DEPARTMENT OF COMPUTER SCIENCE (DIKU)
UNIVERSITY OF COPENHAGEN

Workshop on Breast Image Analysis

In conjunction with
the 16th International Conference on
Medical Image Computing and Computer
Assisted Intervention (MICCAI 2013)

Anne Martel, John Hipwell, Julia Schnabel,
Mads Nielsen, Martyn Nash,
Despina Kontos, Nico Karssemeijer (Eds.)

Technical Report no. 01/2013

ISSN: 0107-8283

ISBN: 978-87-996443-0-8

Preface

In the developed world more than 10% of women are diagnosed with breast cancer during their lifetime and between 1980 and 2010 global breast cancer incidence increased at an annual rate of 3.1%. It is therefore important that improved techniques to detect, treat and manage breast cancer are developed and imaging plays a crucial role in this endeavor.

Well established imaging modalities used in detection, such as mammography, 2D handheld ultrasound, dynamic contrast-enhanced MRI, are now complemented by newer technologies such as whole breast 3D ultrasound and breast tomosynthesis. As cancers are being detected at an earlier stage there is an increased need for breast conserving surgery and image guidance is essential to improve the accuracy of surgical procedures. The role of functional imaging and digital pathology is also becoming more important in guiding treatment options after surgery. To make optimal use of these technologies, there is a strong need for the development of effective image analysis methods. Examples are extraction and combination of information from multimodal imaging, the use of prior images in screening, and relating image characteristics to risk factors. Development of such methods is highly challenging due to large deformations of the breast and subtlety of the abnormalities to be detected and diagnosed.

The second MICCAI workshop on Breast Image Analysis provided an international forum for the presentation and discussion of new advances that address these problems. It was held on 26th of September 2013 in Nagoya, Japan, in conjunction with the 16th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2013). Full-length paper submissions (8 pages, LNCS style) were invited and a total of 19 submissions from 6 continents and 11 countries were received. Each submitted paper received double-blinded reviews from 3 of the 20 members of the Scientific Committee and the Organising Committee, all experts in the field. The organizing committee carried out the selection of papers based on the reviews and a total of 16 papers were accepted and are included in these workshop proceedings. From these, 11 papers were selected for oral presentations and 5 papers were selected for poster presentation. The main themes were: Deformation and Registration; Digital Mammography; and Segmentation and Classification. Authors were asked to take reviewers' comments into account for their final version. Proceedings were made available as a PDF file on the MICCAI Workshop USB stick, as well as ISBN Online proceedings. The purpose of this Workshop was to bring together

researchers in the field of breast imaging to discuss and exchange new ideas and applications, and to provide a platform for breast image analysis methodology across all imaging modalities. The invited keynote lecture was given by Professor Tsuyoshi Shiina, from the Graduate School of Medicine, Human Health Sciences, Kyoto University, on ultrasound tissue elasticity imaging in breast cancer and was very enlightening and well received.

On a more sombre note, we would like to pay our respects to Prof. Dr. Bernd Fischer, Institute of Mathematics and Image Computing, University of Lübeck. In 2010, Bernd founded and headed the Fraunhofer Project Group Image Registration. He successfully introduced abstract mathematical methods into the medical imaging domain and in doing so made an enormous contribution to medical image analysis in general. Many of the groups presenting at this workshop have been greatly influenced by his work on image registration. In addition to his scientific contributions, Bernd was a very warm and generous individual and his cheerful presence will be sorely missed by all those who were fortunate to know him personally.

On behalf of the Organising Committee, we would like to thank all members of the Scientific Review Committee, workshop presenters and contributing authors, as well as all workshop participants for making this second Workshop on Breast Image Analysis a stimulating and successful event!

