A Registration Framework for the Comparison of Mammogram Sequences

Kostas Marias*, *Member, IEEE*, Christian Behrenbruch, Santilal Parbhoo, Alexander Seifalian, and Michael Brady, *Senior Member, IEEE*

***Abstract*—In this paper, we present a two-stage algorithm for mammogram registration, the geometrical alignment of mammogram sequences. The rationale behind this paper stems from the intrinsic difficulties in comparing mammogram sequences. Mammogram comparison is a valuable tool in national breast screening programs as well as in frequent monitoring and hormone replacement therapy (HRT). The method presented in this paper aims to improve mammogram comparison by estimating the underlying geometric transformation for any mammogram sequence. It takes into consideration the various temporal changes that may occur between successive scans of the same woman and is designed to overcome the inconsistencies of mammogram image formation.**

***Index Terms*—Biomedical image processing, image registration, image sequence analysis, image matching.**

I. Introduction

THE reliable diagnosis of abnormalities from a single mammogram is an extremely difficult task even for a skilled radiologist, and so it is increasingly the case that pairs of mammograms are compared. These may be, for example, the left and right mammograms taken at the same session. Equally, when mammograms from an earlier time are available, as is increasingly the case, the radiologist will routinely compare the older and more recent images. For this reason alone, the development of mammogram registration is increasingly important for the early detection of pathology.

However, mammogram registration cannot usefully be approached as a specific instance of a general (nonrigid) medical image registration problem, because of variations in breast compression and in imaging parameters, and because of variations in the shape and amount of the pectoral muscle present in the medio-lateral projection as well as the large amount of textural detail—particularly curvilinear structures (CLS)—which together massively affect the content and appearance of the two images. Herein, we present a novel method for temporal mammogram registration to assist the clinician in comparing temporal mammograms more effectively. The main contribution of this paper is a multiscale curvature-based algorithm for aligning the breast outline combined with a wavelet-based technique for matching a small number of distinct internal regions.

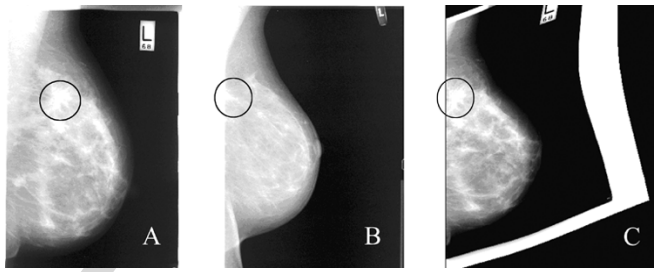


Fig. 1. Although registration can be limited by the actual changes over time, it improves the correspondence of regions; C is mammogram A registered to the most recent one —B. This way, by comparing B and C, the clinician can be reassured that the encircled lesion in B is scar tissue from the excision of a previous cancer in the same region (encircled in the same location in C).

Mammogram registration cannot usually be applied directly for the automatic detection of small abnormalities since the breast is a highly dynamic organ in which numerous normal changes occur regularly. This way, subtracting registered mammogram sequences may lead to a large number of false positives (in this context defined as normal changes and misregistrations), appearing in the difference image. For this reason, we contend that registration should be used primarily as a tool for aiding the clinician to better understand changes over time. This is illustrated in Fig. 1, where two mammograms of the same woman (A and B) are shown together with C, the registered version of A to B. The differences in imaging conditions and the temporal changes (4-yr interval) render the comparison of A and B difficult, while by comparing B (recent mammogram) and C the clinician can be reassured that the encircled region in B is scar tissue from the excision of a previous cancer in C (also encircled in the same location). This aspect of mammogram registration is the basic motivation of our work.

The potential clinical applications are as follows:

1) for women at high risk of developing breast cancer (e.g., women with family history of breast cancer or genetic susceptibility), usually have more frequent mammograms taken in order to detect a malignancy at as early a stage as possible. Previous ("normal") mammograms are used as a baseline for comparison with recent ones.
2) for postmenopausal women who often decide themselves, or are advised by their GP, to undergo hormone replacement therapy (HRT). However, there is a suggestion,

Manuscript received October 22, 2004; revised March 23, 2005. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was C. Meyer. *Asterisk indicates corresponding author.*

*K. Marias was with the Medical Vision Laboratory, Department of Engineering Science, Oxford University, Parks Road, Oxford OX1 3PJ, U.K. He is now with the ICS-FORTH, Vassilika Vouton, P.O. Box 1385, GR 711 10 Heraklion, Greece (e-mail: kmarias@ics.forth.gr).

C. Behrenbruch is with Mirada Solutions Ltd., Beaver House, Oxford, U.K. (e-mail: Christian.behrenbruch@mirada-solutions.com).

S. Parbhoo and A. Seifalian are with the Department of Surgery, Royal Free and University College Medical School, UCL, London NW3 2QG, U.K.

M. Brady is with the Medical Vision Laboratory, Department of Engineering Science, Oxford University, Oxford OX1 3PJ, U.K. (e-mail: jmb@robots.ox.ac.uk).

Digital Object Identifier 10.1109/TMI.2005.848374

0278-0062/$20.00 © 2005 IEEE

IEEE PROOF

based on clinical experience, that significant regional increases in tissue density could be an early indicator of breast cancer for women using HRT. For this reason, it is important to be able to register HRT mammogram sequences, aiming at a more effective comparison for early detection of lesions.

3) for retrospective studies that aim to analyze temporal data in order to assess the accuracy and effectiveness of diagnosis in hospitals/screening centres. Such studies aim to define the rate of missed cancers and interval cancers, as well as to further educate clinicians in the important task of early diagnosis. Mammogram registration could potentially be used to trace back "missed cancers" in previous mammogram sessions.

