior mammograms. As evaluation measure we used the percentage of correctly linked locations. If the selected location on the prior was inside the ground truth, the match was considered correct.

Figure 6 shows the performance of the regional registration performance. The y-axis plots the fraction of correctly matched regions. On the x-axis the radius r of the circular search area is plotted. This radius determines the size of the search area on the prior mammogram. This parameter should be adjusted to the accuracy of the global registration. A very small search area would suffice for an almost perfect global registration. However, as registration is a difficult task in mammography, it may be expected that a larger search area is required in combination with a proper regional registration technique. As match criterion we compared the correlation measure from Sanjay [16] and the cost measure given by Eq.2. The correlation

measure is based on similarity between the region on the prior and the current mammogram. The proposed cost measure prefers suspicious locations that are located near the initial estimate. The highest number of correctly matched regions for the proposed method was 72% for $\alpha = 2.0$ and $r = 20mm$. The method based on correlation linked 69% correct for $r = 16mm$.

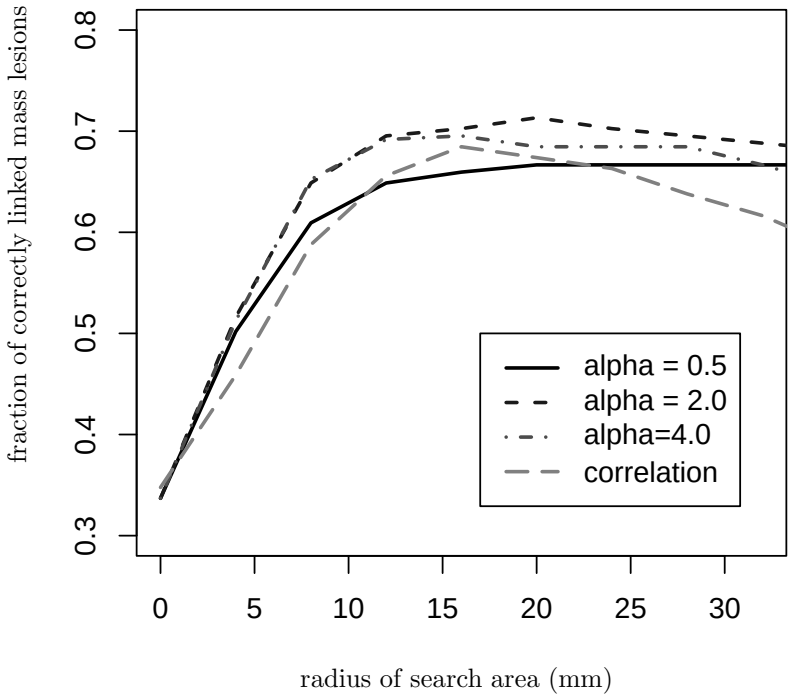


Figure 6: Regional registration results for different values of α compared to the registration method based on template matching. On the horizontal axis the radius of the search area is plotted. The vertical axis shows the fraction of tumors that were correctly linked by each of the regional registration methods.

4.2 Feature Selection

Table 3 and 4 list the number of selected features for each the basic feature set and the temporal feature set. The temporal features that were selected most frequently were difference in contrast, difference in size and difference in initial likelihood. Most of these temporal features were selected instead of their corresponding basic features. For example the feature size is almost always selected in the basic selection, in the temporal selection the difference in size is selected. This can be expected as difference features contain temporal information as well as information about the current region. Table 5 lists some information about the selected temporal features. The mean and standard deviation of the feature value of the current region are calculated for true positives and false positives. Furthermore the difference in feature value between the prior and

Group Name	Number of times a feature is selected
Initial likelihood	20
Location	18
Dense tissue	17
Contrast	16
Spiculation	12
Mass	10
Size	9
Compactness	8
Iso denseness	7
Texture	3

Table 3: Results of the feature-selection process for the original feature set. The first column lists the feature-groups from which at least one feature has been selected. The second column counts the number of features that have been selected from each group.

the current region is displayed. It can be observed that feature values stay more or less unchanged for the false positives, while feature values for true positives change in time. On average true positives are larger in size, have higher likelihood measures and more contrast compared to lesions one screening round earlier. We evaluated the individual performance of each selected temporal feature by calculating the area under the individual ROC curve. For this purpose we first applied our initial detection algorithm to select the most suspicious region in each image. This resulted in 200 false positive regions and 389 true positive regions. We used these regions to construct an ROC curve. The last column in Table 5 gives the area under the ROC curve (A_z value) for each selected temporal feature.