Organizing Committee:

Anne Martel	Sunnybrook Research Institute, Medical Biophysics, University of Toronto, Canada
John Hipwell	Centre for Medical Image Computing, University College London, UK
Julia Schnabel	Institute of Biomedical Engineering, University of Oxford, UK
Mads Nielsen	Department of Computer Science, University of Copenhagen, Denmark
Martyn Nash	Department of Engineering Science, University of Auckland, New Zealand
Despina Kontos	Computational Breast Imaging Group, Department of Radiology, University of Pennsylvania, Philadelphia, USA
Nico Karssemeijer	Department of Radiology Radboud University Nijmegen Medical Centre, The Netherlands

Scientific Committee:

Spencer Bowen	Massachusetts General Hospital, US
Mike Brady	University of Oxford, UK
Karen Drukker	University of Chicago, US
Hiroshi Fujita	Gifu University, Japan
Maryellen Giger	University of Chicago, US
Markus Harz	Fraunhofer Mevis, Germany
Brad Keller	University of Pennsylvania, US
Robert Marti	University of Girona, Spain
Thomy Mertzanidou	University College London, UK
Alison Noble	University of Oxford, UK
Bram Platel	Fraunhofer MEVIS, Germany
Nicole Ruiter	Karlsruhe Institute of Technology, Germany
Chris Tromans	Matakina Technology Ltd

Online proceedings: http://www.diku.dk/BIA2013proceedings.pdf
ISSN: 0107-8283
ISBN: 978-87-996443-0-8

Table of Contents

Temporal and ipsilateral X-ray mammography registration via a 3D patient-specific model

Thomy Mertzanidou[1], John Hipwell[1], Lianghao Han[1], Henkjan Huisman[2], Ulrich Bick[3], Nico Karssemeijer[2], David Hawkes[1]

[1] Centre for Medical Image Computing, University College London, Gower Street, WC1E 6BT London, UK
[2] Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, P.O. Box 9102, 6500 HC Nijmegen, The Netherlands
[3] Department of Radiology, Charite Universitätsmedizin Berlin, 10117 Berlin, Germany
t.mertzanidou@cs.ucl.ac.uk

Abstract. Radiologists routinely identify corresponding features in mammograms that were acquired at different times, to identify changes in breast tissue. Moreover, they often correlate findings between the CC and MLO views of a patient, as the superimposed structures can cause ambiguities regarding the presence and location of a tumour or microcalcifications. Both tasks are challenging due to the projective nature of X-ray mammography and the large deformation that the breast undergoes, which varies between different acquisitions. Therefore an automated method would be valuable for radiologists. For this task, we propose the use of a patient-specific 3D model of the breast that is generated from MRI, when available. The Finite Element (FE) model is used to iteratively deform the MR image of the patient, using a biomechanical simulation, and align it with the two mammograms of the patient via an intensity-based registration framework. The findings in each mammogram can then be matched to the other, via the 3D volume. The framework was tested on six clinical cases (12 registrations) with expert annotations for the temporal registration, giving a mean registration error of 3.5mm. The CC to MLO registration was tested using ten clinical cases and provided a mean error of 8.1mm.

1 Introduction

Determining correspondences between temporal mammograms facilitates the assessment of change in the breast tissue over time and can help radiologists in cancer detection and the assessment of tumour response to treatment. Ipsilateral registration of mammograms has also great clinical importance, as two views are typically acquired: Cranio-Caudal (CC) and Medio-Lateral Oblique (MLO). Relating regions between them is needed for the interpretation of findings, as a single mammographic view provides a 2D projection image of a 3D volume that can obscure regions of interest. In addition to image assessment, both registration

ISBN: 978-87-996443-0-8

tasks can be valuable as components in Computer-Aided Detection algorithms, to reduce the search space and automatically match extracted features.

Finding corresponding regions in temporal mammograms is a difficult task, due to the highly deformable nature of the breast tissue and the differences in positioning before compression. Previously proposed methods primarily use 2D transformations, distance metrics or extracted features. As there are no clear anatomical correspondences in an X-ray mammogram, feature-based registration can be problematic and thus 2D intensity-based techniques have been proposed instead. Between these, previous studies concluded that a 2D affine [1, 2] and a multi-resolution B-splines transformation (Free Form Deformations - FFD) technique [3] perform best.

For CC to MLO view registration, distance and texture measures have been commonly used [4–7]. Kita et al. [8] proposed a technique that reconstructs a simplified breast model from the two views and used the extracted curved epipolar lines to match findings between them.

