Computer Science and Information Systems 12(4):1255–1272 DOI: 10.2298/CSIS141230049M

Computer Aided Mass Detection in Mammography with Temporal Change Analysis

Fei Ma[1], Limin Yu[2], Gang Liu[1], and Qiang Niu[1]

[1] Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, China
{fei.ma, gang.liu, qiang.niu}@xjtlu.edu.cn
[2] Electric and Electronic Engineer, Xi'an Jiaotong-Liverpool University
Suzhou, China. limin.yu@xjtlu.edu.cn

Abstract. This paper presents a method to extract change information from temporal mammogram pairs and to incorporate the temporal change information in the malignant mass classification. In this method, a temporal mammogram registration framework which is based on spatial relations between regions of interest and graph matching was used to create correspondences between regions of current mammogram and regions of previous mammogram. 18 image features were then used to capture the differences (temporal changes) between the matched regions. To assess the contribution of temporal change information to the mass detection, 5 methods were designed to combine mass classification on image features measured on single regions and mass classification on temporal features to improve overall mass classification. The method was tested on 95 pairs of temporal mammograms using k-fold cross validation procedure. The experimental results showed that, when combining two classification results using linear combination or by taking minimum value, the A_z score of overall classification performance increased from 0.8843 to 0.8989 and 0.8863 respectively. The results demonstrated that registering temporal mammograms, measuring temporal changes from matched regions and incorporating the change information in the mass classification improves the overall mass detection.

Keywords: mammography, temporal change analysis, registration, mass detection.

1. Introduction

Breast cancer is the most common cancer in female population. In USA, breast cancer is the second frequently diagnosed cancer in women only next to skin cancer, and is the second leading cause of cancer death in women (after lung cancer) [2]. Early detection of breast cancer is believed to be the key to reduce the mortality. As the most effective method for early breast cancer detection, breast cancer screening has been introduced in many countries to provide intensive monitoring of breast cancer. For example, in Australia, women, primarily those between the ages of 50 to 69, are recommended and invited to attend screening every two years. Breast cancer screening can greatly reduce the risk of dying from breast cancer. Screening trial results conducted in USA suggests that mammography reduces mortality of breast cancer by 15% to 20%, while in Europe the studies suggests that the mammography can reduce the death rate of breast cancer by more than 1 - third [1].

Mammograms are read by radiologists to identify abnormalities. Normally multiple reads are used in order to increase sensitivity and specificity of diagnosing. Computer-aided diagnosis (CAD) programs are developed to assist radiologists in diagnosing disease

by providing processing results to the doctors as second opinion [23]. CAD programs are not designed to replace the radiologists. This is because, on one hand the CAD programs can not promise that the disease is always detected, on the other hand, the false positive rates of current CAD programs are normally high.

When reading the mammograms, radiologists usually compare the current mammogram with a previous one to check the changes between the two mammograms. Mammograms of the same breast but taken in different visits form temporal sequences (temporal mammograms) and are kept for future reference. The temporal change information is crucial for identification of malignant masses. Integrating the change information in CAD programs can potentially help radiologists in deciding if a mammogram contains benign or malignant masses [25]. However, incorporating such information in CAD programs is not a trivial task and has not been entirely successful so far. The difficulties lie in the nature of the mammogram which is a projection of a 3D breast in a 2D X-ray film. Natural changes of breast over time, different positioning of breast during acquisition, various X-ray dosage, different acquisition equipments, and most importantly, the unpredictable deformation of breast under compression during the acquisition contribute to the difficulty and complexity of extracting and utilizing temporal change information.

Many attempts have been made in analyzing the temporal changes contained in temporal sequences and using these information to improve mass detect in mammography. One approach is to simulate rigid or affine deformations between breasts in temporal mammograms. Hasegawa et al. used deformation of a B-spline control point grid to register dense regions of mammogram after a rigid body alignment[10]. Richard and Cohen used a variational formulation to simulate smooth deformation between mammogram pairs and applied the method to bilateral pairs Richard280 [18]. While their methods used warping functions to simulate the deformations, in practice, often the functions are not smooth, and not even well defined in some cases.

Another approach is based on locally searching in one mammogram to find matching regions of interest (ROI) of the paired mammogram. Sanay-Gopal et al. defined a fan-shaped searching region in previous mammogram based on the the nipple and the centroid of the breast for each mass like region in the current mammogram [20]. Sahiner et al. proposed a classification method based searching for matched anomalies in the mammogram for radiologist defined ROIs in the paired mammogram [9]. An extension approach is to integrate both local and global registration. Timp and Karssemeijer used the center of breast to register temporal mammograms globally and performed local searching to associate suspicious regions [24]. Marias et al. analyzed changes in two stages. The first stage was to align the two mammograms based on the alignment of breast boundaries. Multiresolution representations of the internal structure of the breasts was then matched based on thin plate spines in the second stage [17]. Wei et al. combined dual systems to conduct two-view analysis to improve cancer detection [28]. Casti et al. [4] used spherical semivariogram descriptors and correlation-based structural similarity indices in the spatial and complex wavelet domains to capture the changes in mammographic structural information of multiple mammogram views. They received a high detection accuracy through the proposed two-view analysis. A review of mammogram registration technique can be found in [8].

Comparison of different approaches are normally difficult, one of the main reason is different testing data sets used in different studies. Engeland et al. [6] compared the

performance of 4 methods, a nipple alignment based method, a Mutual Information (MI) based method, a method based on alignment of center of mass of the breast area and a warping based method. In this study, the MI based method was found superior to the others. Zheng et al. [33] compared three methods aimed at matching CAD-cued mass regions depicted on two views. Among the three search methods, the study found that the straight strip method required a smaller search area and achieved the highest level of CAD performance.