The novelty of the method presented here derives mainly from the detailed understanding of the temporal changes that can occur between successive acquisitions. Breast compression and imaging condition variability often lead to a nonrigid transform in the image plane (even a small difference in compression can lead to a significant and uneven displacement of the breast structures), and a nonrigid transformation of image intensity, respectively. Temporal changes can also occur due to breast positioning resulting in a rotation of the breast between two mammographic exposures. Our experience leads us to conclude that these are small, certainly less than ten degrees (around an informally defined axis from nipple to pectoral muscle). It is well known from work in computing structure from visual motion, that such small rotations are often indistinguishable from translations. Our assumption is that any such rotations can be accommodated within the nonrigid framework that we propose and that they do not need to be computed explicitly. Larger rotations could lead to significant changes in the two-dimensional (2-D) image and reduce the applicability of registration methods. In any case, a prior affine registration is often used to cope with such changes—the physical limitations on the taking of a mammogram obviate the need for such a prior acquisition in most cases.

In addition, normal changes (e.g., weight gain or loss, involution) as well as affects of HRT or chemotherapy agents such as Tamoxifen can introduce more differences in the architecture of a temporal pair. The intensities of a mammogram pair can be normalized using the h_{int} representation of interesting tissue [1] which results in the standard mammogram form (SMF), a standardized representation of a mammogram computed from the image intensities (film or digital) and imaging parameters of the system used to acquire the image. The method described below deals with the *geometrical alignment* of mammogram sequences. Critically, however, initial results presented herein, indicate significant improvement in the registration when image normalization is incorporated into the registration process.

Previously, promising results have been obtained for mammogram registration using the breast outline, for example. Bowyer and Sallam [2], and Karssemeijer and te Brake [3]. The effectiveness of the boundary as a basis for registration reflects the fact that though the breast is compressed, it tends to be compressed to approximately the same extent in each session, since the breast contents mostly do not vary greatly between those sessions. Highnam and Brady [1], report that the average difference in breast compression is about 0.5 cm. However, neither method calculates consistent points along the boundary; in [2], a curvature linear expansion algorithm is used to match boundary points, while in [3], the whole segmented outline is sampled without establishing boundary correspondences. Due to the nonrigid expansion of the breast during compression, the assumption of linear expansion in the curvature of the breast outline is not always valid. Nevertheless, we have observed that a small number of "pseudoinvariant" points can be detected reliably, an idea that has been exploited in our work [4]. No previous method incorporates a robust method to select consistent boundary landmarks for the *automatic alignment* of the breast boundary. This is a novelty of our method.

To completely solve the registration problem, internal correspondences need to be defined. This is a difficult problem since different kinds of breast tissue move according to their mechanical properties and the compression exerted and in addition, the anatomy may vary significantly between sessions. Bowyer and Sallam [2] use steerable filters to determine a large number of point correspondences in bilateral images. However, implementation of this method in temporal mammogram registration yields a large number of erroneous landmarks, since the image content changes significantly between acquisitions. Vujovic and Brzakovic [5], and more recently Marti *et al.* [6], suggest using curvilinear mammogram structures corresponding to ducts, connective stroma and blood vessels to define point correspondences between mammograms forming a temporal mammogram pair. However, such an approach intrinsically lacks robustness since the curvilinear structures that are visible in a mammogram, and in particular their intersections, are extremely sensitive to the extent of breast compression. Highnam and Brady show that breast compression changes as small as 0.25 cm can significantly alter the curvilinear structures visible in the mammogram, a compression that is half that likely to occur in temporal registration. Kok-Wiles *et al.* [7] developed a method for matching salient regions (iso-contours) between mammograms. The main problem with this method is the large number of salient regions (iso-contours that fulfil the saliency constraints) and heuristics necessary for the matching, though Hong and Brady [8] have recently reported developments that address this issue. Wirth *et al.* [9] use mutual information (MI) to define similar subimages in a mammogram pair before registering them using radial-basis functions. However, mutual information strongly depends on the size of the selected image window (the smaller the window the weaker the statistics) and it is difficult to overcome nonrigid changes of a region's intensity profile over time. Finally, Richard and Cohen [10] present an interesting registration method that minimizes the energy of the linearized elasticity without any boundary constraints. Here again, the method is highly dependent on the degree of preservation of the intensity profiles and tissue architecture of the mammogram pair. In our paper, we have observed that for most temporal pairs of mammograms it is possible to identify a set of corresponding "landmarks"—points or regions—that can be used to refine and improve the registration. This observation stems from the preserved "architectural similarity" of temporal pairs of mammograms. Based on this, we also developed a

IEEE Proof

novel method for defining internal landmarks in mammogram sequences.

It is important to mention that in all previous work discussed so far, there is no significant quantitative assessment reported. However, there have been two recent efforts to quantitatively validate mammogram registration and for this reason there is a special mention of them in Section IV-E where they are compared to our results.

Our mammogram registration method comprises three stages.

1) Boundary Registration through the robust detection of a set of points on the breast boundary (Section II).
2) Calculation of internal correspondences using a wavelet-based multiscale analysis (Section III).
3) Calculation of the transformation using both the boundary and the internal points in a thin-plate spline approximation scheme (Section III).

The three stages of the registration process are presented in detail in the following sections.

II. Boundary Registration

As we noted in Section I, registration of the breast boundary has been shown to be a good basis for temporal mammogram registration. It provides information about the difference in compression between two acquisitions and enables the calculation of landmarks that are the basis for the initial approximation of the nonrigid transformation. The steps that comprise the boundary registration method are as follows:

1) Boundary outline detection (Section A);
2) Curvature analysis of the outline(s) and definition of landmarks (Section B);
3) Alignment using thin-plate spline interpolation (Section C);

These steps are now explained in detail.

A. Breast Outline Detection

In order to obtain the breast outline, we first need to segment the breast boundary. Although the detection of breast area appears to be relatively straightforward (indeed, several methods such as global thresholding and histogram based methods [3] have been proposed), in practice it is considerably more complex: labels which may overlap or be close to the breast area, (scanner and dark current) noise and skinfolds due to breast compression make the segmentation more difficult.

In our work to date, we have used the boundary extraction technique described in [1]. It is based on a combination of the Hough transform (to remove wedges) followed by image gradient operators (to separate the background) and morphology in order to make coherent the breast region part of the image. Finally, to obtain an 8-connected component (which we need to be able to track points along the boundary) we apply mathematical morphology (closing $(\bullet)$ followed by dilation $(\oplus)$ and then subtraction from the original).

