A regional registration method to find corresponding mass lesions in temporal mammogram pairs.

Sheila Timp, Saskia van Engeland, Nico Karssemeijer

Department of Radiology
Radboud University Medical Centre Nijmegen
The Netherlands

Corresponding author:

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

Abstract

In this paper we develop an automatic regional registration method to find corresponding masses on prior and current mammograms. The method contains three steps. In the first, we globally align both images. Then, for each mass lesion on the *current* view, we define a search area on the *prior* view, which is likely to contain the same mass lesion. Third, at each location in this search area we calculate a registration measure to quantify how well this location matches the mass lesion on the current view. Finally we select the best location. To determine the performance of our method we compare it to several other registration methods. On a dataset of 389 temporal mass pairs our method correctly links 82% of prior and current mass lesions, whereas other methods achieve at most 72%.

keywords: registration, interval change, temporal, computer-aided detection (CAD), breast cancer, mammography

.

1 Introduction

Radiologists generally use multiple mammographic views to detect and characterize suspicious regions. Besides images of the left and right breast they usually have views from previous screening rounds and views from different projections. When radiologists discover a suspicious lesion in one view, they try to find a corresponding lesion in the other views. Views from different projections, typically cranio caudal (CC) and medio lateral oblique (MLO) views, allow for a better characterization of the lesion. Prior views are useful to study changes in the appearance of the lesion over time. Contra-lateral views provide a reference to the normal appearance of the breast and help to determine the relative suspiciousness of the lesion. By combining information from all views radiologists estimate the suspiciousness of the lesion and decide whether further investigation is required. Studies report a positive effect on either recall rate or an improvement in mass detection performance when using multiple views in mammography screening compared to single-view mammography, cf. [1–4].

Most current computer aided detection (CAD) systems differ considerably from radiologists in the way they use multiple views. These systems do not combine information from available views but instead analyze each view separately. Given the positive effect of multi-view systems on radiologists' performance we expect that fusion of information from different views will improve CAD systems as well.

Such multi-view CAD programs require regional registration methods to find corresponding regions in all available views. In this paper we concentrate on developing such a method for corresponding mass lesions in *prior* and *current* mammographic views. In other words, starting from a current image containing a mass lesion, the method aims at locating the same mass lesion in the prior image.

Literature Few studies have been done to find corresponding regions in different mammographic views. These studies aim at finding similar structures in either different projections of the same breast [5,6] or mammograms obtained at different points in time [7–10]. The first three studies on temporal registration [7–9] first localize the mass on the current mammogram in a polar coordinate system with the nipple as the origin. Based on these coordinates the location of the mass on the prior mammogram is estimated. The predicted location of the mass centroid on the prior mammogram determines a fan-shaped search region. A similarity measure then determines the best matching location inside this fan-shaped search region. In [8] the usefulness of correlation and mutual information as registration measures is investigated. In a recent study [9] twelve different similarity measures are compared for the task of template matching. That study shows that the best performing similarity measures for matching corresponding regions in temporal mammogram pairs are Pearson's correlation, the cosine coefficient, and Goodman and Kruskal's Gamma coefficient. In the last study we [10] have developed a regional registration method in which the search for correspondence is done in a feature space. For this purpose we first construct this feature space by estimating at each location inside a circular search area the likelihood that a mass is present, called the mass likelihood. Then we select the location with the highest mass likelihood inside this search area as match for the lesion on the current view.

Both registration methods have some disadvantages. A problem with methods based on template matching is that this only works when both regions are more or less similar in appearance. This might be true for some—especially benign—masses that stay more or less constant in time, but is obviously not true for malignant masses that change considerably in time. Registration

methods that work in a feature space and use a mass likelihood as registration measure work well in cases with relatively few potential lesion candidates. These may fail if the prior region does not display enough mass characteristics—has a low mass likelihood—or if the search area contains more than one region with a high mass likelihood. Therefore we develop a new method that combines the above mentioned methods.

Method Our regional registration method comprises three steps. First, we align both images. Second, we define for each mass lesion on the current view a search area on the prior view in which the mass lesion is most likely to be located. In the third step we combine three registration measures to determine the best location inside the search area. Finally, we choose this location as estimate for the center of the prior mass lesion.

More specifically, in the third step we apply the following three registration measures. The first measure represents the likelihood that a mass is present, i.e. the mass likelihood. As second measure we use Pearson's correlation coefficient that measures the similarity between the mass on the current view and a candidate region for the corresponding mass on the prior view. We evaluate the effect of different template shapes on the performance of this correlation measure and select the best performing shape as our second registration measure. The last measure is a distance criterion that gives preference to locations near an initial estimate. Pertaining to the running time of the method, we also provide a fast variant of this registration method in which the measures are applied sequentially.

We compare the performance of this method with methods that use only one registration measure. For this purpose we use a dataset consisting of 389 temporal mammogram pairs all containing a visible mass on both prior and current views. Finally, we investigate possible shortcomings of each method by comparing the registration performance on different subsets including benign and malignant masses, and masses that are subtle cq. obvious on the prior view.

Structure The paper has the following structure. In Section 2 we explain the registration methods in more detail. In Section 3 we describe the experiments to evaluate the different registration methods. Section 4 presents the results with a discussion and conclusion in the last sections.

2 Method

In this section we present the general procedure we follow to register temporal mass pairs. First, in Section 2.1, we describe preprocessing and global registration. Section 2.2 describes the definition of a search area. Then, in Section 2.3, we explain each of the applied registration measures.

2.1 Preprocessing and global registration

Before we can globally register prior and current views we have to pre-process both images. To this end, we first segment each image into breast region, background tissue and pectoral muscle, using a breast boundary and pectoral muscle segmentation algorithm developed previously in our group[11]. We subsequently apply an algorithm that removes additional attenuation from the pectoral muscle[10]. This pectoral equalization method makes the border region more homogeneous, which is advantageous when dealing with masses that develop on the pectoral boundary.

Finally we apply a peripheral enhancement algorithm to the breast area to correct for differences in tissue thickness[10].