This paper presents a method of extracting temporal change information from temporal mammograms and apply the change information to the detection of malignant masses. This study is an extension of the framework introduced in [16], namely that of registering temporal mammograms based on fuzzy spatial relation representation and graph matching. This paper is organized as follows. In section 2, we briefly introduce the method of registering temporal mammograms based on spatial relations. Section 3 describes the image features used in this study, including image features extracted from single mammogram and image features extracted from comparison of temporal mammograms depicting temporal changes. For evaluation, the method is applied to 95 temporal mammogram pairs using k-fold cross validation procedure. Section 4 describes this data set and the experimental results. Finally, discussion and conclusions are made in section 5 and section 6.

2. Related work - temporal mammogram registration

The flowchart shown in Fig.1 describes the framework of registering temporal mammograms. In this framework, the preprocessing step consists of four operations; gamma correction, anisotropic filtering and extraction of both breast boundary and pectoral muscle boundary (for MLO view mammogram). Details of extraction algorithms of both breast and pectoral muscle boundaries can be found in [29] and [14]. After preprocessing, two image segmentation algorithms, adaptive pyramid (AP)([12]) based segmentation and sublevel set analysis([5][21]), were used to segment the images. In our experiment, the adaptive pyramid based segmentation algorithm is capable of isolating ROIs from the image, however, boundaries of the segmented ROIs are generally inaccurate. The sublevel set based analysis, with AP segmentation results as inputs, is used to improve the boundaries of mass like regions. Details of these two segmentation algorithms have appeared previously [15].

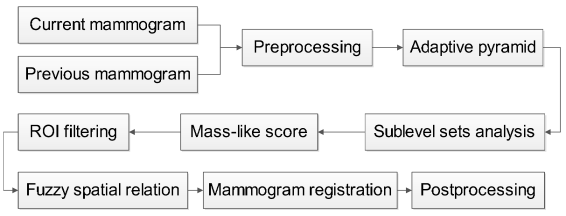

Fig. 1. Flowchart of temporal mammogram registration framework.

Segmentation of mammograms typically results in many small pieces or regions, not all of them correspond to masses, or are ROIs. To remove non-mass pieces, a mass-like score is calculated for every region, and a threshold is set on the mass-like score to remove some of the regions. The mass-like score is calculated using Fisher's linear discriminant and is based on 17 image features (see section 3). The threshold was set sufficiently low to only filter out obvious non-mass pieces, and hence many not-malignant-mass regions were still included. "ROI filtering" in Fig. 1 corresponds to this step.

After the ROI filtering step, a full graph was used to organize the left ROIs for each mammogram with every vertex representing a mass-like region. In the "Fuzzy spatial relation" step, spatial relations between ROIs were calculated and were associated to edges as weights. More explicitly, in this study, for a pair of points p and q, we compute four relationships, including *to the right of, to the left of, below*, and *above*

$$\mu_{\text{right}}(\theta) = \begin{cases} \cos^2\theta \text{ if } \frac{-\pi}{2} \le \theta \le \frac{\pi}{2}, \\ 0 \quad \text{otherwise}, \end{cases}$$

$$\mu_{\text{below}}(\theta) = \begin{cases} \sin^2\theta \text{ if } 0 \le \theta \le \pi, \\ 0 \quad \text{otherwise}, \end{cases}$$

$$\mu_{\text{above}}(\theta) = \begin{cases} \sin^2\theta \text{ if } -\pi \le \theta \le 0, \\ 0 \quad \text{otherwise}, \end{cases}$$

$$\mu_{\text{left}}(\theta) = \begin{cases} \cos^2\theta \text{ if } -\pi \le \theta \le \frac{-\pi}{2}, \frac{\pi}{2} \le \theta \le \pi, \\ 0 \quad \text{otherwise}, \end{cases}$$

where θ is the angle bounded by the line segment connecting p and q and the honrizontal line.

For two regions $C = \{c_1, c_2, \ldots, c_n\}, D = \{d_1, d_2, \ldots, d_m\}$, a multiset Θ was firstly defined as $\Theta = \{\theta_{ij} = \angle(c_i, d_j), c_i \in C, d_j \in D\}$, the histogram was then defined as

$$H_\Theta(C, D) = \{(\theta, f_\theta)\},$$

where f_θ is the count of point pairs (c_i, d_j) having $\angle(c_i, d_j) = \theta$. The f_θ is normalized by $f_\theta = f_\theta / \max(f_\theta)$.

To find the spatial relation between two objects is to evaluate the degree of H_Θ approaching the four spatial relations. For this purpose, the histogram H_Θ is treated as a fuzzy set whose membership function μ_H is defined as

$$\mu_{H_\Theta}(\theta) = f_\theta, \text{ where } (\theta, f_\theta) \in H_\Theta.$$

The degree of H_Θ approaching the four spatial relations was then calculated by finding the compatibility of fuzzy sets. Let G be one of the four spatial relations and μ_G be the corresponding membership function. The compatibility between two fuzzy sets H_Θ and G is a fuzzy set $CP(H_\Theta; G)$ with membership function

$$\mu_{CP(H_\Theta;G)}(v) = \begin{cases} \sup_{\theta, v=\mu_G(\theta)} \mu_{H_\Theta}(\theta) \text{ if } \mu_G^{-1}(v) \neq \emptyset, \\ 0 \qquad\qquad\qquad\qquad \text{if } \mu_G^{-1}(v) = \emptyset. \end{cases}$$

Finally to evaluate the degree to which a spatial relation among two objects holds, the center of gravity of the compatibility fuzzy set is calculated as

$$\frac{\sum_v v \cdot \mu_{CP(H_\Theta;G)}(v)}{\sum_v \mu_{CP(H_\Theta;G)}(v)}.$$

Creating correspondences between ROIs of two mammograms is equivalent to finding matched subgraphs in the two graphs. In the mammogram registration step Ullmann's backtracking algorithm [26] was used for the subgraph matching.