4.3 FROC analysis

Figure 7 shows the mass detection performance for the basic feature set and the temporal feature set. The results are given in FROC curves, where horizontally the number of false positive detections per image is plotted and vertically the sensitivity. The figure shows that temporal features improve the detection performance, especially at a low number of FP detections per image. Table 6 gives the results of the Wilcoxon statistic for the total dataset and the two subsets. The difference in performance between both feature sets is statistically significant.

We furthermore calculated FROC curves for the different subsets. Figure 8 shows the results for masses that are visible cq. not visible on the prior view. We see that masses that are visible on the prior profit more from temporal features than masses that are not visible on the prior. Figure 9 shows the results for mammogram pairs in which the current mammogram is the diagnostic mammogram and pairs in which the current mammogram is a prior I mammogram (see also Figure 5). The overall detection performance for diagnostic mammograms is better than for prior I mammogram. Both subsets show an improvement when temporal features are used. These improvements however are not statistically significant.

Group Name	Number of times a feature is selected
Initial likelihood	20
Location	20
Dense tissue	12
Spiculation	10
Contrast difference	10
Texture	9
Contrast	7
Compactness	7
Size difference	9
Mass	6
Initial likelihood difference	3
Spiculation difference	2
Size	1
Iso denseness	2
Variance	2

Table 4: Results of the feature-selection process for the total feature set. The first column lists the feature-groups from which at least one feature has been selected. The second column counts the number of features that have been selected from each each group.

Feature	false positives				true positives				A_z
	original features		temporal features		original features		temporal features		
	mean	sd	mean	sd	mean	sd	mean	sd	
Size	0.26	0.10	-0.02	0.16	0.54	0.28	0.21	0.34	0.67
Level	1.58	0.11	0.04	0.27	2.04	0.11	0.33	0.25	0.66
Contrast	0.35	0.03	0.01	0.04	0.60	0.09	0.21	0.10	0.75

Table 5: Mean and standard deviation of the selected features for the false positives and the true positives. The last column shows the A_z value for the selected temporal features.

	total dataset	visible on prior	not visible on prior	diagnostic	prior
mean area under FROC basic feature set	0.706	0.762	0.656	0.785	0.45
mean area under FROC temporal feature set	0.721	0.785	0.669	0.797	0.46
p-value Wilcoxon test	0.05	0.05	0.13	0.19	0.38
confidence interval	0.00–0.03	0.02–0.08	-0.01–0.06	-0.01–0.03	-0.01–0.04

Table 6: Results of the Wilcoxon's test for the statistical difference in area under the FROC curve for different datasets. The first two rows gives the mean area under the FROC curve for the temporal and the basic feature set. The third row gives the p-value and the last row the 95% confidence interval.