Following preprocessing we use a simple procedure based on a center of mass alignment to globally register both images. For this alignment we first determine the mathematical center of mass of prior and current images. The center of mass can be determined using the whole breast area including the pectoral muscle or the breast area with the pectoral muscle excluded. In our experiments we determine the center of mass of prior and current images with the pectoral muscle excluded as this improves the registration accuracy (see Table 2 and[12]). Then we horizontally and vertically shift the prior image such that its center of mass coincides with the center of mass of the current image. Figure 1 illustrates this alignment after a vertical shift of ty and a horizontal shift of tx. After alignment of both images we use the center coordinates of the lesion on the current image (cx, cy) as initial estimate of the location of the lesion on the prior image.

2.2 Definition of the search area

After both images have been aligned we define for each mass lesion on the current image a search area on the prior image. In the literature two different shapes of a search area have been proposed: a circular search area[10] and a fan-shaped search area[7]. As prerequisite we consider it important that the definition of a search area can be done completely automatic. A fan-shaped search area, as proposed in[7], requires the position where the nipple is located. This is a disadvantage as the nipple may sometimes be hard to identify, in particular if modern high contrast films screen combinations are used. Therefore we decide to use a circular search area that is independent of the nipple position and is completely determined by specifying its center coordinates and radius. As center we choose the initial estimate of the location of the mass lesion on the prior view.

To ensure that the search area includes most masses we use a large radius of 30 mm. The size of the search area is based on a comparative study Van Engeland et al.[12] performed on several registration methods. In that study they found that the maximum error, i.e. the maximum distance between the estimated mass location and the real mass location, is 30 mm for a center of mass alignment. Figure 1 shows the final search area on the prior view with center (cx, cy) and a radius of 30 mm. After defining the search area we calculate at each location inside this area the regional registration measures that we will describe in the next paragraph.

2.3 Registration Methods

2.3.1 Registration based on mass likelihood

The registration method based on mass likelihood selects the location with the highest mass likelihood value as estimate for the mass location on the prior image. To this end we assign all locations inside the search area a mass likelihood value based on the outcome of a pixel level mass detection algorithm. This algorithm calculates at each location inside the breast area two features for the detection of stellate lesions and two features for the detection of focal masses. Below we will shortly describe these features. For details about the algorithms see[13] and[14].

Features to detect stellate lesions The spiculation features are based on the idea that stellate lesions show a patterns of lines directed toward the center pixel of a lesion. We first estimate the line orientation at each location inside the image using directional second order Gaussian derivatives. We then define a neighborhood around each pixel in the image. We call this pixel the center pixel. To estimate whether a spiculated lesion is present we derive the following two features from the map of line orientations inside this neighborhood. The first feature is a normalized measure for the fraction of pixels with a line orientation directed toward

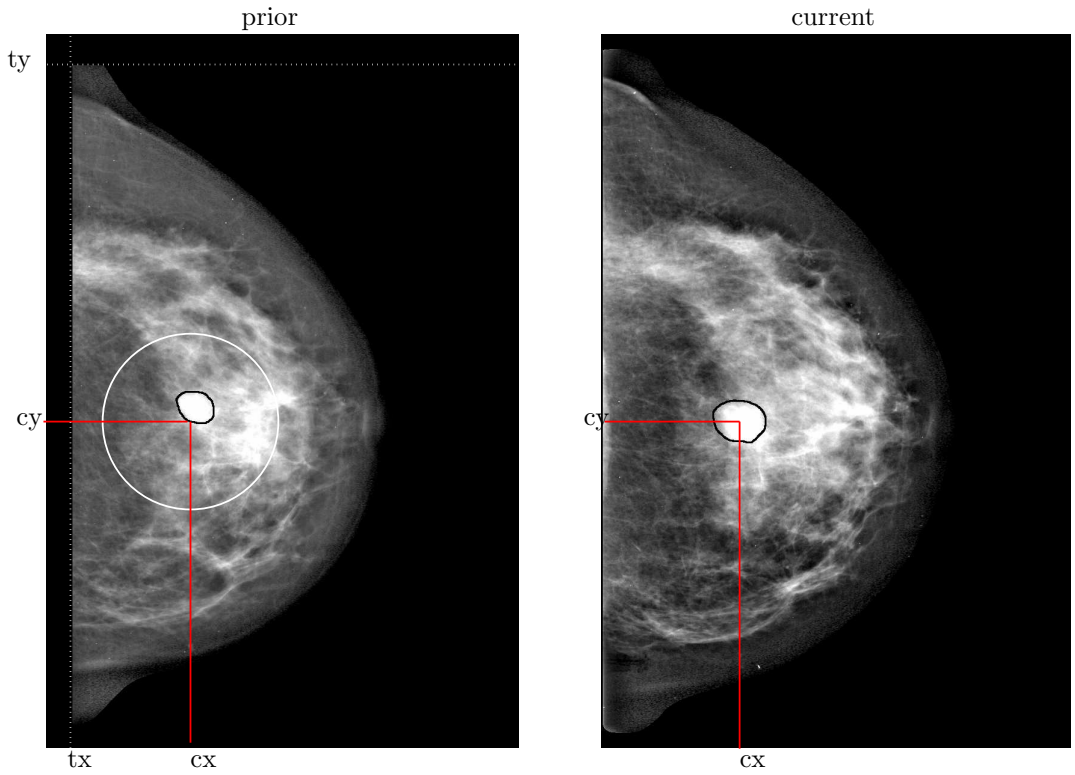


Figure 1: Global alignment and definition of the search area. First both images are aligned by shifting the prior image with tx and ty. The center coordinates of the lesion on the current image (cx, cy) then form the initial estimate for the lesion on the prior image. This initial estimate is the center of a circular search area (white circle) with radius r. We calculate the registration measures at each location inside this search area.

the center pixel. The second feature calculates to what extent the pixels with a line orientation toward the center pixel are equally distributed in all directions.

Features to detect focal mass lesions For the detection of masses we use a similar approach as for the detection of spicules. Instead of determining the line orientations we now calculate the gradient orientation at each location in the image. If a mass is present pixels in a neighborhood of the center pixel will have a gradient orientation toward the center. Otherwise a random gradient direction will be found. We derive the following two features from the calculated gradient orientations. The first feature is a normalized measure of the fraction of pixels with an intensity gradient pointing toward the center pixel. The second feature calculates whether these pixels occur in all directions of the center pixel.

Mass likelihood The spiculation and mass features are used as input for a 3-layer feed-forward neural network trained on known abnormalities. Then we use the classifier output as mass likelihood.