To pick the best matched subgraphs, a match cost function was defined based on the fuzzy spatial relations. For a subgraph $H_c = \{V_c, E_c\}$ with $V_c = \{v_c^1, v_c^2, \ldots, v_c^n\}$ and E_c the set of edges of H_c, and a subgraph $H_p = \{V_p, E_p\}$ with $V_p = \{v_p^1, v_p^2, \ldots, v_p^n\}$ and E_p the set of edges of H_p, suppose $v_c^k \in V_c$ corresponds to $v_p^k \in V_p$, the spatial similarity between v_c^k and v_p^k with regards to spatial relation μ_t is defined as

$$C_t(v_c^k, v_p^k) = \sqrt{\sum_{i=1}^{n} (\mu_t(v_c^k, v_c^i) - \mu_t(v_p^k, v_p^i))^2 + (\mu_t(v_c^k, B_c) - \mu_t(v_p^k, B_p))^2}, \tag{1}$$

where $t \in \Lambda = \{$*to the right of, to the left of, below,above*$\}$ is a spatial relation and $\mu_t(v_c, v_p)$ gives the μ_t spatial relation between two vertices v_c and v_p. B_c and B_p are the breast boundaries of current and previous mammograms. Treating breast boundaries as components and involving them in the matching process provides global references without explicit alignment of breast boundaries.

The final cost function is the sum of match costs over all matches and 4 spatial relations

$$\begin{aligned} \Phi(H_c, H_p) = &\sum_{i=1}^{n} C_{\text{left}}(v_c^i, v_p^i) + \sum_{i=1}^{n} C_{\text{right}}(v_c^i, v_p^i) \\ &+ \sum_{i=1}^{n} C_{\text{above}}(v_c^i, v_p^i) + \sum_{i=1}^{n} C_{\text{below}}(v_c^i, v_p^i). \end{aligned} \tag{2}$$

In a common graph matching problem, the best solution is reached when the longest sequence of matches was found. In this study, however, a longer solution (with more matched pairs) with false matched regions is not better than a shorter solution with more true matched regions. To find the best matches, a strategy is used. The best matches of different length were found firstly and combined to form a final matched sequence. To remove the false matches among the final matched sequence, firstly, the difference of spatial relations to the boundaries is computed. For a matched region (x_j, y_j), the difference of spatial relations to the boundaries d_j is calculated as

$$d_j = \sum_{t \in \Lambda} |\mu_t(x_j, B_x) - \mu_t(y_j, B_y)|. \tag{3}$$

Matched pair (x_j, y_j) is identified as a false match and is removed if $d_j > E(D) + \sigma(D)$. Here $E(X), \sigma(X)$ are the mean and standard deviation of X and $D = \{d_1, d_2, \cdots, d_n\}$ with n being the number of total pairs in the final match.

Due to the difference in mammogram acquisition process, gray levels of mammograms are normally not directly comparable. However, intensities of a mammogram contains important information regarding the existence of malignant mass. Spatial relation based registration is able to ensure compatible spatial relations among matched regions, but does not ensure compatible gray levels between matched regions. To remove false matches with incompatible gray level, such as a bright region in one mammogram matched to a dark region in the other mammogram, the relative gray level difference was used. Firstly, to transform the grey level of the first mammogram to match the gray level histogram of the second mammogram, the cumulative histograms of the intensity inside the breast area is calculated. For intensity level i, the cumulative histogram $f(i)$ of the mammogram is calculated as

$$f(i) = \frac{1}{Q}\sum_{j=0}^{i} H(j), \tag{4}$$

where $H(j)$ is the number of pixels inside the breast area that having intensity j, and Q is the total number of pixels inside the breast area. Let f_1 be the cumulative histogram of the first mammogram and f_2 be the cumulative histogram of the paired mammogram. Each gray level i_1 in f_1 is transformed to

$$i_2 = f_2^{-1}(f_1(i_1)). \tag{5}$$

After the histogram transformation, the relative gray level difference of a matched pair (x_i, y_j) is measured as

$$\text{Diff}(x_i, y_j) = \frac{|E(x_i) - E(y_j)|}{E(x_i) + E(y_j)}, \tag{6}$$

where $E(x)$ is the mean gray level of region x. A threshold ξ is then used and those pairs with Diff $> \xi$ are filtered out. In this paper, empirically, $\xi = 2.45$ is used.

Fig. 2 and 3 show matching results of two temporal mammogram pairs.

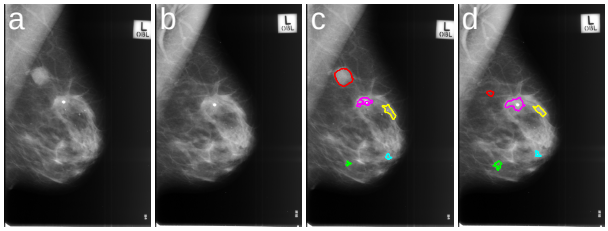

Fig. 2. Examples of match results. (a)(b) show a pair of temporal mammograms. (c) and (d), (e) and (f), and (g) and (h) show the same pair of temporal mammograms with matched regions depicted with white boundaries.

3. Image features

As mentioned in section 2, a mass like score was calculated for each ROI based on 17 image features. The following list describes these 17 image features.

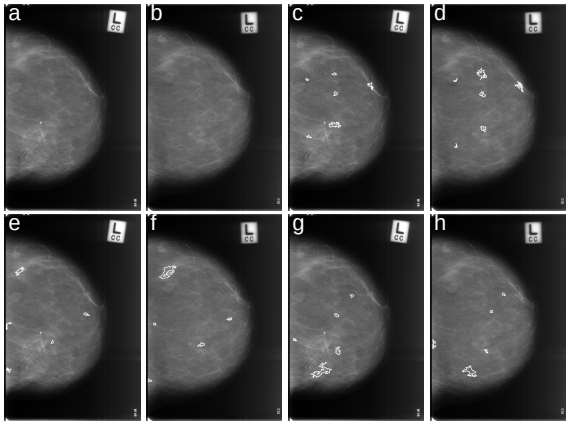

Fig. 3. Examples of match results. (a)(b) show a pair of temporal mammograms with 5 matched regions (matched regions are in same color).