A 3D breast model has been used in the literature to compute a transformation between two 2D projection images for different applications, such as the validation of mammogram registration techniques [9], and the matching between ipsilateral tomosynthesis views [10].

MRI is commonly used as a complementary modality to mammography to investigate symptomatic patients and high risk younger women. When an MRI of the patient is available, we propose the use of the pre-contrast image to build an FE model, that can subsequently be used to simulate the physical compression of the breast that occurs in a mammographic acquisition. By incorporating this into an MRI to X-ray registration algorithm, and registering the MRI to a pair of temporal mammograms, or both CC and MLO views, correspondence between the mammograms can be computed by combining the resulting deformation fields. Various FEM-based MRI to X-ray mammography registration algorithms have been proposed [11–13] but to our knowledge none have been applied to the problem of registering pairs of either temporal or ipsilateral X-ray mammograms. In this paper we use an iterative, solely intensity-based technique [13] to solve this problem and demonstrate that correspondence can be achieved to within a few millimetres, if an MRI is available.

2 Methodology

2.1 Temporal and CC to MLO registration framework

Our methodology for temporal registration is illustrated in Figure 1. Initially, we use the X-ray mammogram of a patient that is acquired at time-point t_1 and extract the position of interest, for example the centre of mass of a detected lesion, shown in green in step (1). This is back-projected through the deformed MR volume using the FEM registration between the MRI and the CC view mammogram. This MR locus is illustrated in red in step (2) and is a straight line, corresponding to the ray projected onto the lesion. Using the displacement

ISBN: 978-87-996443-0-8

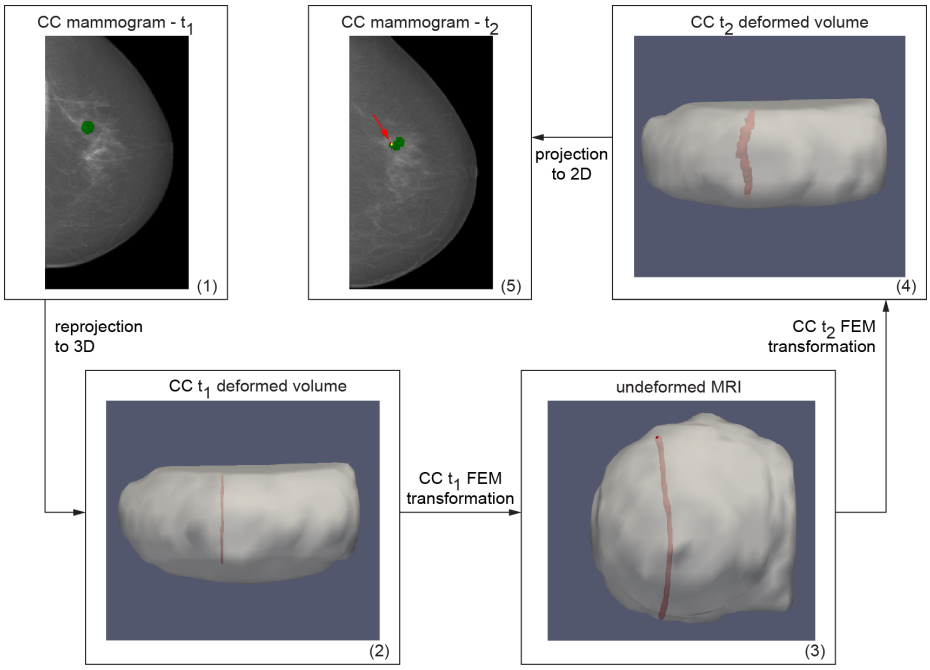


Fig. 1: The process of relating findings between the temporal CC view mammograms of a patient, through the MRI. For displaying purposes, the corridors have been dilated, using a binary structuring element, and subsequently smoothed using Gaussian smoothing, in order to produce a single connected mesh.

field of the CC FEM registration, we map this ray to the undeformed MRI of the patient, shown in step (3). At this stage we know which voxels of the original MRI correspond to the lesion identified in the mammogram acquired at t_1. We then deform the MR volume and the region of interest using the FEM registration result between the MRI and the CC mammogram at time-point t_2, illustrated in step (4). Finally this deformed corridor of interest is projected onto the patient's X-ray mammogram at t_2. Step (5) shows this mammogram overlaid with the projected area in green and in this particular clinical case a clip inserted after biopsy in red. Their overlap is yellow. As expected, the correspondence between the temporal mammograms of the patient is not one-to-one, but one-to-many. This is also clear in the results shown in section 3.