2.3.2 Registration based on gray scale correlation

Method Registration methods based on gray scale correlation calculate the pixel correlation between a template image of the current mass—the current mass template—and a candidate region on the prior image. We first select one of the templates described below and put this template over the current mass lesion to obtain the current mass template. Then we obtain candidate regions for the prior mass by putting the template at each location inside the search area on the prior image. Finally we calculate Pearson's correlation measure between the current mass template and the candidate region centered at (x, y) on the prior image:

$$C(x,y) = \frac{\sum_{(m,n)}(p(m,n) - \overline{p})(q_{(x,y)}(m,n) - \overline{q})}{\sqrt{(\sum_{(m,n)}(p(m,n) - \overline{p})^2)(\sum_{(m,n)}(q_{(x,y)}(m,n) - \overline{q})^2)}} \quad (1)$$

The gray level at location (m, n) in the current mass template is given by $p(m, n)$ and the gray level of the candidate region with center (x, y) at the same relative location by $q_{(x,y)}(m, n)$. The summation is performed over all locations (m, n) inside the current mass template. The average gray level values in the mass template and the candidate region are given by $\overline{p}$ and $\overline{q}$. We select the location with the highest correlation as estimate for the location of the mass on the prior. The next paragraph describes different template shapes.

Mass templates We design different templates for the registration method based on correlation: an inner mass template, an outer mass template, and three extended templates. These templates cover different parts of the underlying mass lesion and its surrounding tissue. Figure 2 illustrates the templates for a benign mass. For all templates we first determine the contour of the current mass with a segmentation algorithm based on dynamic programming [15].

The inner mass template, as illustrated in Figure 2(b), consists of all pixels inside the contour and exactly represents the underlying mass lesion. A candidate region on the prior image highly correlates with this mass template when the mass lesion is similar in appearance on prior and current views. On the other hand, if the inner part of the mass changes considerably between two consecutive screening rounds, for example in size or contrast, the correlation between both regions will be low.

Figure 2(c) illustrates the outer mass template. This template consists of all pixels surrounding the mass with a distance of less than 6 mm from the border of the mass. Consequently the

correlation only depends on the similarity between the outer border region of the current mass and a similar region on the prior view. This can be an advantage for masses that change significantly in appearance. For these masses the outer border region will stay more or less similar in appearance between both views. On the other hand, this can cause problems when the gray level characteristics of the outer border region are not unique. An example is a tumor completely embedded in fatty tissue. The outer mass template will represent gray level characteristics of fatty tissue and thus show little variation. Consequently, the correlation between this template and a candidate region on the prior image will be high when the candidate region is homogeneous as well. In uniform breasts this may result in many candidate regions all correlating equally well with the current mass template.

Figure 2(d), 2(e) and 2(f) show the extended templates. These templates consists of an inner region as well as an outer border region. The first extended template, Figure 2(d) is a simple extended template that consist of the whole inner region and an outer border region. The second one—the growing mass template—only contains the most inner part of the mass lesion and an outer border region. We design this template for masses that grow between two screening rounds. We assume that for these masses the most inner part and the outside border part will be more or less similar for the prior and the current mass. To determine an appropriate size for the inner region we make use of the observation that for most masses in our database the projected size at most doubles between two consecutive screening rounds. Consequently, the inner part of this combined mass template consists of all pixels within a distance of $r/\sqrt{2}$ from the center of the mass lesion. The last extended template is the circular template. For this template we first calculate the area A of the region inside the segmentation and then determine the effective radius $R = \sqrt{(}A/\pi)$. The radius of the circular template is $R + b$ where b is the size of the outer border region. For all three extended templates the size of the outer border region is 3 mm.

2.3.3 Combined registration methods

The last registration method combines the mass likelihood with the correlation measure and a distance criterion. We develop two variants of this combined method. In both methods we first determine the individual registration measures at selected locations inside the search area. After calculating the individual measures at each location we normalize each measure v using the minimum and maximum values found in the dataset

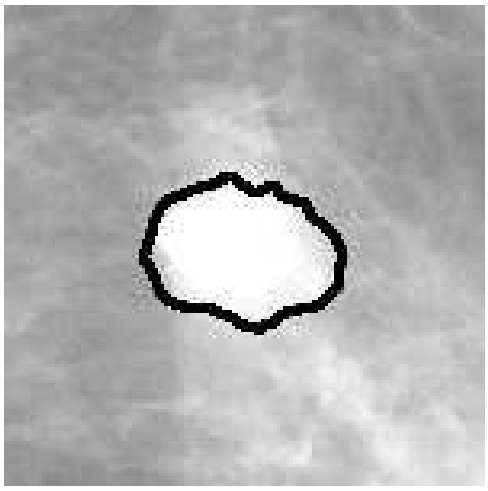
$$\overline{v} = \frac{v - \min(v)}{\max(v) - \min(v))} \tag{2}$$

and then linearly combine them:

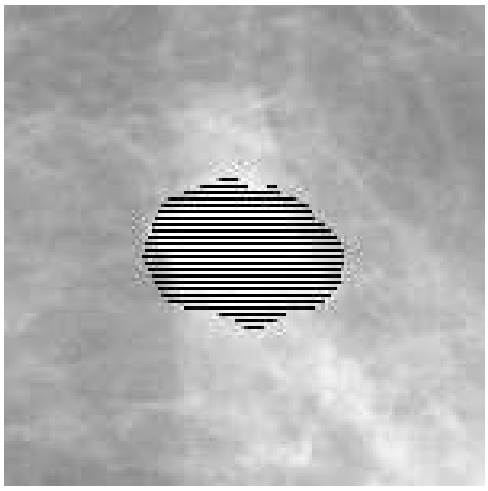
$$R(x, y) = \alpha C(x, y) + \beta l(x, y) - \gamma d(x, y), \tag{3}$$

where $R(x, y)$ is the combined registration measure, $C(x, y)$ the gray scale correlation measure for the best performing template shape, $l(x, y)$ the mass likelihood and $d(x, y)$ the distance to the initial estimate. We then use the whole dataset to determine the weights α, β and γ to achieve maximum registration performance. To find the optimal weights we vary the coefficients α and β between zero and 100 and keep γ fixed at 51.