- *solidity*. The number of pixels in the intersection of convex hull and the component divided by the number of pixels of the component. The convex hull refers to the smallest convex polygon that covers the component.
- *axis ratio*. The major axis of the ellipse whose normalized second central moments equals to the component divided by the minor axis of the same ellipse.
- *std radi*. Standard deviation of radial distances. The radial distance is the distance between a edge point to the centroid of component.
- *iv*. Variance of intensities along the component boundary.

$$iv = \frac{|E(R) - E(Q_2)|}{E(R) + E(Q_2)},$$

  Here $E(R)$ is the average intensity value of component R. Q_t is the set of pixels within t pixels distance to the component R but not in R.
- *c2*. A intensity contrast measurement.

$$c2 = \frac{(E(R) - E(Q_d))^2}{\sigma(R) + \sigma(Q_d)},$$

  where d is

$$d = \sqrt{\frac{area}{\pi}}.$$

  $\sigma(R)$ is the standard deviation of pixel intensities of R.
- *c3*. A intensity contrast measurement.

$$c3 = \sum_i |O(R, i) - O(Q_d, i)|,$$

  here $O(R, i)$ is the proportion of pixels in R that having intensity i, and d is the same as defined in *c2*.

– *int entropy*. Entropy of the intensity distribution. It is calculated as

$$\text{int entropy} = \sum_{k=1}^{256} O(R,k)\log(O(R,k))$$

– *Energy*

$$\text{energy} = \frac{\sum_{i,j} p(i,j)^2}{\text{size of ROI}},$$

here $p(i,j)$ is the intensity of pixel (i,j).

– *Inertial momentum* The luminosity inertial momentum is defined as

$$inertial = \sum_{i,j} p(i,j)d(i,j)^2$$

where $d(i,j)$ is the Euclidean distance between the (i,j)th pixel and the luminosity center (x_{lc}, y_{lc}), which is calculated by

$$x_{lc} = \frac{\sum_{i,j} p(i,j)i}{\sum_{i,j} p(i,j)} \quad y_{lc} = \frac{\sum_{i,j} p(i,j)j}{\sum_{i,j} p(i,j)}$$

– *Anisotropy*. Anisotropy is the distance between the geometric center and the luminosity center. It measures the irregularity. If the shape of a ROI have an isotropic symmetry this distance is close to 0 while it increases at high distortion.

– *Moment-based Shape Factors*
Let s_{pq} be the two-dimensional $(p+q)th$-order moment of a component and is defined as

$$s_{pq} = \sum_{i,j} i^p j^q p(i,j),$$

here $p, q = 0, 1, 2, \ldots$, the center of the component is defined as $\bar{x} = s_{10}/s_{00}$, $\bar{y} = s_{01}/s_{00}$. Seven low-order, central invariant, second- and third-order central moments are defined as:

$$\begin{aligned}
s_1 &= t_{20} + t_{02},\\
s_2 &= (t_{20} - t_{02})^2 + 4t_{11}^2,\\
s_3 &= (t_{30} - 3t_{12})^2 + (3t_{21} - t_{03})^2,\\
s_4 &= (t_{30} + t_{12})^2 + (t_{21} + t_{03})^2,\\
s_5 &= (t_{30} - 3t_{12})(t_{30} + t_{12})[(t_{30} + t_{12})^2\\
&\quad - 3(t_{21} + t_{03})^2] + (3t_{21} - t_{03})(t_{21} + t_{03})\\
&\quad [3(t_{30} + t_{12})^2 - (t_{21} + t_{03})^2],\\
s_6 &= (t_{20} - t_{02})[(t_{30} + t_{12})^2 - (t_{21} + t_{03})^2]\\
&\quad + 4t_{11}(t_{30} + t_{12})(t_{21} + t_{03}),\\
s_7 &= (3t_{21} - t_{03})(t_{30} + t_{12})[(t_{30} + t_{12})^2\\
&\quad - 3(t_{21} + t_{03})^2] - (t_{30} - 3t_{12})(t_{21} + t_{03})\\
&\quad [3(t_{30} + t_{12})^2 - (t_{21} + t_{03})^2],
\end{aligned}$$

where

$$\begin{aligned} t &= s_{00}, \;\; t_{10} = t_{01} = 0, \\ t_{20} &= (s_{20} - t\bar{x}^2), \\ t_{02} &= (s_{02} - t\bar{y}^2), \\ t_{11} &= (s_{11} - t\bar{x}\bar{y}), \\ t_{30} &= (s_{30} - 3s_{20}\bar{x} + 2t\bar{x}^3), \\ t_{03} &= (s_{03} - 3s_{02}\bar{y} + 2t\bar{y}^3), \\ t_{21} &= (s_{21} - s_{20}\bar{y} - 2s_{11}\bar{x} + 2t\bar{x}^2\bar{y}), \\ t_{12} &= (s_{12} - s_{02}\bar{x} - 2s_{11}\bar{y} + 2t\bar{x}\bar{y}^2). \end{aligned}$$

Each factor was normalized by dividing by t_r,where $r = 1 + (p + q)/2$.

The features were normalized according to the following formula:

$$f = \frac{f - \text{mean}(f)}{\text{std}(f)},$$

where mean and std are the average and standard deviation of f.

The mass like score is calculated by linearly combining the features measured on each ROI by

$$\text{mass-like score} = \sum_{t=1}^{17} a_t f_t,$$

where f_t is the tth feature. Parameters $a_t, t = 1 \ldots 17$ are optimized using linear discriminant analysis (LDA).

3.1. Temporal features

To depict the difference or changes between a pair of matched regions, a different set of 18 image features were used and the feature difference between the pair of matched regions is taken as the temporal feature value for the pair. For a region R_c of a current mammogram and a region R_p of the previous mammogram, for feature f, the temporal feature f_t of R_c and R_p is defined as

$$f_t = f_{R_c} - f_{R_p}.$$

In this paper, temporal features include 5 image features as listed below plus 12 image features which were also used in calculating mass like number. The mass like score calculated for ROI filtering is also used as the 18_{th} temporal feature. Details of the image features used as temporal features were listed in table 1.