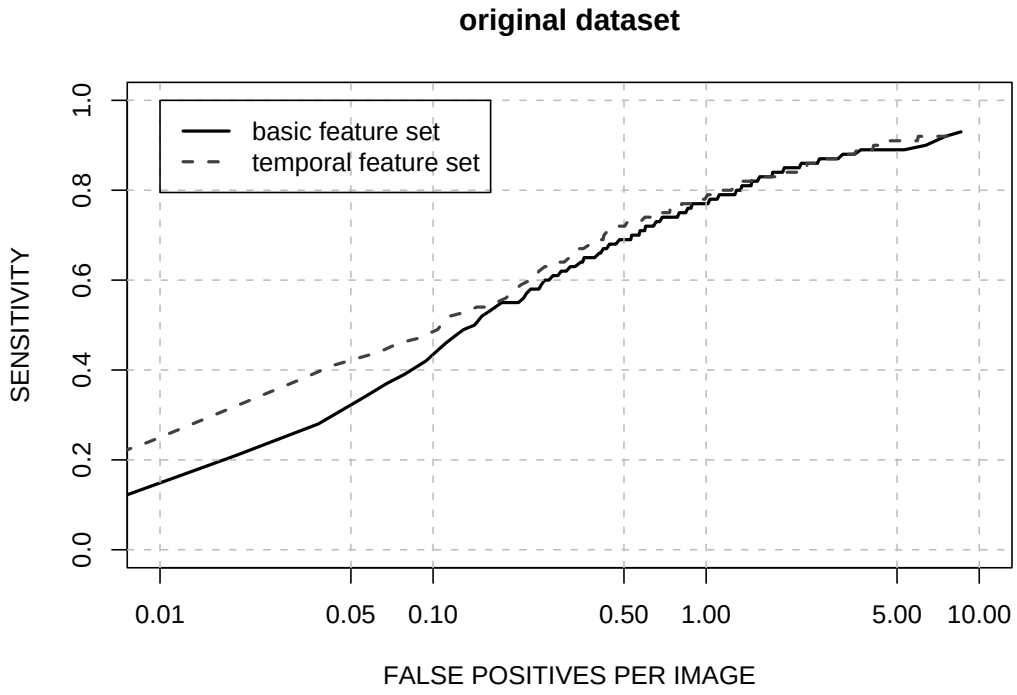
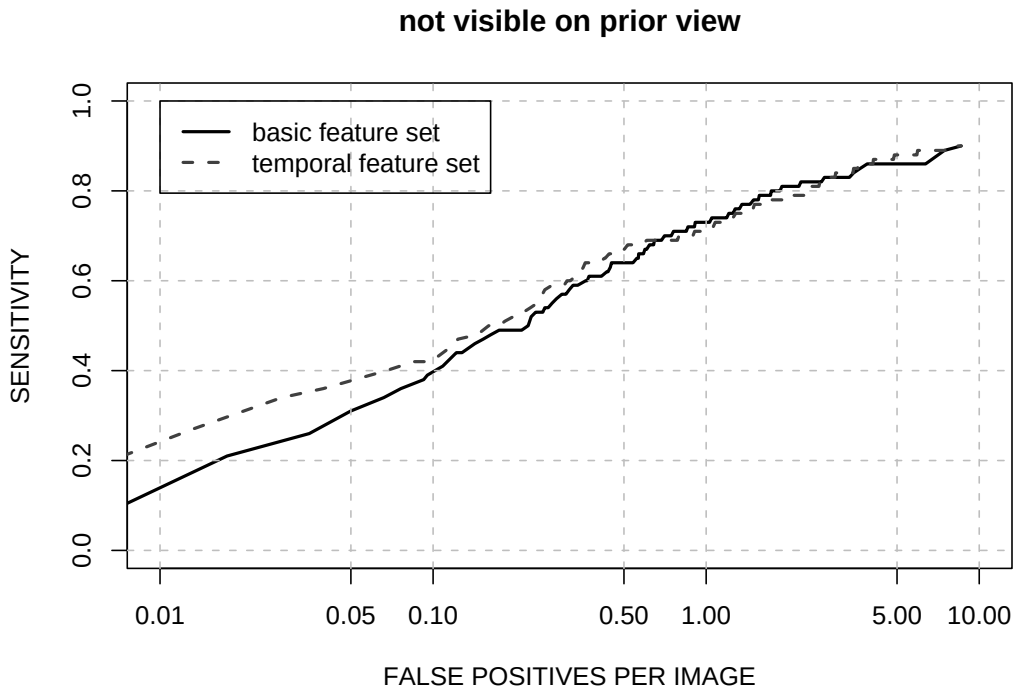
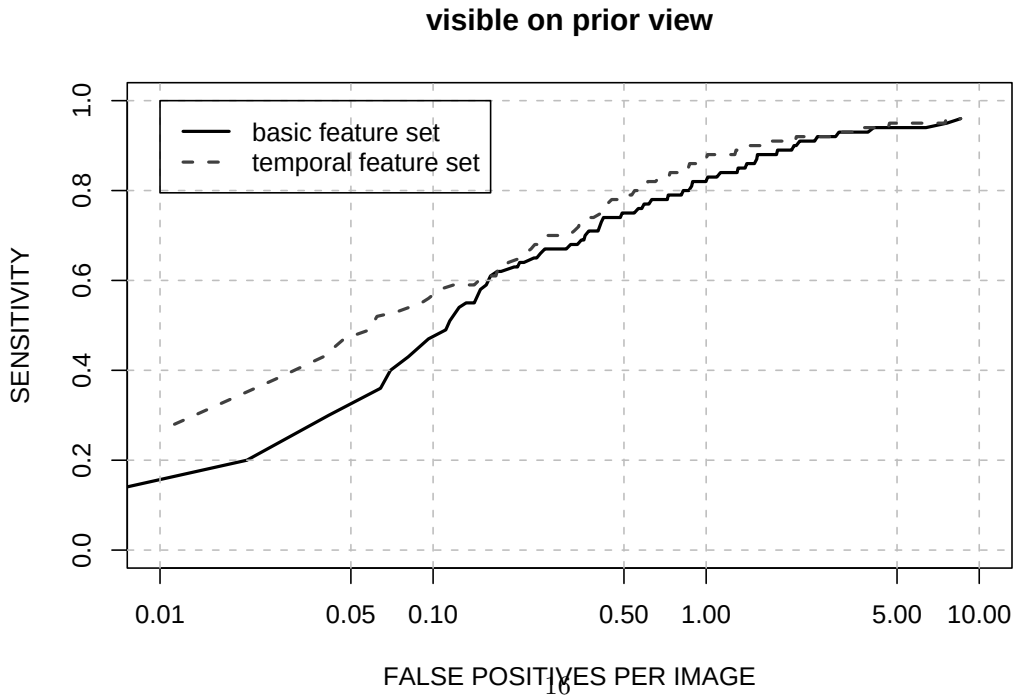