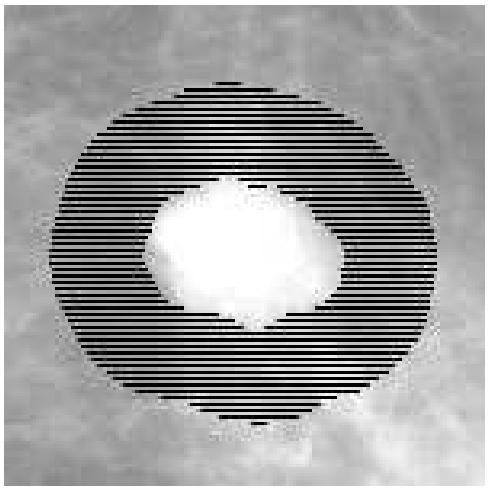
The difference between both methods is the selection of the locations where the measures are calculated. The first variant simply calculates the three registration measures at each location inside the search area. As all measures are used simultaneously we call this method the simultaneous combination method. To reduce the computational effort we developed a second variant in which we calculate the registration measures sequentially. This method first selects all locations inside the search area with a mass likelihood above a certain threshold. If two selected locations



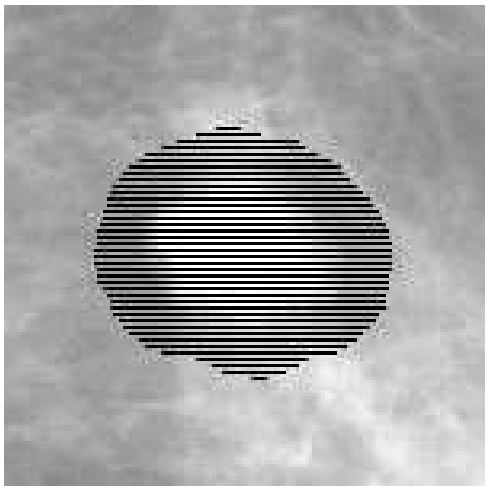
(a) Segmentation of a benign mass



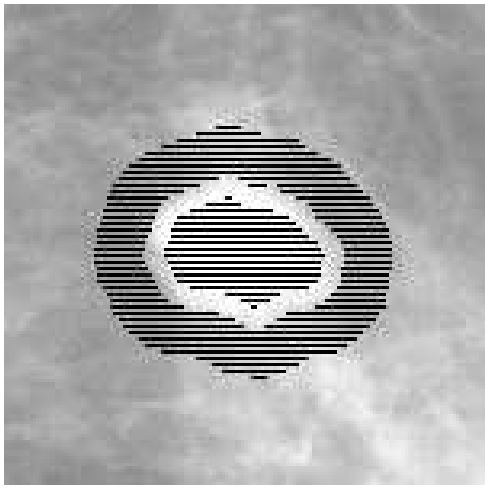
(b) Inner mass template



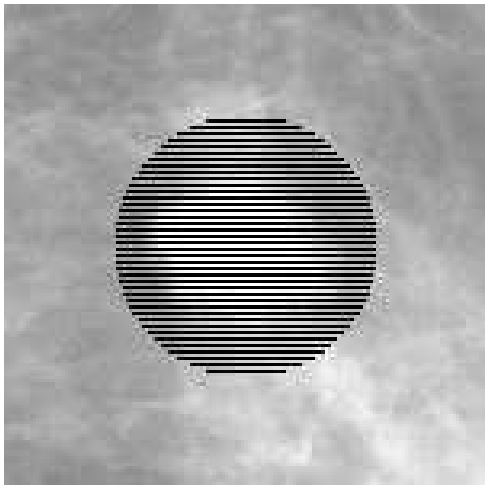
(c) Outer mass template



(d) Simple extended mass template



(e) Growing mass template



(f) Circular mass template

Figure 2: Figure 2(a) shows the segmentation of a benign mass. The other figures show the different mass templates.

are less then one millimeter apart we remove the one with the lowest mass likelihood. This procedure results in an average of 100 selected locations for each search area. As the measures are calculated sequentially we call this method the sequential combination method.

We compare both variants with respect to registration performance and computational efficiency. An important difference between both methods is that the sequential method only processes locations that show mass characteristics. This has positive and negative consequences. A negative consequence is that a correct location will be missed when its mass likelihood is below the threshold, independent of the value of the correlation measure. This may result in a decrease of the registration performance for (benign) lesions with few mass characteristics. A positive consequence is that the sequential method will skip locations with accidental high correlation when they display not enough mass characteristics. This may increase the probability that a correct match occurs.

The computational efficiency is determined by the number locations at which the correlation measure is calculated. This corresponds with on average 100 locations for the sequential method. The simultaneous method calculates the correlation measure at each location inside the search area. This corresponds with almost 71,000 locations for a search area with radius 30 mm and a pixel resolution of $200\mu m$. The computational effort is thus reduced about a thousandfold by using the sequential method compared to the simultaneous method. For the sequential registration method, the whole procedure, including the calculation of the mass likelihood, takes less than one minute per image. As we determine the mass likelihood already in our standard CAD program for the task of mass detection, the extra time needed for the registration is based solely on the calculation of the correlation measure. This takes a few seconds per image in the sequential registration method. This means that the method can be implemented into a CAD system without much additional time costs.

3 Experiments

3.1 Dataset

The mammograms used in this study all come from the Dutch Breast Cancer screening Program. All women aged 50-75 are invited bi-annually 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. A temporal image pair consists of the mammograms from two consecutive screening rounds. We call the most recent image in a temporal pair the current mammogram and the image obtained in the previous screening round the prior or previous mammogram.

We construct the dataset for the experiments by collecting all temporal image pairs from two different datasets in which a mass is visible on both prior and current views. Information about both datasets is summarized in Table 1. The first dataset consists of 155 image pairs with a malignant mass on both views. This dataset contains 281 images from 87 patients. The number of temporal pairs is larger than half of the number of the images since for some women the mammograms of three consecutive screening rounds were available. The images were digitized with a Lumisys 85 digitizer at a pixel resolution of $50\mu m$ x $50\mu m$.

The second dataset consists of 234 image pairs in which 94 contain a malignant mass and 140 a benign mass. This dataset contains 434 images from 155 patients. These images were digitized with a Canon CFS300 laser scanner at a pixel resolution of $50\mu m$ x $50\mu m$. A radiologist rated all masses in this dataset for their visibility on a scale from 1 to 5. A rating of 1 corresponds to masses that are clearly visible on the prior view and a rating of 5 corresponds to masses that are

very subtle on the prior view and that can only be detected when the current view is available. All images were averaged to a resolution of $200\mu m$ maintaining the original gray value resolution of 12 bits. Combination of the two sets results in 389 temporal pairs of images, 140 benign and 249 malignant.

	dataset I	dataset II
number of image pairs	155	234
number of images	281	434
number of patients	87	155
malignant image pairs	155	94
benign image pairs	0	140
MLO views	124	194
CC views	31	40

Table 1: Composition of the datasets used for the experiments.