Although it not important in this context to give more detail, we note that our implementation of the algorithm works at pixel accuracy, so is prone to jaggedness (e.g., staircase effect). This complicates the detection of landmarks (strong curvature extrema) as is discussed in the next section.

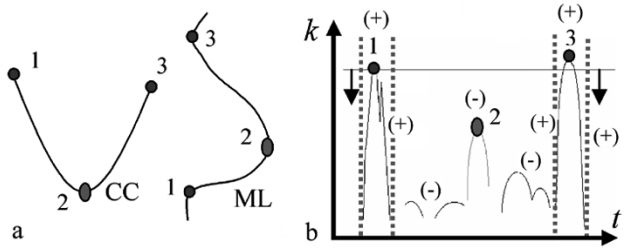


Fig. 2. (a) Consistent landmarks in the CC and ML "idealized" outlines. (b) The corresponding curvature profile showing the positive $(+)$ and negative $(-)$ segments overlaid.

B. Curvature Analysis of the Breast Outline

Fig. 2 outlines our method for detecting invariant points along the breast boundary in the "ideal" breast outlines, showing three points of characteristic curvature. In the cranio-caudal case (CC), the points 1 and 3 are assumed (as in [2]) to be near in the chest wall (and thus invariant) and are approximated by the first and last points of the breast outline, respectively. While the chest wall was always visible in our dataset, it is possible that in a particular session the breast is imaged at a distance from the chest wall; in these cases only point 2 can be computed and image warping may lead to errors. Point 2, is the maximum curvature point (negative curvature by convention). Most frequently this point correspond to the nipple; however, even if the nipple is not visible or detectable by curvature the method still computes and uses the maximum negative curvature point. In the mediolateral (ML) case, the 3 invariant points can be described as two maxima of positive (by convention) curvature (points 1 and 3) and 1 point of maximum negative curvature (point 2).

As shown in Fig. 2(a), the detected landmarks very often correspond to the anatomical location of the rib (point 1 in Fig. 2), the nipple (point 2), and the axilla (point 3).

To build a robust detection algorithm for the three points aforementioned, we need to calculate the curvature profile of the breast outline. In the remainder of this section, we will refer to the ML case only, as it is more difficult and general than the CC (providing the chest wall is visible). The breast outline may appear quite jagged and this can result in a large number of local maxima (and minima) of curvature along the breast outline. Additionally, the calculation of curvature involves the estimation of second order derivatives resulting in "noisy" curvature profiles. To overcome such problems, Asada and Brady [11], and others, have suggested a Gaussian multiscale analysis of features in 2-D curves. The multiscale method we developed is described in [12] and [14], and can be summarized in three steps.

1) First, we automatically detect the three suggested points based on an "idealized" model of the breast outline in Fig. 2(a). This is done by analysing separately the opposite signed curvature segments [shown overlaid in Fig. 2(b)], and performing 1-D detection of the maximum curvature points in the positive curvature profile (for points 1 and 3), and in the negative one (for point 2). After the detection of each maximum the signal is gradually "erased" to avoid the detection of local maxima as is shown in Fig. 2(b).

IEEE Proof

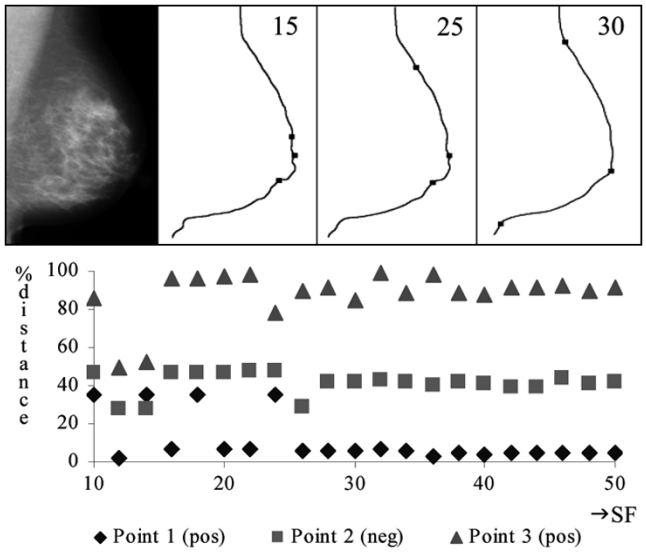


Fig. 3. The graph shows the optimum sampling S opt estimation as the sampling frequency (SF) along the breast outline where the detection of the 3 curvature maxima converges to a steady solution (S opt $= 40$ in this case). The detection of points 1, 2, 3 is shown above for $\mathrm{SF} = 15, 25$ and 30.

2) Then, we repeat the previous step at different scales (each scale being a spline approximation of the curve at a given sampling of points along the boundary S). This way, we can detect the optimum sampling rate (S_{opt}) for which the originally jagged-shaped outline approaches the "ideal" on the basis that the detection of the three boundary points becomes stable. This concept is illustrated in Fig. 3

C. Thin-Plate Spline Interpolation to Align the Boundaries

We have observed that for temporal mammogram registration a good initial alignment can be achieved using at least five points along the breast boundary. For greater accuracy we use seven points uniformly sampled between boundary landmarks 3 and 2, and another seven between landmarks 2 and 1 (total of 17 points). Using these points, the transformation that aligns the boundaries is computed using thin-plate spline interpolation [13]. Once the interpolating function has been calculated, "warped" images are produced by forcing every point (x, y) in a mammogram to take the intensity value of the point where the interpolating function maps the (x, y) point of the previous mammogram. Bilinear interpolation can be used to calculate intensities outside the pixel grid (as in [2]). Boundary registration seems to vastly improve the correspondence of a mammogram pair. Generally, it accounts well for the major differences between the images, correcting for scaling (due to differences in compression or breast size) translation and small rotations due to breast positioning and orientation of the glandular structures (since the nipple is included as a landmark). However, as described in [7], different breast compressions (which are almost inevitable), tend to make denser structures (e.g., glandular tissue, masses) move more than less dense tissues resulting in disruption to an otherwise generally smooth motion field. This is illustrated in Fig. 4 where the boundary registration does not calculate with good accuracy the location of the fibroadenoma (bright region of interest in the center of mammograms A and

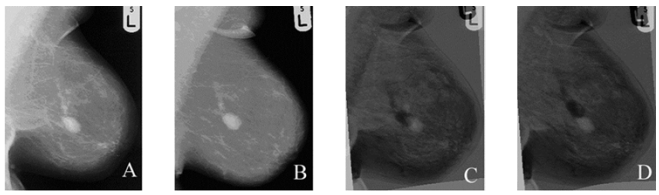


Fig. 4. Boundary registration for a mammogram pair (A), (B). The difference image after registration (C), shows that internally, registration is incomplete. By aligning the pectoral muscle in (D), the internal correspondence is reduced.