Figure 7: FROC detection results for the basic feature set and the temporal feature set. The image-based results are given.

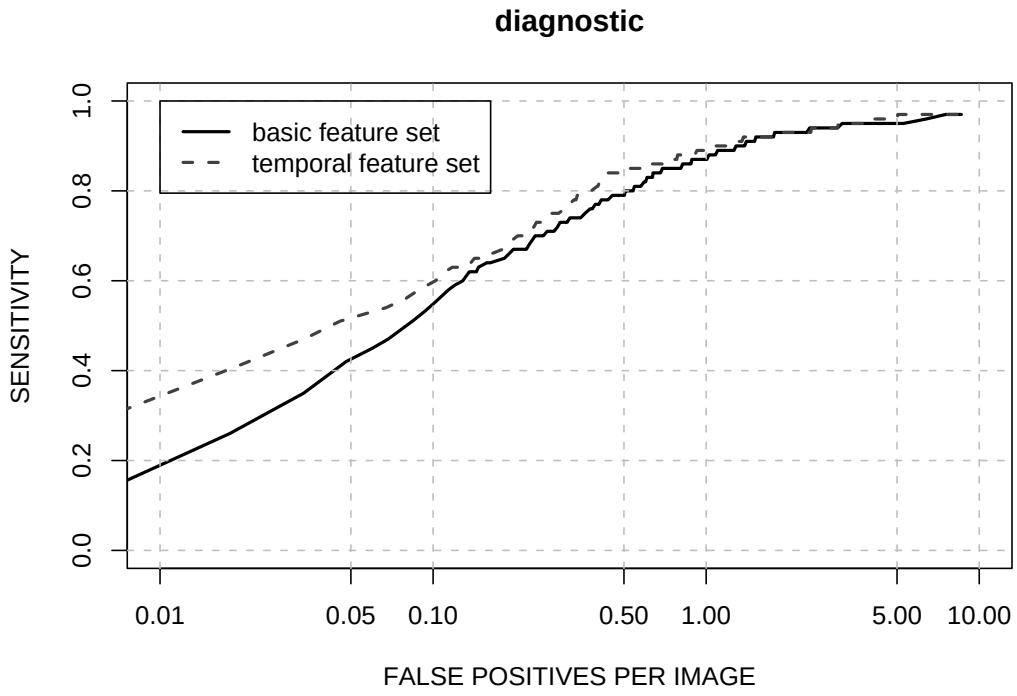


(a)