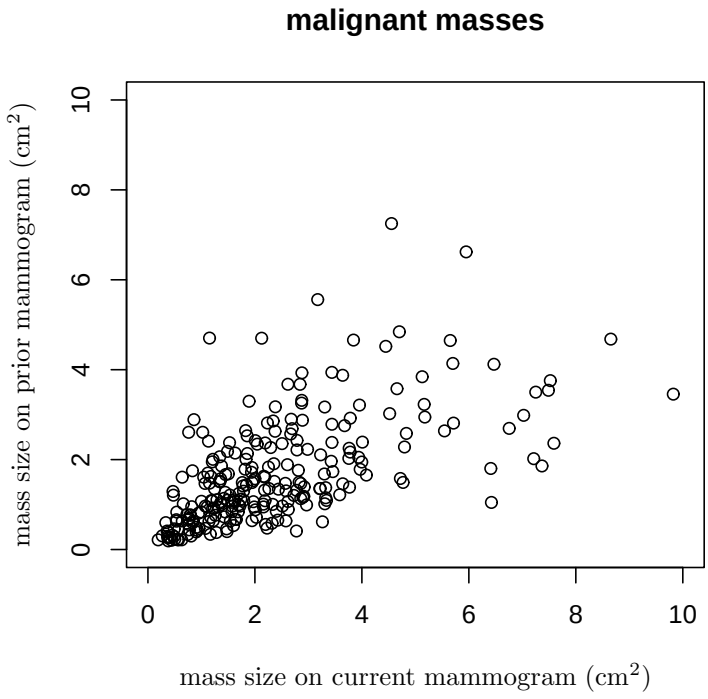
We annotated all mass lesions on prior and current views under supervision of an expert radiologist. For this purpose we used specially designed software on a dedicated mammographic review station. We determined the mass size on both current and prior mammograms as the area inside the annotation. Figure 3 shows the distribution of the mass size for benign and malignant masses. The mean size of benign masses in the current mammogram is 2.2 cm^2 versus 1.8 cm^2 in the prior mammogram. The average growth of the benign masses, defined as the ratio between the current and the prior mass size, is 1.4. The mean size of malignant masses in the current mammogram is 2.4 cm^2 versus 1.7 cm^2 in the prior mammogram. The average growth of malignant masses is 1.66.

3.2 Subsets

We test the performance of each registration method on the whole dataset and on several subdivisions of the original dataset. These subdivisions contain different mass types and the performance on these subsets informs us about specific shortcomings of each method. The first subdivision is between benign and malignant masses. We use this subdivision to test our assumption that correlation measures are more suited for benign masses and measures based on mass likelihood for malignant masses. We base this assumption on the fact that correlation measures work best for masses that stay more or less constant in time, which is often the case for benign masses. Malignant masses, on the other hand, can change considerably in time, not only in size but also in contrast and overall appearance. We use the set of malignant masses that have been rated for their visibility to make the second subdivision between masses that are clearly visible on the prior view and masses that are very subtle on the prior view. To this end we put all masses with a visibility rating below 5 in the group of "obvious priors" and masses with a visibility rating of 5 in the group of "subtle priors". We expect that masses from the group of subtle priors are very different in appearance on prior and current views and will thus be less suited for methods based on correlation.

3.3 Validation

As evaluation measure for the registration methods we use the fraction of correctly matched lesions. We count a match as correct when the selected location is inside the annotation of the



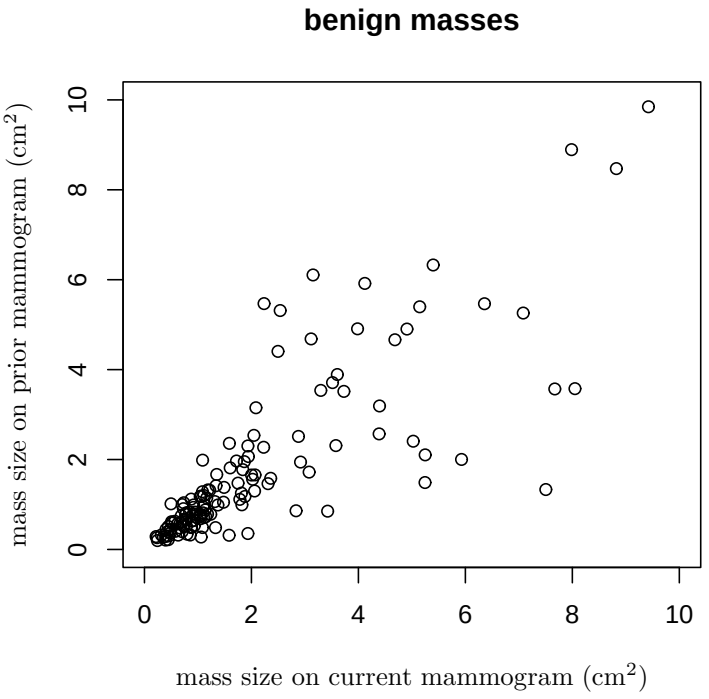


Figure 3: Mass size of masses on prior and current mammogram

radiologist. Besides evaluating the performance of each registration method we determine the optimal search radius by varying the radius of the search area between 0 and 30 mm.

4 Results

In the first two paragraphs we present the results for the global and regional registration methods. We used the complete dataset of 389 temporal image pairs to evaluate the different registration methods. In the third paragraph we give the performance of the proposed registration method for different subsets. The last paragraph describes cases where the registration method failed to establish a correct link.

4.1 Global registration

The results for the global registration procedure are summarized in Table 2. The second column gives the percentage of correctly linked masses. The last column gives the mean distance from the initial estimate (the center of the search area on the prior mammogram) to the center of the ground truth. From Table 2 we see that the global registration method improves when we exclude the pectoral muscle for determining the center of mass.

	fraction correct	mean distance to ground truth (mm)
with pectoral muscle	0.30	11.9
without pectoral muscle	0.37	9.9

Table 2: Results for the global registration procedure where the center of mass has been determined with and without the pectoral muscle. The first column gives the fraction of correctly linked masses. The second column gives the mean distance from the chosen location to the center of the ground truth.