B). As we have previously reported ([4], [12], [14]) the pectoral muscle should be ignored in the registration process because of the (relatively) weak control on its placement and tension (it largely depends on the woman's arm position). This is also illustrated in Fig. 4(D) where the alignment of the pectoral muscle (by placing landmarks manually) leads to additional displacement of the mass. Nevertheless, in many cases the registration result can be improved by defining internal correspondences in the mammogram pair. This is discussed in the next section.

III. Internal Landmarks and Final Registration

To complete the registration process, we need to identify internal landmarks to compensate for the complex internal deformation of the breast caused by the compression. The most logical choice for internal landmarks is isolated regions of dense tissue (bright regions in the mammogram). These can be dense glandular tissue, masses or prominent calcifications since they all give rise to "bright" isolated regions. To improve robustness, calcifications are not considered as potential matches due to their limited spatial extend which can lead to multiple matches. In addition, we mainly consider distinct regions of dense tissue that move closer (most of the times overlap) when applying the initial, boundary-based transformation and have similar size/scale characteristics. This way, the number of internal matches is reduced (depending also on the tissue structure), but there is a corresponding increase in the robustness of the method.

The mammogram pair is analyzed using a nonlinear wavelet scale-space to isolate significant regions of interest. The algorithm used for segmenting and matching internal structures in mammograms is described in [14]. Here, we summarize the basic steps of the scale-space analysis.

1) The mammogram pair is decomposed using the Coiflet wavelet packets. This particular wavelet was chosen because it yields good frequency and spatial localization (e.g., it is edge preserving) and has compact support.
2) After each mammogram is decomposed into a set of high- and low-frequency images (with good spatial localization of features), these are ranked using an information cost function in the context of a "best basis" algorithm [15]. For this, an entropy measure is used and each wavelet subspace (filter superposition) is then cumulatively adjointly convolved, in order, with respect to the best-basis assessment of the decomposition. As shown in Fig. 5, the result is a "stack" of reconstructions from minimum to maximum information content. The top of the stack contains the fully reconstructed image while coarser features

IEEE Proof

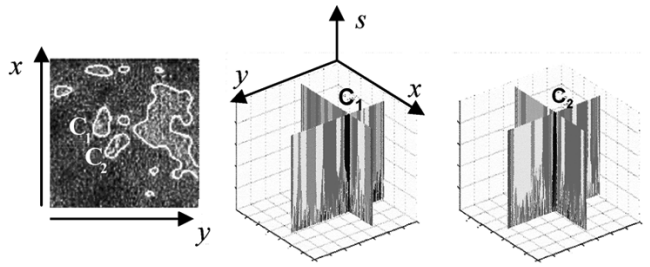


Fig. 5. Scale-space analysis performed to detect a pair of calcifications in a small section of a 50 μm mammogram

appear toward the bottom. This construction is used to track significant features through scale space and forms the basis of the feature segmentation.

3) Based on the constructed "stack," important features are detected, namely those that persist through the scale-space. The scalogram method used, has been designed for multimodality registration (X-ray/MRI) and it is described in detail in [14]. First, region growing is applied from the lowest scale toward the highest. A merge operator tracks the feature to the highest scale so that each feature can be represented with more "detailed" information. The region growing starts with a collection of seed pixels at the low-scale part of the stack. A merging operator is applied to adjacent pixels $l_{i+1}, l_{i-1}, l_{j+1}, l_{j-1}$, and l_{k-1}, l_{k+1} in the scale-stack volume $V(x, y, \lambda)$ using

$$\Delta E = \frac{L_i L_j L_k}{L_i + L_j + L_k} \times \left\| \frac{\sum_{n\in X} V(n)}{L_i} - \frac{\sum_{n\in Y} V(n)}{L_j} - \frac{\sum_{n\in \Lambda} V(n)}{L_k} \right\|^2 - \Gamma \quad (1)$$

and

$$\Delta E \begin{cases} <0, l_i, l_j, l_k & \text{merge} \\ >=0, l_i, l_j, l_k & \text{unchanged} \end{cases} \quad (2)$$

where

$$L_i = \sum_{n\in l_i} L(n), \quad L_j = \sum_{n\in l_j} L(n), \quad L_k = \sum_{n\in l_k} L(n). \quad (3)$$

$L(n) = 1$ and L_i gives the vector of intensities of l_i. This process is iteratively repeated for a threshold value Γ until no more merges are possible (i.e., the region is completely grown). Γ is gradually increased until either the desired number of regions is produced (the m most significant regions) or the number of regions converges. Once the regions have been segmented they are fitted with a cubic spline contour and the image labeled into generalized areas of "region of interest" and "noise." As is explained in [14], region competition is also used to "fine tune" the segmented regions.

In Fig. 5, calcifications are detected as a well-localized propagation through the sliced scale-space (C_1 and C_2). The scalogram's intensities reflect relative saliency of the grown region from the wavelet packet descriptor set. In our experimentation, we have found that it is more robust to use fewer good quality landmarks, therefore, this number m can either be controlled interactively by the user (typically $m < 5$) or the landmark selection criteria (described later in this section) used to reject features that have less significant scale localization.