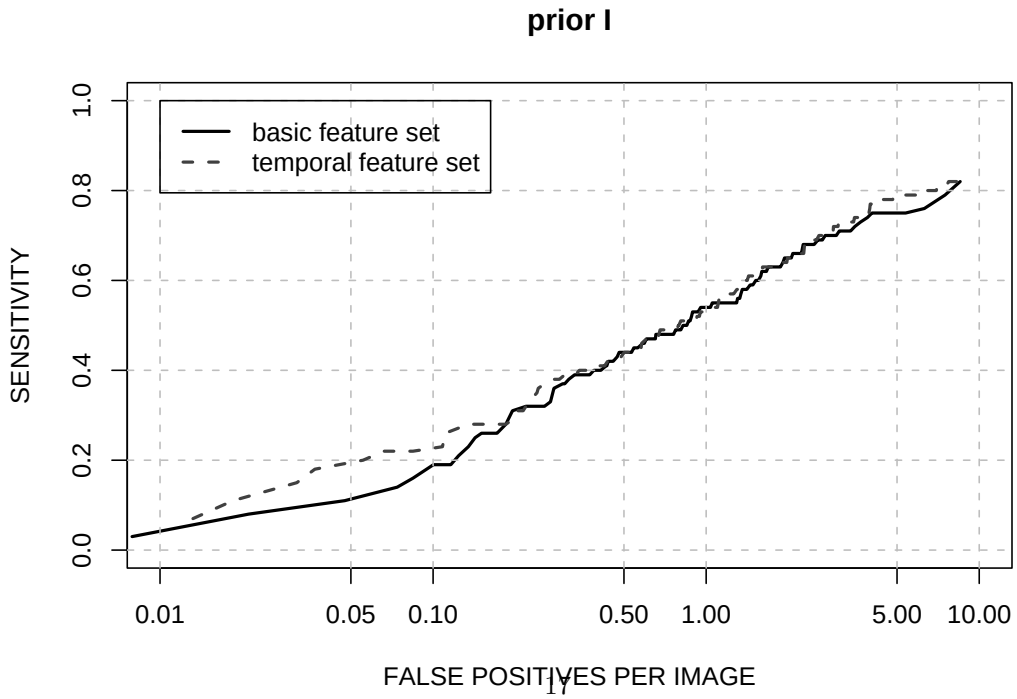


(b)

Figure 8: FROC detection results for the subsets in which the mass was visible cq. not visible on the prior mammogram.



(a)



(b)

Figure 9: FROC detection results for the subsets in which the current mammogram is the diagnostic mammogram and prior I mammogram.

5 Discussion

In this study we investigated the additional value of temporal features in our detection scheme. For this purpose we developed a regional registration technique that linked each region on the current mammogram to a corresponding region on the prior. This technique correctly linked 72% of all tumors. There are some differences between this technique and other techniques described in literature. The most important difference is that the search for a corresponding location is performed in a feature space. Although malignant tumors can change considerably between two consecutive screening rounds, other linking techniques use gray level correlation measures to find a corresponding lesion on the prior mammogram. Sometimes subtle signs of malignancy on the previous mammogram can change into an obvious mass on the current mammogram. Algorithms based on gray level matching may fail in these cases. Furthermore, gray level matching techniques are unable to deal with different compression and imaging techniques. A spiculated lesion that is present on both mammograms will have a high correlation if the orientation of both spiculated lesions is identical. A small change in orientation can change correlation measures considerably, even if both regions are nearly similar.

In our proposed algorithm we use a measure of suspiciousness to find the mass lesion on the prior mammogram. This measure corresponds with the presence of tumor characteristics at a certain location including the presence of a central mass and the degree of spiculation. This makes this method suitable to find small and subtle masses on the prior mammogram. A disadvantage of the proposed method is that it only relies on the presence of tumor characteristics, without considering the appearance of the mass on prior and current views. Consequently the method may fail if either the location on the prior does not display enough tumor characteristics or if another nearby location is more suspicious.

Compared to correlation measures the proposed method is very fast. A correlation measure is calculated at each location in the search area. This makes this algorithm very time consuming. The suspiciousness level we used to find a corresponding region was already calculated in the first detection step. It should also be mentioned that our method is fully automatic, whereas the method from Sanjay-Gopal et al relies on manual identification of the nipple. They use the position of the nipple as basis for the global and regional registration method.

In this study we used feature selection to find the best features out of the basic and temporal feature set. The feature selection method selected three temporal features: contrast difference, size difference and difference in initial likelihood. The initial likelihood is a feature from the first detection step and detects the presence of a central mass or spiculated lesion. In a previous study [11] we examined mammographic changes in masses with respect to size and contrast. In that study we found that contrast features and size features both increase during time and can be used as tumor markers . In this study we found that the feature selection method selected similar features.