4.2 Registration measures

Table 3 and Figure 4 show the results for the different registration measures. The best performance for the measure based on mass likelihood is 0.71 for a search radius of 12 mm. Considering the correlation measure we find that the inner and outer mass template have a significantly lower performance than the extended mass templates. The best performing extended mass template is the growing mass template, although the difference with the other extended templates is not statistically significant. We furthermore studied the influence of the outer border region by varying the size of this region in the simple extended template between 0 and 8 mm. From Table 4 we see that the fraction of correctly linked masses is 0.60 for the simple extended template without an outer border region, that this fraction increases up until 0.68 for an outer border region of 1.4 mm and then stays more or less constant.

We select the growing mass template as template for the gray scale correlation measure in the combined registration methods. The difference between the performance of the combined registration methods and the individual registration measures is statistically significant. Figure 4 shows that the performance of both combined methods increases up until 0.82 for a search radius of 20 mm and then stays more or less constant. The weights for α, β and γ were 33, 29 and 51 for the simultaneous combination method and 51, 75, 51 for the sequential combination method.

registration measure	fraction of correctly linked masses ± sd	radius of search area	mean distance to ground truth
mass likelihood	0.71 ± 0.02	12	3.6
inner mass template	0.60 ± 0.02	16	4.2
outer mass template	0.48 ± 0.03	8	4.6
simple extended mass template	0.69 ± 0.02	20	3.6
growing mass template	0.71 ± 0.02	20	3.5
circular mass template	0.69 ±0.02	16	3.7
simultaneous combination	0.82 ±0.02	20	2.6
sequential combination	0.82 ± 0.02	20	2.8

Table 3: Registration results for the different methods. The first column shows the registration measure. The second column gives the fraction of correctly linked masses and the standard deviation. The third column shows the radius of the search area where the maximum performance has been obtained and the last column the mean distance to the ground truth.

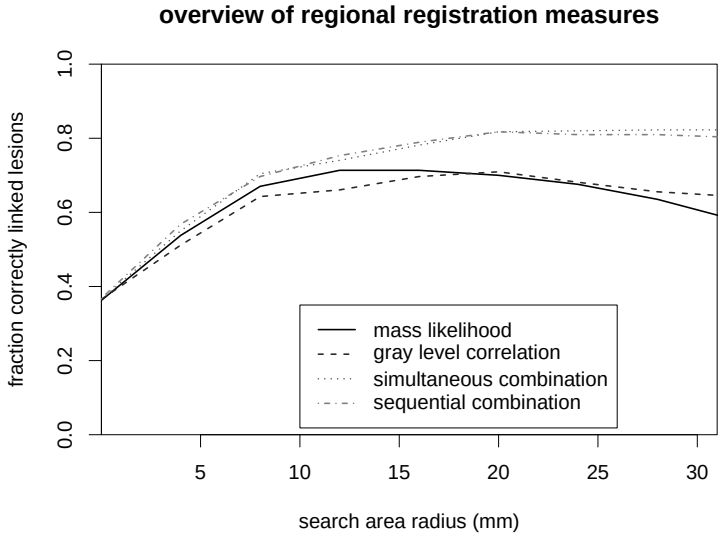


Figure 4: Fraction of correctly linked masses for the measure based on mass likelihood, the measure based on gray level correlation and the combination methods.

outer border size (mm)	0	0.6	1.0	1.4	2.0	3.0	4.0	6.0	8.0
fraction correctly linked	0.60	0.63	0.66	0.68	0.68	0.69	0.69	0.69	0.69

Table 4: Fraction of correctly linked masses for different outer border sizes.

From these coefficients we see that the distance measure is more important for the simultaneous method than for the sequential method. For the sequential method the mass likelihood measure has a lower weight than the correlation measure. This was expected as all processed locations in the sequential method already have a relatively high mass likelihood. The choice for a location then mainly depends on the gray scale correlation measure. For both methods we find that small variations in the coefficients have little influence on the results. For example, when the coefficients α, β and γ have equal weights, the performance of both methods is 0.80.

Figure 5 shows a scatter plot for the measure based on mass likelihood versus gray scale correlation for correctly linked masses and masses that were linked incorrect. The correlation between both measures is 0.34 for correctly linked masses and 0.22 for incorrect matches. From the figure we see that most correctly linked masses have a high correlation measure and a high mass likelihood. However, there is also a large number of masses with either a low correlation or a low mass likelihood. This explains the increased performance of the combination methods compared to the performance of the individual measures.

Figure 6 shows the histogram of the distance between the selected location and the center of the ground truth. The mean distance for correctly linked masses is 1.2 mm. For incorrect matches the mean distance is 10.0 mm. This is more or less equal to the mean distance measured after the global registration step. There are a few outliers among the incorrect links. In these cases the global registration failed and the true mass lesion was located outside the search area.

4.3 Subsets

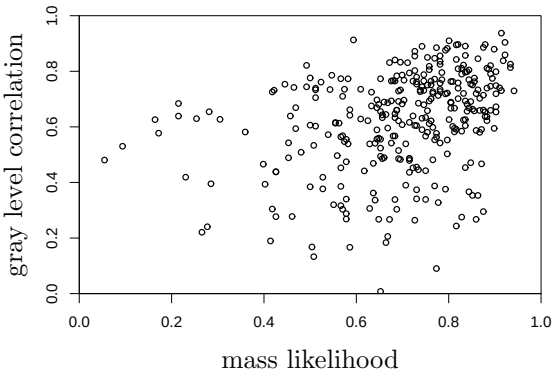
Table 5 gives the fraction of correctly linked masses for the different subsets. This table shows that the mass likelihood performs best on malignant masses and the gray scale correlation measure on benign masses. The combination methods perform satisfactory on both subsets. Table 5 also shows that the individual registration measures perform similarly for masses that are subtle cq. obvious on the prior view. Finally, we find that the sequential combination method performs better than the simultaneous combination method on the group of masses that are subtle on the prior view.

list	number of images	mass likelihood	gray scale correlation	simultaneous combination	sequential combination
original dataset	389	0.71	0.72	0.82	0.82
benign masses	140	0.66	0.74	0.82	0.79
malignant masses	249	0.75	0.69	0.82	0.84
subtle on prior view	37	0.76	0.68	0.78	0.84
obvious on prior view	57	0.74	0.67	0.84	0.84