To match internal regions in each sequence, for each region in the first image we define a "search" window in the second, whose size is defined by the displacement of the centroid when applying the boundary deformation function. Due to the small number of segmented significant regions, matching becomes a relatively easy task. If more than one potential match is present in this window, we define a "match rejection filter" based on the size and orientation of the region, scale information, and relative motion of each potential match when applying the boundary based registration (Section II). This way, we ensure that spatially localized features are not matched to larger ones. In addition, if the distance between a candidate pair is reduced after boundary registration, the correspondence is preferred to another for which the distance increases. A small number of landmarks (depending on the composition of the breast as is described in the results section) is then defined as the centroids of the matched regions. In most cases, the segmented salient regions in the partially warped and the target image move toward each other and exhibit a significant degree of overlap facilitating the matching process. As aforementioned, the pectoral muscle is ignored, even if it is a scale persistent feature, since it introduces an often different motion field analogous to the temporal variation of its tension. In addition, although the method can detect microcalcifications we only use features that have a major axis larger than 10 mm to minimize the matching search space and increase robustness.

The boundary points and the internal landmarks (computed by the wavelet analysis) together control a thin-plate spline approximation technique that allows smoothness control and the *individual weighting* for each landmark. For example, a pair of internal landmarks that correspond to "large areas" of tissue should play a more important role than those that correspond to a smaller one. To implement this, it is necessary to weaken the interpolation condition in such a way that the resulting transformation is a compromise between smoothness and data adaptation. As described in [16], the new functional $J_\lambda(\mathbf{t})$ to be minimized is a function of λ, a regularization parameters that controls the balance between the smoothness and the approximation of the transformation

$$J_\lambda(\mathrm{t}) = \sum_{i=1}^{n} \frac{|\mathrm{q_i} - -\mathrm{p_i}|^2}{\sigma_i^2} + \lambda \cdot J_2^2(\mathrm{t}) \quad (4)$$

where $\mathbf{q_i}$ and $\mathbf{p_i}$ are the landmark pairs, σ_i is the individual weight for each landmark pair i, $J_2(\mathbf{t})$ is the bending energy of the transformation $\mathbf{t}$ and λ is a parameter that controls the tradeoff between the smoothness of the transformation and adaptation to the local transformations induced by the data. λ may vary between 0 (very good approximation of the data) to ∞ (very smooth transformation with very little adaptation to local deformations). This is illustrated in Fig. 6, where λ increases from 0 to 10^4, resulting in a smooth—but not close to the data-transformation. In our paper, λ was set to 1 since we individually weighted each landmark pair. After matching the

IEEE Proof

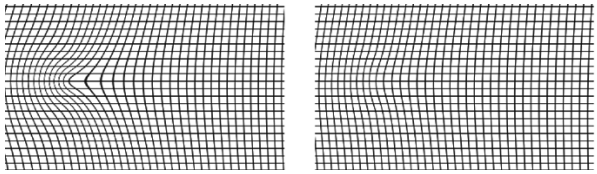


Fig. 6. Increasing λ from 0 (left) to 10^4 (right) leads to a smooth transformation that does not conform any more to the data. The landmarks used are the same in both cases.

segmented structures, the "goodness" of each match is characterized by its size and persistence of the regions in the wavelet stack (i.e., the "volume" of the region *through* the scale-space stack). These scores are used (rescaled between 1 and 100 and the higher value is also assigned to the boundary landmarks) as the landmark weights (the inverse of σ^2) in the approximation scheme. This way, greater deformation influence can be attributed to features that have stronger spatial localization. It is worth mentioning that the registration results are stable to small perturbations in the value of λ and change significantly only when it is increased by an order of magnitude.

IV. Results and Validation

The difficulty in validating a mammogram registration method is that after registration the image pair is not expected to be identical due to frequent changes in breast composition. For this reason, the success of the registration technique has to be assessed *both* with quantitative measures and by the judgement of an expert clinician. In this section, we present registration results and validation including discussions concerning the use of "artificial data" for validation as well as the need to include intensity normalization in the process.

A. Artificial Data

Most researchers use "artificial data" to evaluate complex registration problems such as temporal mammography. An example is illustrated in Fig. 7, in which we apply a known nonrigid transform (taken from another mammogram pair so that it reflects a *realistic* breast deformation) to Fig. 7(a) thus obtaining the deformed mammogram shown in Fig. 7(b). By applying our registration technique the original mammogram is recovered from (b), as is shown in Fig. 7(c) (warped image) and Fig. 7(d) (difference image after registration). This kind of experiment offers a basis for evaluation (the images have to be identical). Although the registration is satisfactory, in such an experiment we cannot mimic the actual anatomical changes that occur in the breast during acquisitions and are amplified by differences in compression and imaging parameters. Furthermore, in reality the breast structure may change significantly over time and register pairs are not generally expected to be identical! For this reason we use *real temporal data* in order to validate the proposed registration technique, even if this makes establishing a ground truth a more difficult task.

B. Description of the Data-Set

Our main dataset consisted of 50 "normal" temporal pairs of mammograms taken from the Royal Free Hospital, London. It should be noted that the data was not taken from a screening

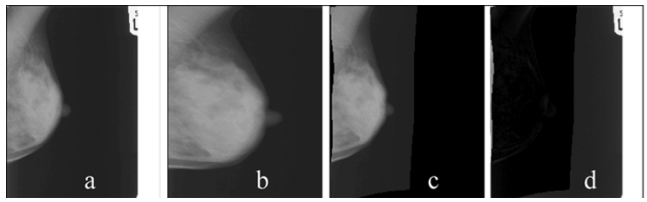


Fig. 7. An artificial experiment for evaluating registration. (a) The original mammogram. (b) A random mammogram transform is applied and the mammogram is significantly deformed nonrigidly. (c) Using the mammogram registration technique the mammogram is aligned to its original shape. (d) The difference image shows that the images are almost identical, however such artificial experiments do not model well the actual mammographic changes encountered in real temporal data.

program and, therefore, time intervals were long (average interval 30 mo, range: 12–48 mo). The mammograms were digitized using a Lumisys scanner into 50-μm 12-bit images and then downsampled to 200 μm (as in [17]). In 10 of these pairs, the breast consisted mostly of fatty tissue with no ducts visible or with prominent ducts in the anterior portion. In 25 cases, the breast was involuted with prominent ductal pattern of moderate to severe degree (usually scattered densities), occupying more than one-fourth of the breast volume. In the remaining cases, the breast parenchyma was very dense and prominent ductal patterns were not visible. Densities were selected as control points for validation (Section IV-D).