Similarly, we can use this framework for the alignment of CC to MLO view mammograms. For this application, steps (4) and (5) in Figure 1 are different. To produce the image in step (4), we use the transformation result of the MLO FEM registration of the mammogram also acquired at time-point t_1. This transformation incorporates a rotation of the breast of approximately 45°, to account for the different positioning of the X-ray source and the detector. Finally, in step (5) we obtain the MLO view of the patient's X-ray mammogram overlaid with the projection of the lesion position from the CC view, in step (1).

ISBN: 978-87-996443-0-8

2.2 FEM-based MRI to X-ray registration

The key component of our mapping process between different X-rays is the FEM-based registration between the MRI and the X-ray mammograms. We use an intensity-based approach previously proposed for this task [13], as it has the advantage of using the full intensity information provided by the images, rather than extracted features, and also follows an iterative optimisation scheme, whereby both pose and biomechanical model parameters are updated simultaneously. For completeness, we summarise the pipeline below. Further implementation details can be found in [13].

Breast models are generated using the MRI of the patient. The breast volume is segmented from the image and a volumetric mesh is extracted. We are using a nearly incompressible, anisotropic and hyperelastic neo-Hookean model for modelling, that is transversely isotropic, to account for the reinforcement of biomechanical properties from fibre-like connective tissues in a preferred direction. The plate compression is simulated with a frictionless contact model and the position of the pectoral muscle nodes is approximated to be planar, with movement of the nodes constrained to lie within that plane.

There are 7 iteratively updated transformation parameters. Four account for the rigid position of the breast before compression: the translation vector in the plane perpendicular to the source-detector normal, and two rotations, one for rolling of the breast about the anterior-posterior axis and one for in-plane rotation of the breast about the superior-inferior axis. The modelling and simulation parameters that are optimised are the distance between the two compression plates (amount of compression), the ratio of tissue enhancement coefficient (anisotropy) in the preferred direction and Poisson's ratio. We consider the breast tissue to be homogeneous with a Young's modulus of $4kPa$.

Before registration, the MRI intensities are transformed to X-ray attenuation, using the method described in [14], to produce a volume that when transformed and projected into 2D, using a perspective ray-casting projection, gives a simulated mammogram. The similarity measure used is normalized cross correlation. At each iteration of the hill climbing optimisation scheme, one parameter is updated, that which results in the largest increase in the similarity measure.

3 Experiments

For validation we used clinical X-ray mammograms of patients for whom an MRI was also acquired. The mammograms included annotated lesions of various pathologies. We also included cases that had an X-ray compatible clip inserted after biopsy. The clip was inserted at the position of the lesion and consequently the annotations and clip positions were used as gold standard correspondences across the mammograms. The inter-observer error calculated from 4 radiologists on 4 of the mammograms used in the experiments was $2.8mm$.

For the evaluation, we consider the centres of mass of the annotations and/or clips. Our error metric is the minimum of the 2D Euclidean distances between