- *solidity2*. The proportion of pixels in the smallest rectangle containing the region that are also in the component.
- *int*. Intensity of the component.
- *relint*. Intensity of the component divided by the average intensity of the whole breast area.
- *circularity*. $p^2/area$, here p is the perimeter of the component.
- *radi*. Average of the radial distance.

Single features	Matched features
solidity	solidity
axis ratio	solidity2
std radi	axis ratio
iv	circularity
c2	int
c3	relint
int entropy	std radi
energy	radi
inertial momentum	c2
anisotropy	c3
m1 - m7	int entropy
	energy
	inertial momentum
	anisotropy
	m2, m3, m7
	mass like number

Table 1. Single and temporal features.

4. Experiments

To investigate if the temporal features measured from the matched region pairs can contribute to the cancer detection [19], the temporal features will be applied to the malignant mass classification together with the image features measured from single regions. Section 4.1 firstly describes the data set to be used in the experiment. Section 4.2 introduces the k-fold cross validation process that the experiment will follow. In section 4.3, four different methods, Linear combination, Minimum value, Multiple classifiers and Combine to one, are used to combine the two kinds of features for the classification.

4.1. Experimental data

95 pairs of temporal mammograms were used to evaluate the performance of the proposed method in this paper. The images were from the database of BreastScreen SA (Adelaide, South Australia) and were selected subject to two conditions, firstly the current image contains histopathology proofed malignant mass, and secondly the previous image taken 2 - 3 years earlier is available and has no malignant mass found. Within the 95 pairs, 44 are CC view mammogram pairs and 51 are MLO views. 44 pairs contain histopathology confirmed malignant masses. A Vidar Diagnostic Pro Advantage digitizer was used to digitize all images at 48μm resolution and 12 bit depth. All images were subsampled and every 8$\times$8 pixel patches were replaced by one pixel whose intensity equals the average intensity of the patch.

4.2. K-fold cross validation

To assess the performance of classifications with two sets of image features, we applied the k-fold cross validation procedure. Each mammogram in the data set was firstly labelled as malignant if it contains malignant mass or normal otherwise. The two labels

split the whole data set into two classes. Next, both classes were randomly and evenly divided into k subgroups. Of the k subgroups, k-1 subgroups are joined together to form a training set while the left one is retained as the test set. The cross-validation process is then repeated k times. Each time a different subgroup is used as test set. For k times, all subgroups were used as test set for exactly once. In the end, the k results were combined as the final classification results. In this paper, considering the size of the sample data, we use k=5.

4.3. Mass classification with combination of two sets of image features

Classifying a image region as malignant or benign, based on a series of image features, is generally the final but most important step in medical image analysis [11][22][3][27][30][31][32]. In this study, the process of segmenting a mammogram into regions, extracting 17 image features for each region, using Fishers linear discriminant to calculate a mass-like score for each region, and classifying each region as benign or malignant using the calculated mass-like score, comprises a mass detection scheme. We refer to this scheme as single classification. Registering a pair of temporal mammograms, extracting 18 temporal features for each pair of matched regions, and using the temporal features to classify the corresponding regions as malignant or benign, also constitutes a mass detection scheme. We refer to this scheme as temporal classification. Note that the 18 temporal features were only calculated on the ROIs that have formed matches during the mammogram registration, thus not all ROIs have been calculated temporal features. The ROIs involved in the temporal classification were much less comparing to that in single classification.

To utilize the temporal change information to aid the detection of malignant masses means to combine the single and temporal features to classify a ROI as a malignant or benign mass. In this study, to assess the contribution of the temporal features to the malignant mass detection, four different methods were used to combine the result of single classification and temporal classification.

Linear combination In this method, the two classification results were combined linearly to generate a new set of classification score. For a region of interest R_i, the new score c_i is chosen as:

$$c_i = \begin{cases} \alpha s_i + \beta t_i & \text{if } R_i \text{ has } t_i, \\ s_i & \text{otherwise,} \end{cases} \tag{7}$$

where s_i is the single classification score and t_i the temporal classification score of R_i. α and β are two coefficients with properties $\alpha \geq 0, \beta \geq 0$ and $\alpha + \beta = 1$.

To decide the optimal α and β, a heuristic searching procedure was used. α was varied from 0.1 to 1 with increment 0.01 while $\beta = 1 - \alpha$. Some combinations of α and β were listed in table 2. From the table, it can be seen that the detection rate measured by A_z score increases with the increase of α until $\alpha = 0.83$ and $\beta = 0.17$ where the maximum detection rate $A_z = 0.8867$ was reached. Hence $\alpha = 0.83$ and $\beta = 0.17$ was used in this experiment.

Furthermore, for this method, based on the temporal classification results, a safe upper threshold and low threshold were chosen. Regions whose temporal scores were below the safe low threshold were identified as non masses, and regions whose temporal scores were

α	β	A_z	α	β	A_z
0.1	0.9	0.8689	0.55	0.45	0.8827
0.2	0.8	0.8729	0.65	0.35	0.8848
0.3	0.7	0.8755	0.77	0.23	0.886
0.41	0.59	0.8782	0.83	0.17	0.8867
			0.89	0.11	0.8866

Table 2. Different combinations of α and β and their corresponding detection performance.

above the safe upper threshold were identified as true masses. Finally, the adjusted scores were used to evaluate the performance of the classification. Fig. 4 shows the results of the linearly combination classification together with single and temporal classifications.

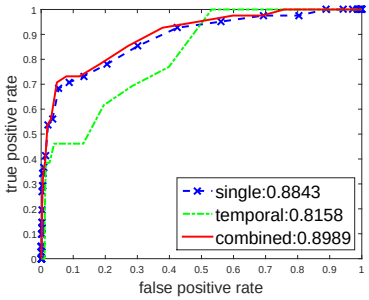

Fig. 4. ROC curves of single, temporal and linearly combined classifications.

Minimum value In this method, for a region of interest R_i, a new score c_i is chosen as:

$$c_i = \begin{cases} \min(s_i, t_i) & \text{if } R_i \text{ has } t_i, \\ s_i & \text{otherwise,} \end{cases} \tag{8}$$

where s_i is the single classification score and t_i the temporal classification score of R_i.