Figure 7 gives the detection performance of the total dataset with and without the use of temporal features. The detection performance significantly improved when using temporal features. In the current study we calculated temporal features for all regions, regardless of whether they were visible on the prior or not. Figure 8 shows the results for existing and new lesions. We see that masses that are visible on the prior profit more from the use of temporal features. New masses, however, also show a small (not significant) improvement when temporal features are used. It might be better, however, first to classify all regions on the current as new or existing and then decide which features to calculate for each group. Some features might be calculated for both new and existing lesions. An example is contrast. If a tumor is not visible on the prior we could define an artificial region and calculate contrast measures inside this region. Then we can compare the contrast of this region with the contrast of the region on the current.

Other features can only be calculated if the region is visible on the prior mammogram such as growth of a lesion. Growth is calculated as the difference in size between the prior and current region. If nothing is visible on the prior mammogram this feature can not be calculated. If we could correctly classify regions as new or existing, we could calculate different features for both groups. This might lead to a better performance. Future work will concentrate on developing features to discriminate between new and existing lesions .

The features we used in this study are simply calculated by taking the difference between the feature value from the current region and the feature value from the prior region. We observed an improvement in detection performance with this simple technique. It will be important to evaluate other similarity measures that can characterize small difference in image features of the object of interest. If regions on current and prior images are very similar, it is more likely that one deals with a false positive or a slowly growing benign mass. On the other hand, if a region has changed considerably, this is more suspect for a malignant lesion.

6 Conclusion

We performed a study to evaluate the effectiveness of interval change analysis for the detection of malignant tumors. To this end we developed a regional registration technique that correctly linked 72% of all tumors. After the locations on the prior and the current had been linked, difference features were calculated to study interval changes. In comparison with the classification based on single view image information the temporal change information improved the detection accuracy of our CAD algorithm. Future research will concentrate on the development of specially designed temporal features and methods that discriminate between new lesions and already existing abnormalities.

Appendix

This section describes the features for both the general and the temporal CAD program in more detail. We first define a region inside and outside the contour. We use these regions for the calculation of several features. The inside region contains all pixels inside the segmentation. The outside region consists of all pixels outside the segmentation with a distance of less than $0.6R$ from the contour, where R is the effective radius of the inside region

$$R = \sqrt{\frac{\text{area of the inside region}}{\pi}}$$

The outside region is approximately twice the size of the inside region.

In this section we introduce the following notation. We denote the set of pixels in the inside region by I, the set of pixels on the contour by C and the set of pixels in the outside region by O. The set of pixels in the fatty and dense region is indicated by F and D respectively. We use $E(X)$ to denote the mean gray level of the pixels in the set X. $\mathrm{Var}(X)$ and $\sigma(X)$ represent the variance and the standard deviation of the gray level of the pixels in set X. $N(X)$ denotes the number of pixels in set X.

Dense Tissue Features We use a Gaussian mixture model to estimate the distributions of fatty and dense tissue inside the breast area. Based on this model we segment the breast into fatty tissue and dense tissue. For both regions we calculate the mean gray value and the standard

deviation. Finally we determine for each gray value y the log likelihood ratio between both tissue types.

$$L(y) = \frac{(y - E(F))^2}{\mathrm{Var}(F)} - \frac{(y - E(D))^2}{\mathrm{Var}(D)}$$

The first two dense features represent the fraction dense tissue in the breast region B and in the outside region O:

$$D_B = \frac{N_d(B)}{N_d(B) + N_f(B)}$$

and

$$D_O = \frac{N_d(O)}{N_d(O) + N_f(O)},$$

where $N_d(X)$ and $N_f(X)$ are the number of dense and fatty tissue pixels in the set X. The third $L(O)$ and the fourth feature $L(I)$ give the mean value of the likelihood ratio in the outside region O and the inside region I. The last feature gives the difference between $L(I)$ and $L(O)$:

$$Ldiff = L(I)/L(O)$$

Spiculation Features The spiculation features $f1$ and $f2$ are calculated at each location inside the breast area and represent the likelihood for the presence of a stellate pattern [21]. Two other features *f1mean* and *f2mean* are the mean value of *f1* and *f2* inside the segmented region.

Mass features The mass features $g1$ and $g2$ represent the likelihood for the presence of a focal mass at each location inside the breast area [20]. The features *g1mean* and *g2mean* are the mean value of *g1* and *g2* inside the segmented region.