ISBN: 978-87-996443-0-8

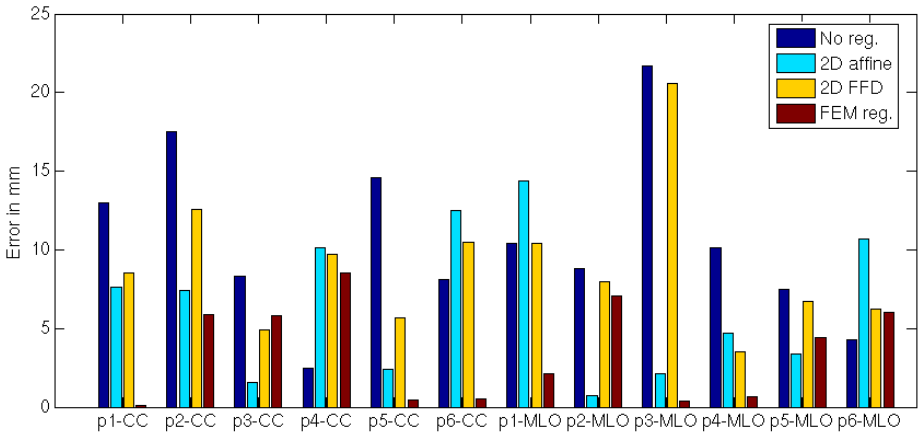


Fig. 2: 2D registration errors for the temporal CC and MLO view registrations of 6 patients. Our method is illustrated in red and it is compared against a 2D affine transformation (light blue), a 2D multi-resolution FFD (yellow) and the initial 2D error without registration (blue). The clip cases are patients p5 and p6.

the centre of mass of the annotation/clip in the second mammogram and the projection of the uncompressed and recompressed ray locus corresponding to the centre of the annotation in the first mammogram. The same metric has been previously used for the evaluation of a CC to MLO registration task [8].

3.1 Temporal registration

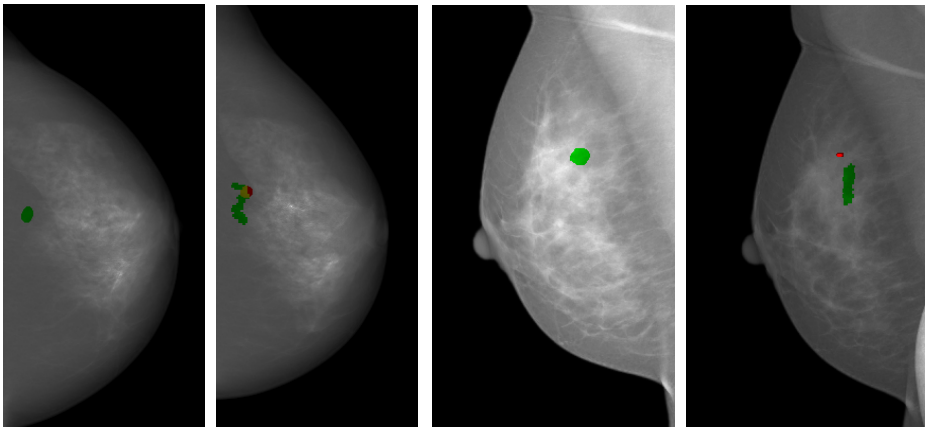
For the temporal registrations, we used data from six patients, four with annotated lesions at two time-points (mean radius: $6.8 \pm 2.7mm$) and two with annotated diagnostic mammograms that were mapped to X-rays obtained after biopsy, with a clip inserted at the lesion's position. Our method was compared against a 2D affine and a multi-resolution FFD registration, that have been proven previously to outperform other methods [1–3]. We also provide the errors before registration.

All results are given in Figure 2. The mean registration error of our approach is $3.5 \pm 3mm$, the 2D affine registration $6.5 \pm 4.6mm$, the 2D FFD registration $8.9 \pm 4.5mm$, and no registration $10.5 \pm 5.4mm$. As we can see, our method outperforms both 2D registration methods, giving the lowest error in 9 out of 12 registration tasks. It also provides good registration results for all cases, with a maximum error of $8.5mm$. Two example cases, one good registration result and one with a larger error, are illustrated in Figure 3. The CC temporal registration of patient p5 is illustrated in Figure 1.

3.2 CC to MLO registration

In these experiments, we demonstrate the use of the same framework for the matching of corresponding regions between CC and MLO view mammograms.

ISBN: 978-87-996443-0-8



(a) p1 CC (t_1) (b) p1 CC (t_2) (c) p6 MLO (t_1) (d) p6 MLO (t_2)

Fig. 3: Temporal registration results for 2 patients. The X-ray annotation at time-point t_1 is shown in green, as well as its projection on the second mammogram. The X-ray annotation of p1 and clip of p6 at time-point t_2 is shown in red. Their overlap is yellow.