In addition, for the quantitative assessment in Section IV-D, 16 temporal mammograms from postmenopausal women with 17 consistent, annotated masses were also used for validation. The average interval here was 12 months (range: 10–13) and the tissue composition was fatty. These images were not used to assess the method for cancer detection (an initial report proposing a framework for reducing the large number of false positives is reported in [18]).

C. Clinical Assessment of the Results

The correspondence based on visual perception was "good" (very few misregistration "shadows") in 70% of the cases and "average" (regions of poor overlap in the difference image resulting in "shadows"), in 25%. The "poor" correspondences (important features such as the nipple were not aligned well), represented the 5% of the cases and mainly corresponded to fully involuted mammogram pairs, without prominent, "bright" regions while in most of the "very dense pairs" registration was characterized as average. In the next section, we suggest a quantitative validation method based on the "preserved" tissue architecture.

D. Quantitative Assessment of Registration

Fig. 8 shows the normalized (via division with the sum of the image entropies), mutual information measure $(\mathrm{MI}(\mathrm{A},\mathrm{B})/[\mathrm{H}(\mathrm{A})+\mathrm{H}(\mathrm{B})])$, calculated for the main database of 50 mammogram pairs before (nMI Unregistered curve in Fig. 8), after boundary registration (nMI Boundary curve in Fig. 8) and after the final inclusion of internal landmarks (nMI Registered curve in Fig. 8). It should be noted that MI as a second order statistical similarity measure, is largely dependent on the preservation of the intensity profile and is used here

IEEE Proof

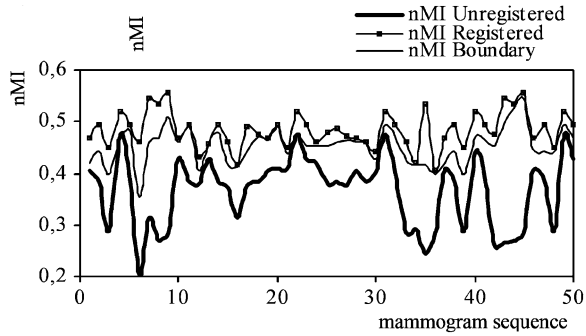


Fig. 8. Normalized MI is calculated for the 50 mammogram pairs of the main database. "nMI Unregistered" is the MI curve before registration, "nMI Boundary" is the MI curve after boundary-only registration and "nMI Registered" is the MI curve after the final registration.

only to visualize the "general trend" of improvement which however, is not obvious when comparing boundary and internal landmark registration in Fig. 8. It should also be mentioned that the pectoral muscle had to be completely removed to obtain this result since in some cases due to image warping the overlap of the muscle was poor, resulting in a random fluctuation of the statistical measure used.

Since fiducials cannot be used in the breast, (except maybe the possibility to use skin markers for assessing the detection of the nipple), we selected 58 landmarks from 20 sequences of our main dataset where the architectural structure was preserved, so that reliable correspondences could be used as "ground truth," according to the clinician. It should be noted that in very dense or involuted pairs it is difficult to define reliable controls. In each pair, three "obvious" landmarks were selected on average (depending on the similarity), as shown in Fig. 9(a) resulting in 58 "ground truth" landmarks. These landmarks were most often characteristic regions of dense tissue or benign masses. In addition, for the main database of 50 mammograms, the selected landmarks had to be significantly different (>0.4 cm) from the landmarks used for the registration in order to assess the *global* improvement in correspondence. Also, a different dataset with 17 annotated masses in 16 sequences was also used. The center of mass of corresponding annotations was used to define 17 additional controls [last 17 points in Fig. 9(b)]. Fig. 9(a) illustrates a "ground truth" landmark pair $(\mathrm{P_S}, \mathrm{P_T})$ in the source and target image. By applying first the boundary transformation and then the final one (if there are internal landmarks), the source point $\mathrm{P_S}$ is transformed to $\mathrm{T_B}(\mathrm{P_S})$ and $\mathrm{T_I}(\mathrm{P_S})$, respectively. By comparing the initial distance error, $(\mathrm{e} = \mathrm{d}(\mathrm{P_S}, \mathrm{P_T})$ to the boundary registered one $(\mathrm{eB} = \mathrm{d}(\mathrm{T_B}((\mathrm{P_S}), \mathrm{P_T})$ and to the final registration one $(\mathrm{eI} = \mathrm{d}(\mathrm{T}_I(\mathrm{P_S}), \mathrm{P_T})$ the improvement of landmark correspondence is assessed for the 75 "ground truth" landmarks of our experiment. The error curves e, eB, and eI are shown in Fig. 9(b).

E. Comparison of Results With Previous Work

The average and standard deviation for the error curves in Fig. 9(b) are: 1.5 and 1.3 cm before registration, 4, 7, and 3, 1 mm after boundary based registration and 3 and 2, 7 mm after the registration with the internal landmarks. It should be noted, that without the inclusion of the annotated database (that was

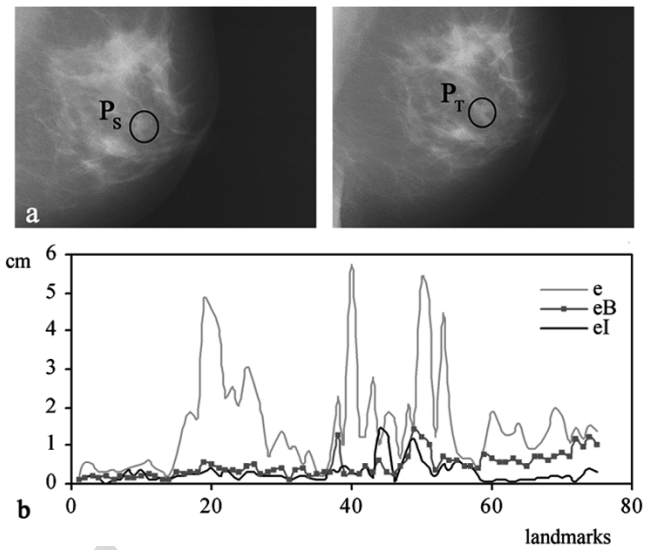


Fig. 9. (a) Selected landmarks $(\mathrm{P_S}, \mathrm{P_T})$ for assessing the accuracy of the method. (b) e is the initial error, eB is the error after boundary registration, and eI the error after including internal landmarks.

mostly comprised of fatty tissue mammograms with a "bright" mass present), the average errors after partial and complete registration were 4 and 3, 5 mm. This indicates that the final step of registration is beneficial mainly in "fatty tissue" mammograms with clear masses present. In addition, in all cases there was partial overlap of the corresponding regions after registration. Since these results are based in quantitative assessment of controls, they can be compared with the ones reported in [17] and [19].