In our testing images the previous mammogram taken 2 -3 years earlier was judged as normal, hence if a ROI in the current mammogram were matched to a ROI in the previous mammogram, then more likely the ROI in the current mammogram is not malignant. This method incorporates this knowledge. It decreases the scores of matched ROIs. The safe upper and low thresholds were also applied in this method. Fig. 5 shows the performance of the new classification together with single and temporal classifications.

Multiple classifiers In this method, the two classification results were used as input into a third classifier to produce a new set of classification score. Because the temporal classification results contain less data compared to the single classification results. To compensate the shorten length, the temporal classification scores were expanded to the same size

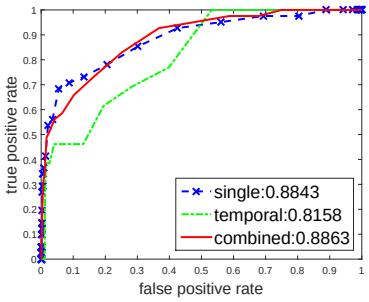

Fig. 5. ROC curves of single, temporal and combination with minimum value classifications.

as the single classification scores. A ROI appearing in the temporal classification retains the temporal classification score while a ROI not appearing in the temporal classification is treated as a potential mass and all its temporal features are set to 1.

In this experiment, two classifiers were tried as the third classifier, the Fisher's linear discriminant and the support vector machine (SVM) [13, 7] with gaussian radial basis function (RBF) kernel. Fig. 6 and Fig. 7 show the ROC curves and A_z scores of the multiple classifications with Fisher's linear discriminant and SVM with RBF kernel as the third classifier respectively.

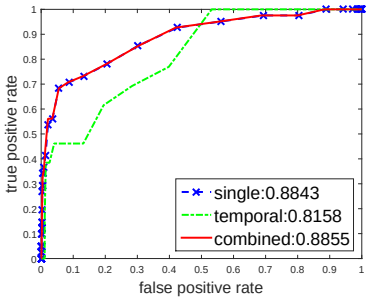

Fig. 6. ROC curve and A_z score of single and temporal classifications and multiple classification with Fisher's linear discriminant as the third classifier.

Combine to one In this method, two sets of image features were combined together to form a set of 35 image features. ROIs that have no temporal feature values will have all their temporal features to be 0. The Fisher's linear discriminant was applied to these 35 image features to generate a new classification score. The classification evaluation results are shown in Fig. 8.

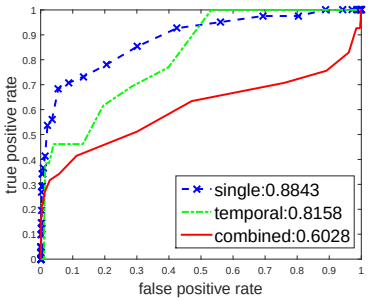

Fig. 7. ROC curve and A_z score of single and temporal classifications and multiple classification with SVM with RBF kernel as the third classifier.

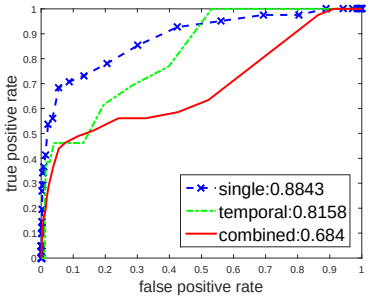

Fig. 8. ROC curve of single and temporal classifications and classification on combined feature sets.

5. Discussion

The A_z score made by the classification based on the temporal features alone is 0.8158, which is lower than 0.8481, the A_z score made by the classification based on the mass like number. The temporal features by themselves do not contribute to better classification. This is because on one hand temporal features can only be extracted from matched regions, and hence not all ROIs were associated with temporal features. On the other hand, in this study, the temporal mammogram pairs were so selected that masses were found on the current mammogram but were not identified on the previous mammogram, so true malignant masses in many current mammograms do not correspond to any regions in the previous mammogram. Many ROIs corresponding to true malignant masses were missing in the temporal classification, which contributed to the low score of classification with temporal features alone.

The experiments show that incorporating the temporal features in the mammogram mass detection is able to improve the overall performance. The detection rates (measured by A_z score) with linear combination and combination by taking minimum value both obtained a small gain. With linear combination, the A_z score increased from 0.8843 to 0.8989 (see Fig. 4). Combining the two sets of classification results by taking the minima

also improves the overall classification with A_z score increased from 0.8843 to 0.8863 (see Fig. 5). However, not all combinations result in improvement. In the last method when two sets of image features were combined into one set, the classification performance dropped sharply with A_z score decreased from 0.8843 to 0.684. Image features measured on single ROIs and features measured on the matched region pair are different types of image features, the last method suggests that mixing these two types of image features does not improve the detection performance.

Multiple classifiers or multilayered classifiers are popular ways in handling classifications with multiple sets of features and have been employed in many different applications, such as sentiment classification, texture classification and image recognition. Multiple classifiers, however, are not explored much in cancer detection in mammography. In this study, the multiple classifiers method firstly used 2 classifiers on the temporal and single feature sets and a third classifier on the results of the first 2 classifiers. Both Fisher's linear discriminant and SVM with RBF were used as the third classifier. The results, however, are very different. With Fisher's linear discriminant, detection rate was increased sightly with A_z score increased from 0.8843 to 0.8855, but when using SVM with RBF as the third classifier, the A_z score dropped to 0.6028.