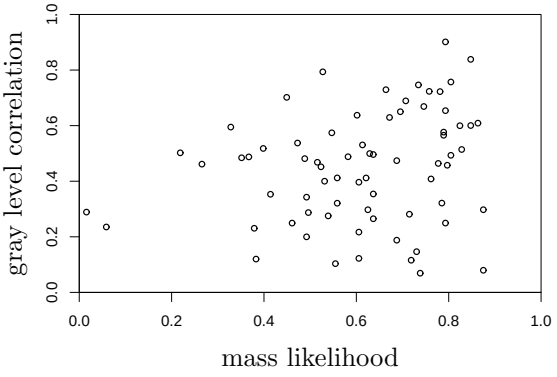
Table 5: Registration performance for different subsets. The second column gives the number of temporal image pairs in each subset. The other columns give the registration performance for each measure.

4.4 Linking errors

One of the best performing methods—the simultaneous combination method—links 18% of all lesions incorrect. To obtain insight into the possible causes of these linking errors we compare each of the three registration measures—correlation, mass likelihood and distance—at the selected location with the same measure at the correct location. If one measure performs substantially

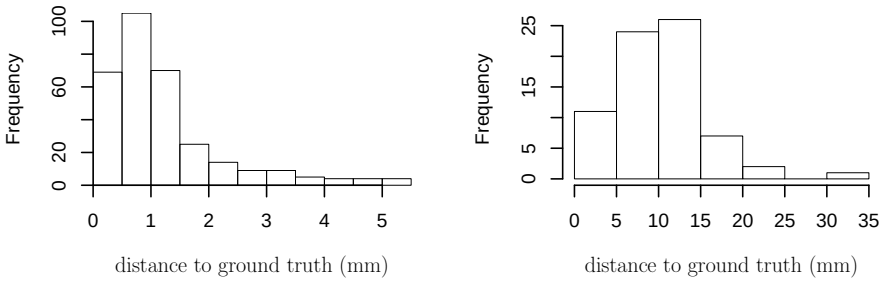


(a) correct link



(b) incorrect link

Figure 5: Scatter plot for correct and incorrect links



(a) Correct links

(b) Incorrect links

Figure 6: Histogram for the distance from the selected location to the ground truth for correct and incorrect links.

better at the selected location, we choose failure of this measure as the most important cause of the incorrect match. Table 6 shows that a combination of a low correlation and a low mass likelihood is the most common cause for an incorrect match. Most of these cases consist of a

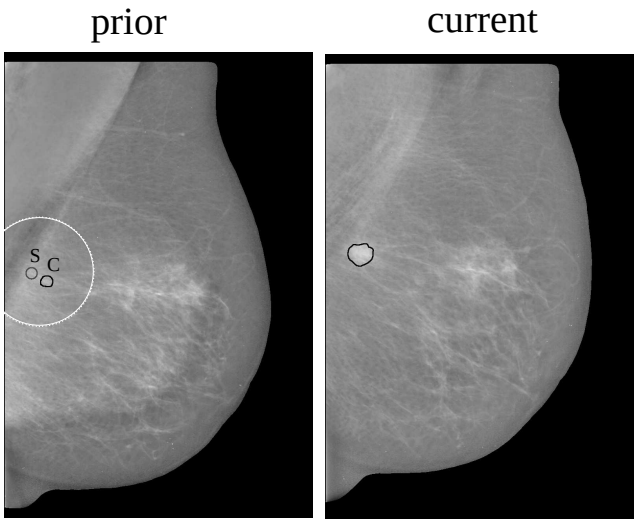
reason	percentage
combination of low correlation and low mass likelihood	38%
far from initial estimate	25%
low mass likelihood	21%
low correlation	17%

Table 6: Summary of the most important causes of linking errors

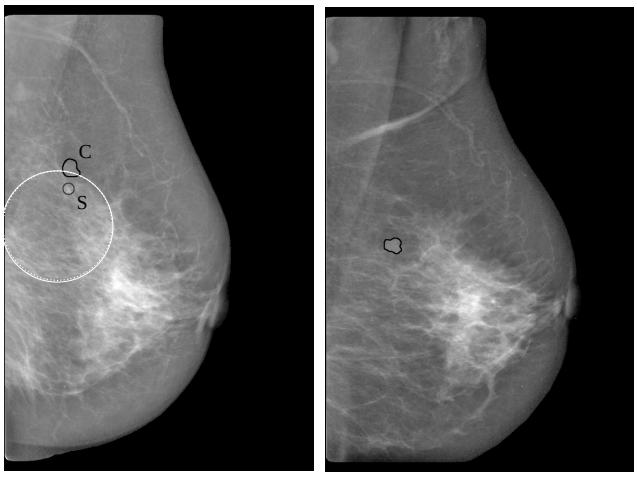
very subtle mass on the prior mammogram, which has consequences for both the correlation and the mass likelihood. The second most important cause of linking errors is a large distance to the initial estimate. In these cases the global registration method did not work very well. We find a low mass likelihood as cause for the linking errors for benign masses that are subtle on the prior image. Finally, we see that masses with a low correlation often change considerably between two consecutive screening rounds. Figure 7 gives some examples where the combined registration method failed. The letter C indicates the correct location, S the selected location. Figure 7(a) shows a very subtle mass on the prior view, in Figure 7(b) the mass on the prior view is outside the search area, and in Figure 7(c) the selected location shows a spiculated pattern resulting in a higher mass likelihood than the correct location.

5 Discussion

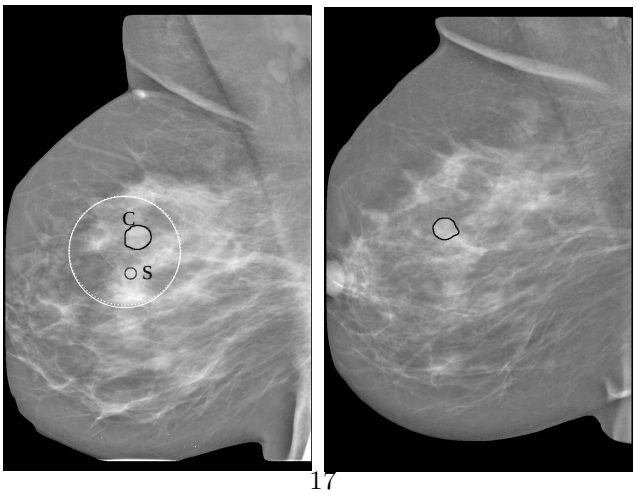
In this paper we present an automatic regional registration method that finds corresponding mass lesions in temporal mammogram pairs. This method combines three different registration



(a) Subtle mass on prior view



(b) Prior mass outside search area



(c) Selected location with higher mass likelihood than correct location

Figure 7: Examples of linking errors. The white circle indicates the search area. C is the correct location, S the selected location.

measures: a measure based on correlation, a measure based on mass likelihood and a distance criterion.