In [17], four methods are compared and the best results (average error 7.9 mm using densities as controls) are obtained by optimizing an affine transformation based on mutual information maximization. In [19], a quantitative assessment based only on lesions shows average error 8.4 mm reducible to 4.2 mm when additional local affine matches are used. It should be noted that although the results from the tree methods are comparable, the variability in the used datasets can be very significant. This is also illustrated in Fig. 9(b), where it is obvious that the improvement in registration is largely dependent on the initial displacement. Comparing to [17], the average initial errors in this paper where comparable while in the main dataset the average interval was 30 months, comparing with 23 months in [17]. We argue that the initial displacement (that is largely dependent on the quality of the temporal acquisition), the interval time between sessions and the mammographic tissue type (not reported in [17] and [19]) of the data, are some important parameters that influence the registration result. All these parameters should be taken into consideration for cross-study result comparison of mammogram registration techniques while a common dataset should be defined as a benchmark for validating such heterogeneous methods.

F. Combining Registration and Normalization

In this final experiment, we address the issue of intensity correction in mammogram sequences. The geometrical alignment method presented in this paper does not account for dis-

IEEE Proof

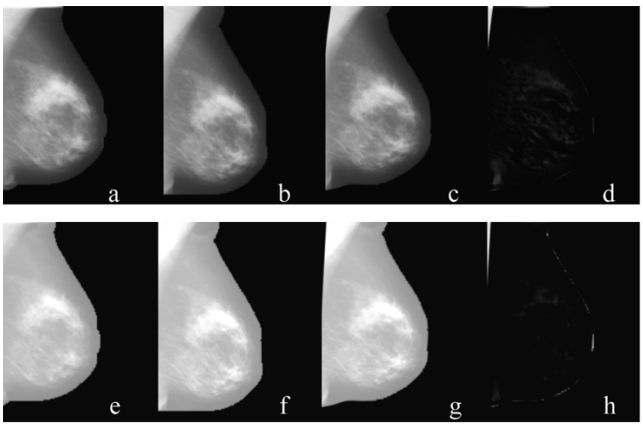


Fig. 10. (a)-(d) The original mammogram registration process (a,b is the mammogram pair, c is b registered in a and d is the difference image). (e)-(h) The same registration process for the normalized images. Note that the initial geometry and the transformation is exactly the same in both cases.

tortions in the mammogram intensities due to imaging conditions. In fact, variations in imaging parameters often result in a nonrigid transform in the mammogram intensity between successive mammograms of the same person. In this section we use a small subset of our data to assess the improvement in the registration result if the SMFs (based on the $h_{\rm int}$ representation [1]) of the data set is used instead of the original mammograms. The reason for not using all our data was the lack of image acquisition information (mainly, height of compression) that is needed for calculating $h_{\rm int}$ [1]. For this reason the results are indicative and further validation should be performed. The hypothesis of this experiment is that the difference image of a temporal mammogram pair after registration will be closer to zero ("zero" is the ideal result) in the SMF pairs.

Fig. 10 shows an example that is in agreement to this hypothesis; a temporal mammogram pair is registered using our algorithm (10.a–10.d). If the same geometrical transformation is applied in the normalized mammograms (10.e-10.f) the registration result (difference image) is improved (misregistration shadows disappear). It is important to note that the $h_{\rm int}$ representation calculates the height of nonfatty tissue that corresponds to each pixel and it has been shown to preserve and most often increase the contrast [1]. In addition, it is noted that the presence of the pectoral muscle can strongly influence statistical measures. However, in this experiment the geometry is exactly the same for both SMF and original data pairs since we are interested in the *differences* in statistics and not in their absolute values. This way, the mean and standard deviation of the difference image were calculated for both representations and the results are shown in Fig. 11 (mean is the center and the SDEV the diameter of each circle). It is clear that in all cases the results are improved as the mean and SD in the difference images are significantly reduced. The results indicate that the residual error after registration in the normalized pairs is a better approximation of the *actual* changes in breast tissue composition over time. Such a method could potentially reduce false positives. However, this hypothesis requires extensive experimentation.

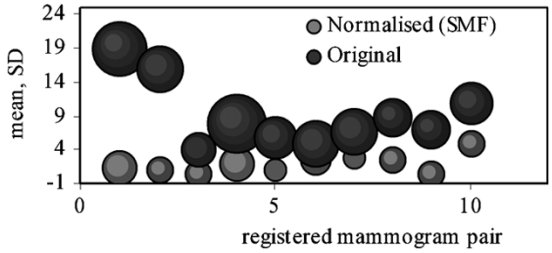


Fig. 11. Intensity mean and SD for the difference images of 10 mammogram pairs before and after normalization.

V. Discussion

In this paper, a registration method for aligning mammogram sequences was presented. The method first aligns the breast boundaries by calculating salient curvature points that conform to the "idealized" model (shown in Fig. 2(a) This algorithm (described in Section II-B) has also the potential to be the basis of a CAD system for the automatic detection of the "nipple," "axilla," and "rib" anatomical locations. In the 50 mammogram pairs used it was confirmed that the "rib" coincided with "point 1" (see Fig. 2) in 85% of the cases, the "nipple" with "point 2" in 75% of the cases and the "axilla" with "point 3" in 80% of the cases. In the remaining of the cases the nipple was not visible; however, the curvature method detected the maximum negative point. Our method does not locate the nipple when there is no curvature maximum near its location. More specialized algorithms for its detection can improve the accuracy of the method. It is also important to mention that in a clinical system implementation of the presented method, the user must be able to confirm the "correctness" of the calculated landmarks to ensure that the transformation will be computed correctly.