6. Conclusion

In this paper we proposed a method to incorporate temporal change information in mass detection in mammography and investigated how the temporal change information can contribute to the improvement of mass detection. In the proposed method, temporal mammograms were firstly registered under a mammogram registration framework which was based on spatial relations between segmented ROIs and graph matching. The temporal mammogram registration creates correspondences between regions of current mammogram and regions of previous mammogram. Based on the registration, changes between temporal mammograms were extracted using a set of 18 temporal features. To assess the contribution of change information to the mass detection, 5 methods were designed to incorporate temporal features containing temporal changes for malignant mass classification. In the experiment, k-fold cross validation was used to evaluate the classifications. Our experiments show that, when combining the single classification and temporal classification results linearly or by taking the minimum value of two classification, the overall classification results were improved, demonstrating that temporal changes information improves mass detection. However, when using multiple Fisher linear discriminant classifiers or applying Fisher linear discriminant classifier directly to the combined image feature set, no improvement or even decreasing on the overall classification performance were observed.

The experiment shows a small gain on the overall detection rate on the data set consisting 95 pairs of real temporal mammograms. A larger set of mammogram pairs or more sets of mammogram pairs from different sources can help to better identify the contribution of the temporal change information to the cancer detection. Unfortunately such data are unavailable for this study. On the other hand, the focus of this study is to investigate if change information extracted from temporal mammogram pairs can improve the cancer detection. For this purpose, the experiment indicated that the change information is able to improve the overall detection. The detection scheme involves many steps, such as

image preprocessing, image segmentation, feature extraction, feature selection and classification. Each of these steps affects the final detection rate. This study used only 5 initial methods in incorporating temporal image features in the detection. A further study could be to explore different ways in combining the two different feature sets.

In this study, Fisher's linear discriminant was mainly used to assess the performance of different classifications. A thorough assessment may include other classification methods that are popular in mass classification in mammography, such as k-nearest neighbor (KNN), kernel Fisher discriminant (KFD) and relevance vector machine (RVM), to better evaluate the contributions of temporal change information to the malignant mass detection in mammography.

In many classification tasks, with carefully chosen parameters, SVM has been reported to be able to produce better results, comparing to many other classification methods. In this study, SVM with RBF was used as the third classifier with results from Fisher's linear discriminant classifier on single and temporal feature sets as input. The result was much lower comparing to the results made by using the Fisher's linear discriminant as the third classifier on the same input. This may relate to the input data, which were produced by Fisher's linear discriminant on single and temporal feature sets, while the temporal feature set was stretched to be compatible with the single feature set. On the other hand, only few different values of parameters were tried for the SVM method used in this study. With better chosen parameters, using SVM as the third classifier could produce better results. This will be investigated in the future study.

Acknowledgment. This work was supported partly by the National Natural Science Foundation of China (Grant No. 11301420).

References

1. American Cancer Society: Breast Cancer Facts & Figures 2013-2014. Atlanta: American Cancer Society, Inc. (2013)
2. American Cancer Society: Cancer Facts & Figures 2015. Atlanta: American Cancer Society (2015)
3. Bai, X., Fang, Y., Lin, W., Wang, L.P., Ju, B.F.: Saliency-based defect detection in industrial images by using phase spectrum. IEEE Trans. Industrial Informatics 10(4), 2135–2145 (2014)
4. Casti, P., Mencattini, A., Salmeri, M., Rangayyan, R.M.: Analysis of structural similarity in mammograms for detection of bilateral asymmetry. Medical Imaging, IEEE Transactions on 34(2), 662–671 (Feb 2015)
5. Eltonsy, N.H., Tourassi, G.D., Elmaghraby, A.S.: A concentric morphology model for the detection of masses in mammography. IEEE Transactions on Medical Imaging 26(06), 880–889 (June 2007)
6. Engeland, S.V., Snoeren, P., Hendriks, J., Karssemeijer, N.: A comparison of methods for mammogram registration. IEEE Trans. Med. Imaging 22(11), 1436–1444 (Nov 2003)
7. Fu, X.J., Wang, L.P.: Data dimensionality reduction with application to simplifying rbf network structure and improving classification performance. IEEE Trans. System, Man, Cybern, Part B-Cybernetics 33(3), 399–409 (2003)
8. Guo, Y., Sivaramakrishna, R., Lu, C., Suri, J., Laxminarayan, S.: Breast image registration techniques: a survey. Med. Biol. Eng. Comput. 44, 15–26 (2006)
9. Hadjiiski, L., Sahiner, B., Chan, H.P., Petrick, N., Helvie, M.A., Gurcan., M.: Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses. Med. Phys. 28(11), 2309C2317 (2001)