For the measure based on correlation we design different template shapes and investigate the influence of the template shape on the registration performance. For all shapes we use Pearson's correlation coefficient as similarity measure. In a recent study Filev et al.[9] found that this measure works best among a selection of twelve different similarity measures. Results for the different template shapes show that the best performing template is the growing mass template. We designed this template for masses that either grow or stay constant in time. Registration performance of the other two extended mass templates, the simple extended template and the circular template, was only slightly lower. This shows that the correlation measure is not very sensitive for small changes in template shape. The low performance of both inner and outer mass templates shows that both regions are necessary to obtain good registration results. We furthermore find that the minimal size of the outer border region is 1.4 mm and that the registration performance is similar for a larger outer border region. This observation differs from the study from Sanjay et al.[7] in which they found a decrease in registration performance with an increase in the size of the outer border region. The main difference between both studies is that Sanjay et al.[7] use a bounding box as template whereas our template depends on the contour of the mass. As most masses are more or less circularly shaped a bounding box will always contain surrounding tissue along some—but not all—parts of the contour. This might influence the registration results.

The registration method we propose in this paper combines the best performing correlation measure with a mass likelihood and a distance criterion. We call this method the combined registration method. On the complete set of masses this combined method links 82% of the masses correctly, compared with 71% for both individual measures.

To obtain information about the performance of each method on specific mass types we divide the original dataset into several subsets. The first subdivision of the original dataset is between benign and malignant masses. This subdivision shows that a correlation measure is more suited for benign masses. This may be explained by the fact that benign masses stay more or less constant over time resulting in a good correlation between both views. A measure based on mass likelihood is more suited for malignant masses. This measure will select a correct location on the prior view even when the mass has changed considerably, provided that the mass lesion on the prior is at the location with the highest mass likelihood. Furthermore, this measure also takes spiculation into account, which is a frequent sign of malignant masses.

The second subdivision is between masses that are very subtle and masses that are obvious on the prior view. The combined registration methods show that the sequential method is more suited to find subtle priors than the simultaneous method. This is in agreement with the observation that the correlation between subtle masses on the prior and the corresponding masses on the current is often quite low. Consequently, making a pre-selection of locations with a high mass likelihood increases the probability that the correct mass location is selected, on condition that this location has enough mass characteristics to be selected. When all locations are processed—like in the simultaneous combination method—some incorrect locations accidentally may have a high correlation, increasing the probability that an incorrect match occurs.

A computer aided detection (CAD) system can use a regional registration method to link selected regions on the current mammogram to corresponding locations on the prior mammogram. Combination of features from linked regions gives information about temporal changes. In a previous study we found that temporal features improve the detection performance compared to single view mammography, even when the current lesion is not visible on the prior view. We expect that the detection performance may profit even more from the use of temporal features when the CAD system first classifies all lesions on the current mammogram as new (not visible

on the prior view) or existing (visible on the prior view) and then calculates different temporal features for each group. We can use the combined registration measure as a threshold to discriminate between both groups. A low registration value then indicates a low probability that a corresponding lesion is present on the prior view and vice versa. Consequently we can classify regions with a low value as new and regions with a high value as existing.

6 Conclusion

In this study we compare different registration methods. We find that methods that combine several registration measures perform better than methods that use only one registration measure. The choice between both combination methods depends 1) on the number of regions initially detected by a CAD program and 2) on whether the CAD program aims at detecting all kinds of masses or only malignant ones. When the number of initial regions is quite large, what is common for CAD programs, the sequential combination method is preferred because it is very fast compared to the simultaneous method. The sequential method also performs better on the subset of malignant masses. On the other hand, we might choose the simultaneous method when the CAD program mainly aims at detecting benign lesions. Future research will concentrate on the development of temporal features that provide information about the temporal behavior of suspicious regions. We expect that this information will improve the detection performance of CAD programs.

References

[1] N. Wald, P. Murphy, P. Major, C. Parkes, J. Townsend, and C. Frost. "UKCCCR multicentre randomised controlled trial of one and two view mammography in breast cancer screening." BMJ **311**, 1189–1193 (1995).

[2] E. Sickles, W. Weber, H. Galvin, S. Ominsky, and R. Sollitto. "Baseline screening mammography: one vs two views per breast." AJR Am J Roentgenol. **147**(6), 1149–1153 (1986).

[3] M. Thurfjell, B. Vitak, E. Azavedo, G. Svane, and E. Thurfjell. "Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms." Acta Radiol. **41**(1), 52–56 (2000).

[4] M. Callaway, C. Boggis, S. Astley, and I. Hutt. "The influence of previous films on screening mammographic interpretation and detection of breast carcinoma." Clinical Radiology **52**, 527–529 (1997).

[5] W. Good, B. Zheng, Y. Chang, X. Wang, G. Maitz, and D. Gur. "Multi-image CAD employing features derived from ipsilateral mammographic views." In "SPIE conference on Image Processing," volume 3661, pages 474–485 (1999).

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

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

[8] L. Hadjiiski, H. Chan, B. Sahiner, N. Petrick, and M. Helvie. "Automated registration of breast lesions in temporal pairs of mammograms for interval change analysis - local affine transformation for improved localization." Med Phys **28**(6), 1070–1079 (2001).

[9] P. Filev, L. Hadjiiski, B. Sahiner, H. Chan, and M. Helvie. "Comparison of similarity measures for the task of template matching of masses on serial mammograms." Med Phys **32**(2), 515–529 (2005).

[10] S. Timp and N. Karssemeijer. "Interval change analysis to improve computer aided detection in mammography." Accepted for Medical Image Analysis **xx**(xx), xx (2005).

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

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

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

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

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