Likelihood features The first likelihood measure $l1$ is the output of the classifier of the first detection step. This measure assigns a measure of suspiciousness to all pixels based on the above described features $f1, f2, g1$ and $g2$. We obtain two other features $l2$ and $l3$ by normalizing the likelihood measure $l1$ with the likelihood measure of other suspicious locations in the image. Feature $l2$ determines the relative suspiciousness of a lesion with respect to five very suspicious sites and feature $f3$ determines the suspiciousness of a lesion with respect to five other locations that were moderately suspicious.

Intensity and Contrast features The contrast of a region is a useful feature since tumor tissue absorbs more x-rays than fat and slightly more than glandular tissue. Most contrast features are calculated as a difference in intensity between an area inside the contour and an area outside the contour. We use $H(X, i)$ to denote the fraction of pixels in set X with intensity value i. To limit the number of entries, we divide the intensity range into 82 bins, each containing a range of 50 gray values. The first feature is the mean intensity inside the segmented area,

$$\text{Intensity} = E(I).$$

We then define five different contrast measures. The first contrast measure is the difference in intensity between the inside and outside region,

$$C1 = E(I) - O(I).$$

The second contrast feature is the square of the difference between the mean gray value of the inside and outside region, divided by the sum of the standard deviations of both areas,

$$C2 = \frac{(E(I) - E(O))^2}{\sigma(I) + \sigma(O)}.$$

The third contrast feature represents the distance between the histograms of the inside and outside region,

$$C3 = \sum_i |H(I,i) - H(O,i)|.$$

The fourth and fifth measure calculate the contrast relative to some dense tissue parameters,

$$C4 = \frac{E(I) - E(O)}{E(D) - E(F)}, \qquad C5 = \frac{E(I) - E(O)}{\sigma(F)}.$$

Variance We define some features based on the gray level variance in the inside and outside region. The first feature is the gray level variance in the segmented region,

$$\text{var1} = \text{Var}(I).$$

The second feature is the ratio between the variance in the inside region and the outside region,

$$\text{var2} = \frac{\text{Var}(I)}{\text{Var}(O)}.$$

The third and fourth feature calculate the variance of the inside region relative to the variance inside the fatty cq. dense tissue region,

$$\text{var3} = \frac{\text{Var}(I)}{\text{Var}(F)}, \qquad \text{var4} = \frac{\text{Var}(I)}{\text{Var}(D)}.$$

Texture As normal breast tissue often contains linear structures we developed a texture measure that captures these structures in the inside region. First we calculate at each location inside the segmentation a vector representing line orientation and magnitude, see [21]. Then we sum all vectors using the double angle representation, giving a final "total" vector. The first texture feature is the magnitude of this vector. The second texture feature is this magnitude of this vector divided by the total magnitude of all vectors inside the segmented area. We calculated the line orientation at two different scales using second order Gaussian derivatives, with a sigma of 0.2 mm and a sigma of 0.6 mm. The last texture feature calculates the magnitude of the "total" vector in the outside area O.

Iso denseness We developed a feature that determines whether dark areas are present in the inside region. It is then more likely that this region is a normal structure than a malignant mass. This feature is described in [24].

Size Total area of the inside region.

Compactness Compactness is the perimeter of the segmented region divided by the area of the segmented region.

Location features

- shortest distance to the skin line
- shortest distance to the pectoral edge
- overlap percentage of segmented region with the pectoral muscle

Wolfe class We automatically estimate the Wolfe class as described in [18].

References

[1] W. Good, B. Zheng, Y.-H. Chang, X. Wang, M. G.S., D. Gur, Multi-image CAD employing features derived from ipsilateral mammographic views, in: Proc SPIE Medical Imaging, Vol. 3661, 1999, pp. 474–485.

[2] S. Paquerault, N. Petrick, H.-P. Chan, B. Sahiner, M. Helvie, Improvement of computerized mass detection on mammograms: fusion of two-view information, Med Phys 29 (2) (2002) 238–247.

[3] F. Yin, M. Giger, K. Doi, C. Metz, C. Vyborny, R. Schmidt, Computerized detection of masses in digital mammograms : Analysis of bilateral substraction images, Med Phys 18 (5) (1991) 955–963.