10. Hasegawa, A., Neemuchwala, H., Tsunoda-Shimizu, H., Honda, S., Shimura, K., Sato, M., Koyama, T., Kikuchi, M., Hiramatsu, S.: A tool for temporal comparison of mammograms: Image toggling and dense-tissue-preserving registration. In: Digital Mammography 9th International Workshop, IWDM 2008. p. 447C454. Springer, Tucson, AZ, USA (2008)
11. Hundza, S.R., Hook, W.R., Harris, C.R., Mahajan, S.V., Leslie, P.A., Spani, C.A., Spalteholz, L.G., Birch, B.J., Commandeur, D.T., Livingston, N.J.: Accurate and reliable gait cycle detection in parkinson's disease. IEEE Transactions on Neural Systems and Rehabilitation Engineering 22, 127–137 (2014)
12. Jolion, J.M., Montanvert, A.: The adaptive pyramid: a framework for 2D image analysis. Computer Vision, Graphics, and Image Processing 55(3), 339–348 (May 1992)
13. Kecman, V.: Learning and Soft Computing, Support Vector machines, Neural Networks and Fuzzy Logic Models. The MIT Press, Cambridge, MA (2001)
14. Ma, F., Bajger, M., Slavotinek, J.P., Bottema, M.J.: Two graph theory based methods for identifying the pectoral muscle in mammograms. Pattern Recognition 40, 2592–2602 (2007)
15. Ma, F., Bajger, M., Bottema, M.J.: Automatic mass segmentation based on adaptive pyramid and sublevel set analysis. In Proceedings of DICTA09: Digital Image Computing: Techniques and Applications pp. 236–241 (Dec 2009)
16. Ma, F., Yu, L., Bajger, M., MurkJ.Bottema: Incorporation of fuzzy spatial relation in temporal mammogram registration. Fuzzy Sets and Systems 279, 87–100 (2015)
17. Marias, K., Behrenbruch, C., Parbhoo, S., Seifalian, A., Brady, M.: A registration framework for the comparison of mammogram sequences. IEEE Trans. Med. Imaging 24(6), 782–790 (2005)
18. Richard, F.J.P., Cohen, L.D.: A new image registration technique with free boundary constraints: application to mammography. Computer Vision and Image Understanding 89, 166C196 (2003)
19. Samulski, M., Karssemeijer, N.: Optimizing case-based detection performance in a multiview cad system for mammography. IEEE Trans. on Medical Imaging 30, 1001–1009 (2011)
20. Sanjay-Gopal, S., Chan, H., Wilson, T., Helvie, M., Petrick, N., Sahiner, B.: A regional registration technique for automated interval change analysis of breast lesions on mammograms. Med. Phys. 26(12), 2669–2679 (1999)
21. Shiffman, S., Rubin, G.D., Napel, S.: Medical image segmentation using analysis of isolable-contour maps. IEEE Transactions on Medical Imaging 19(11), 1064–1074 (Nov 2000)
22. Stone, E.E., Skubic, M.: Fall detection in homes of older adults using the microsoft kinect. IEEE Journal of Biomedical and Health Informatics 19, 290–301 (2015)
23. Tai, S.C., Chen, Z.S., Tsai, W.T.: An automatic mass detection system in mammograms based on complex texture features. Biomedical and Health Informatics, IEEE Journal of 18(2), 618–627 (March 2014)
24. Timp, S., Karssemeijer, N.: Interval change analysis to improve computer aided detection in mammography. Medical Image Analysis 10(1), 82–95 (2006)
25. Timp, S., Varela, C., Karssemeijer, N.: Computer-aided diagnosis with temporal analysis to improve radiologists interpretation of mammographic mass lesions. IEEE Trans. on information technology in biomedicine 14(3) (May 2010)
26. Ullmann, J.R.: An algorithm for subgraph isomorphism. Journal of the ACM 1(23), 31–42 (1976)
27. Wang, C.W., Huang, C.T., Hsieh, M.C., Li, C.H., Chang, S.W., Li, W.C., Vandaele, R., Maree, R., Jodogne, S., Geurts, P., Chen, C., Chu, G.Z.C., Mirzaalian, H., Hamarneh, G., Vrtovec, T., Ibragimov, B.: Evaluation and comparison of anatomical landmark detection methods for cephalometric x-ray images: A grand challenge. IEEE Transactions on Medical Imaging 34, 1890–1900 (2015)
28. Wei, J., Chan, H.P., Sahiner, B., Zhou, C., Hadjiiski, L.M., Roubidoux, M.A., Helvie, M.A.: Computer-aided detection of breast masses on mammograms: Dual system approach with two-view analysis. Med. Phys. 36(10), 4451–4460 (2009)

29. Yu, L., Ma, F., Jayasuriya, A., Sigelle, M., Perreau, S.: A new contour detection approach in mammogram using rational wavelet filtering and MRF smoothing. In: 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007). pp. 106–111. IEEE., Glenelg, South Australia (2007)
30. Zhang, F., Song, Y., Cai, W., Lee, M.Z., Zhou, Y., Huang, H., Shan, S., Fulham, M., Feng, D.: Lung nodule classification with multilevel patch-based context analysis. IEEE Transactions on Biomedical Engineering 61, 1155–1166 (2014)
31. Zhang, L., Wang, L.P., Lin, W.: Conjunctive patches subspace learning with side information for collaborative image retrieval. IEEE Trans. Image Processing 21(8), 3707–3720 (2012)
32. Zhang, L., Wang, L.P., Lin, W.: Semi-supervised biased maximum margin analysis for interactive image retrieval. IEEE Trans. Image Processing 21(4), 2294–2308 (2012)
33. Zheng, B., Tan, J., Ganott, M.A., Chough, D.M., Gur, D.: Matching breast masses depicted on different views: A comparison of three methods. Acad. Radiol. 16(11), 1338–1347 (2009)

Fei Ma received his M.Sc. from the Xiamen University, China in 2002 and his Ph.D. Degree in Applied Mathematics from Flinders University, Australia in 2008. Currently he is a lecturer at the Mathematical Sciences Department of Xi'an Jiaotong-Liverpool University. His research interests include image processing, medical image analysis, forecasting methods and nonnegative matrix.

Limin Yu received the B.Eng. degree in telecommunications engineering and the M.Sc. degree in radio physics/underwater acoustic communications from Xiamen University, China. She received the PhD in telecommunications engineering from the University of Adelaide, Australia. She worked at ZTE Telecommunication Company, Limited, Shenzhen, China as a software engineer. She also worked at the South Australia University and University of Adelaide as a research fellow and research associate. Dr. Yu is currently a lecturer at the Electrical and Electronic Engineering Department of Xian Jiaotong-Liverpool University, Suzhou, China. Her research interest includes sonar detection, wavelet analysis, multirate filter banks, sensor network and medical image analysis.

Gang Liu received the B.S. degree in mathematics (with honors) from Liaoning Normal University, the Research M.S. degree in topology on lattice from the Ocean University of China and the Ph.D. degree for research on nonlinear control systems from The University of Sheffield. He worked as a lecturer at Ocean University of China and Research Associate at the University of Sheffield. He joined Xian Jiaotong Liverpool University in 2010 and was promoted to Associate Professor in 2014. His research interests include sliding-mode control, differential geometric control, differential geometry, topological controlled dynamics, and topology and mathematical modeling of complex systems.

Qiang Niu completed his Ph.D. Degree in Computational Math in Xiamen University in 2008. He is currently an Associate Professor in Xi'an Jiaotong-Liverpool University. Qiang Nius main research area is Numerical Linear Algebra. He is interested in matrix computation problems, particularly, the large sparse matrix eigenvalue problems, the high performance preconditioning techniques for large system of linear equations, iterative methods, information retrieval and image processing.

Received: December 30, 2014; Accepted: June 12, 2015.