The second part of the presented work strongly depends on the temporal preservation of tissue architecture; a small number of robust internal landmarks is defined using the wavelet scale-space analysis presented in Section III. The inclusion of an intensity normalization scheme can potentially improve the registration result as was demonstrated in Section IV.F.

The method has the potential to assist radiologists in comparing mammograms either on a breast screening scenario or in therapy monitoring (we have reported relevant work for HRT users in [20]) but extensive validation is needed to confirm this.

One of the basic problems for applying the method for the automatic detection of cancers (i.e., prompting), is the large number of false positives. This is illustrated in Fig. 12, where a temporal mammogram pair is registered (C is B registered to A). A small lesion (arrow L) can be detected from the difference image [Fig. 12(D)]. While the architectural similarity is more pronounced when looking at the registered mammogram sequence (A and C) the difference image also generates false positives due to the actual temporal changes that take place as well as the registration error. For this reason the presented method should be primarily used as a tool to improve temporal correspondence, thus aiding the clinician to track and compare specific regions in previous mammograms. In [18], initial results in 15 temporal sequences indicate that a significant resection (up

IEEE Proof

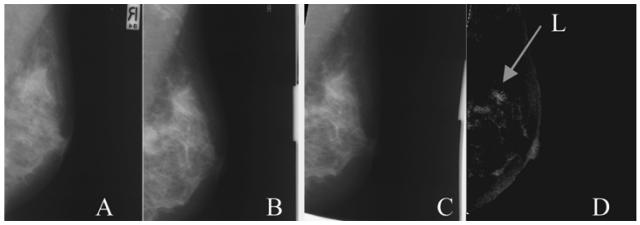


Fig. 12. Temporal detection of changes using registration. B is aligned with A (resulting in C), while the difference image D conveys information regarding regions of normal or abnormal change.

to 90%) of false positives can be eliminated using more sophisticated post-processing techniques. We hope that this work will further inspire research in this area.

REFERENCES

[1] R. P. Highnam and J. M. Brady, *Mammographic Image Analysis*. Norwell, MA: Kluwer Academic, 1999.

[2] M. Sallam and K. Bowyer, "Registration and difference analysis of corresponding mammogram images," *Med. Image Anal.*, vol. 1, no. 1, pp. 73–91, 1996.

[3] N. Karssemeijer and G. T. Brake, "Combining single view features and asymmetry for detection of mass lesions," in *Proc. IWDM*, 1998, pp. 95–102.

[4] K. Marias, C. P. Behrenbruch, M. Brady, S. Parbhoo, and A. Seifalian, "Multi-scale landmark selection for improved registration of temporal mammograms," in *Proc. IWDM*, 2000, pp. 580–586. Med. Phys..

[5] N. Vujovic and D. Brzakovic, "Establishing the correspondence between control points in pairs of mammographic images," *IEEE Trans. Med. Imag.*, vol. 6, no. 10, pp. 1388–1398, Oct. 1997.

[6] R. Marti, R. Zwiggelaar, and C. Rubin, "Automatic registration of mammograms based on linear structures," *Inf. Process. Med. Imag.*, pp. 162–168, 2001.

[7] S. L. Kok-Wiles, M. Brady, and R. Hignam, "Comparing mammogram pairs for the detection of lesions," in *Proc. IWDM 98*, 1998, pp. 103–110.

[8] B. Hong and M. Brady, "A topographic representation for mammogram segmentation," in *Proc. MICCAI*, 2003, pp. 730–737.

[9] M. A. Wirth, J. Narhan, and D. Gray, "Mammogram registration using mutual information," *SPIE Medical Imaging: Image Process.*, vol. 4684, pp. 562–573, 2002.

[10] F. Richard and L. Cohen, "A new image registration technique with free boundary constraints: Application to mammography," in *Proc. ECCV*, 2002, pp. 531–545.

[11] H. Asada and M. Brady, "The curvature primal sketch," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, pp. 2–14, 1986.

[12] K. Marias, "Registration and quantitative comparison of temporal mammogram sequences with application to local tissue changes quantification hormone replacement therapy (HRT) patients," Ph.D. dissertation, Univ. College London, Univ. London, London, U.K., 2002.

[13] F. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.

[14] C. P. Behrenbruch, K. Marias, P. A. Armitage, M. Yam, N. Moore, R. E. English, P. J. Clarke, and M. Brady, "Fusion of contrast-enhanced breast MR and mammographic imaging data," in *Med. Image Anal. (MedIA)*. New York: Elsevier, Sept. 2003, vol. 7, pp. 311–340.

[15] R. R. Coifman, Y. Meyer, S. R. Quake, and M. V. Wickerhauser, "Signal processing and compression with wavelet packets," in *Proc. Int. Conf. Wavelets*, 1992, pp. 77–93.

[16] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. H. Kuhn, "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Trans. Med. Imag.*, vol. 20, no. 6, pp. 526–534, Jun. 2001.

[17] S. van Engeland, P. R. Snoeren, J. Hendriks, and N. Karssemeijer, "A comparison of methods for mammogram registration. A comparison of methods for mammogram registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 11, pp. 1436–1444, Nov. 2003.

[18] M. G. Linguraru, K. Marias, and J. M. Brady, "Temporal mass detection," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2002, Proc. IWDM, pp. 347–350.

[19] L. Hadjiiski, H. Chan, B. Sahiner, N. Petrick, and M. A. Helvie, "Automated registration of breast lesions in temporal pairs of mammograms for interval change analysis—Local affine transformation for improved localization," *Med Phys.*, vol. 28, no. 6, pp. 1070–1079, Jun. 2001.

[20] K. Marias, C. P. Behrenbruch, R. P. Highnam, S. Parbhoo, A. Seifalian, and M. Brady, "A mammographic image analysis method to detect and measure changes in breast density," *Eur. J. Radiol.*, vol. 52, no. 3, pp. 276–282, Dec. 2004.

IEEE Proof