[4] T. Lau, W. Bischof, Automated detection of breast tumors using the asymmetry approach, Comp and Biomed Research 24 (1991) 273–295.

[5] K. Bovis, S. Singh, J. Fieldsend, C. Pinder, Identification of masses in digital mammograms with MLP and RBF nets, in: Proceedings of the IEEE-INNS-ENN, Vol. 1, 2000, pp. 342–347.

[6] S. Timp, N. Karssemeijer, Use of temporal features to improve mass detection, in: 7th International Workshop on Digital Mammography, Chapel Hill, 2004.

[7] N. Vujovic, D. Brzakovic, K. Fogarty, Detection of cancerous changes in mammograms using intensity and texture measures, Proc SPIE 2434 (1995) 37–47.

[8] S. Kok-Wiles, J. Brady, R. Highnam, Comparing mammogram pairs in the detection of lesions, in: N. Karssemeijer, M. Thijssen, J. Hendriks, L. van Erning (Eds.), Digital Mammography, Kluwer, Dordrecht, 1998, pp. 103–110.

[9] L. Hadjiiski, B. Sahiner, H.-P. Chan, N. Petrick, M. Helvie, M. Gurcan, Analysis of temporal changes of mammographic features: computer-aided classification of malignant and benign masses, Med Phys 28 (11) (2001) 2309–2317.

[10] K. White, K. Berbaum, W. Smith, The role of previous radiographs and reports in the interpretation of current radiographs, Invest Radiol 29 (3) (1994) 263–265.

[11] S. Timp, N. Karssemeijer, J. Hendriks, Analysis of changes in masses using contrast and size measures, in: H. Peitgen (Ed.), 6th International Workshop on Digital Mammography, Bremen, Germany, Springer-Verlag, 2002, pp. 240–242.

[12] M. Sallam, K. Bowyer, Registering time-sequences of mammograms using a two-dimensional unwarping technique, in: A. Gale, S. Astley, D. Dance, A. Cairns (Eds.), 2th International Workshop on Digital Mammography, York, USA, Elsevier, Amsterdam, 1994, pp. 121–131.

[13] N. Vujovic, D. Brzakovic, Establishing the correspondence between control points in pairs of mammographic images, IEEE Trans Med Imaging 6 (10) (1997) 1388–1399.

[14] F. Richard, C. Graffigne, An image matching model for the registration of time sequence or bilateral mammogram pairs, in: M. Yaffe (Ed.), 5th International Workshop on Digital Mammography, Toronto, Canada, Med Phys Publishing, Madison, 2000.

[15] S. van Engeland, P. Snoeren, N. Karssemeijer, J. Hendriks, A comparison of methods for mammogram registration., IEEE Trans Med Imaging 22 (11) (2003) 1436–1444.

[16] S. Sanjay-Gopal, H.-P. Chan, T. Wilson, M. Helvie, N. Petrick, A regional registration technique for automated interval change analysis of breast lesions on mammograms, Med Phys 26 (12) (1999) 2669–2679.

[17] L. Hadjiiski, H.-P. Chan, B. S. et al, Automated registration of breast lesions in temporal pairs of mammograms for interval change analysis – local affine transformation for improved localization, Med Phys 28 (6) (2001) 1070–1079.

[18] N. Karssemeijer, Automated classification of parenchymal patterns in mammograms, Phys Med Biol 43 (2) (1998) 365–378.

[19] D. Ballard, C. Brown, Computer Vision, Prentice-Hall, 1982, Ch. 4.

[20] G. te Brake, N. Karssemeijer, Single and multiscale detection of masses in digital mammograms, IEEE Trans Med Imaging 18 (7) (1999) 628–639.

[21] N. Karssemeijer, G. te Brake, Detection of stellate distortions in mammograms, IEEE Trans Med Imaging 15 (1996) 611–619.

[22] S. Timp, N. Karssemeijer, A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography, Med Phys 31 (5) (2004) 958–971.

[23] P. Pudil, J. Novovicova, J. Kittler, Floating search methods in feature selection, Pattern Recognition Letters 15 (11) (1994) 1119–1125.

[24] G. te Brake, N. Karssemeijer, J. Hendriks, An automatic method to discriminate malignant masses from normal tissue in digital mammograms, Phys Med Biol 45 (10) (2000) 2843–2857.