We used ten cases. Six are the same subjects as those in the previous section. For the remaining four we had both MR and X-ray images acquired approximately at the same time-point. Two patients had inserted clips that were used for validation and eight had annotated lesions with mean radius $6.9 \pm 3.4mm$.

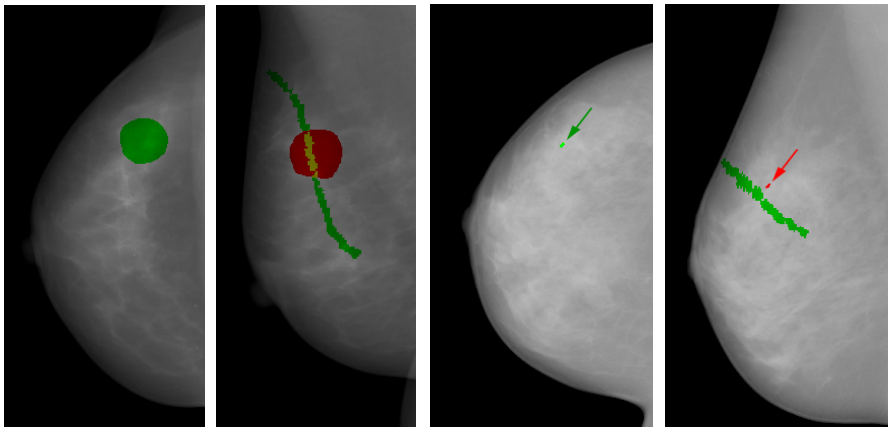
The registration errors are given in Table 1. The mean registration error for this task is $8.1 \pm 5.9mm$. Figure 4 illustrates two example cases, where it is clear that the projected area on the MLO view is not a straight line, as has been frequently approximated in the literature by distance-based metrics. Also its length can vary, depending on the position of the lesion inside the breast. The advantage of our method compared to other techniques that use a distance metric from the nipple and the chest wall, is that it can represent each correspondence case individually, as a patient-specific model is used.

The quantitative results indicate that the errors for the CC to MLO registrations are larger than those of the temporal registration. This was expected, as the difference in breast deformation between the same view (temporal) mammograms is smaller than the difference between the CC and MLO views. Compared to other methods proposed in the literature for the same task, the method of Kita

Table 1: Registration errors for the CC to MLO registrations on ten cases. The clip cases are p9 and p10.

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	mean	std
FEM CC-to-MLO	15.6	11.5	0.1	15.6	2.3	8.9	9.4	0.2	13.0	5.1	8.1	5.9

ISBN: 978-87-996443-0-8



(a) p8 CC (b) p8 MLO (c) p10 CC (d) p10 MLO

Fig. 4: CC to MLO registration results for 2 patients. The X-ray annotation/clip on the CC view its projection on the MLO are shown in green. The X-ray annotation/clip on the MLO view is shown in red.

et al. [8] is the one that is most closely related to our approach, as the authors used a simplified breast model reconstructed from the two mammographic views. The mean registration error in their study was $6.8mm$, which is comparable to our results, although the data sets used for validation are different.

4 Conclusion

We have demonstrated the use of a FEM-based registration framework for the mapping of findings between X-ray mammograms that have been acquired from different views or at different time-points. It is the first time that a patient-specific 3D model of the breast with a physically realistic biomechanical simulation is used for this task. The results showed that our method outperforms both the 2D affine and the multi-resolution FFD transformation that have previously been used for intensity-based temporal alignment. In our experiments we assumed that the breast tissue of the patient has not changed significantly over time. In future work and if an MRI is available at both time-points, an additional step could be added in the pipeline, that would align the two MRIs, to measure any changes in breast volume and structure.

Our proposed framework was also tested for the matching between CC and MLO mammograms, achieving a comparable accuracy to the use of a simplified geometrical model of the breast reconstructed from the two views. Future work includes validation on a larger data set and further investigation of the effect that the biomechanical modelling has on the registration accuracy, such as the use of more tissue types and more accurate modelling of the pectoral muscle.

ISBN: 978-87-996443-0-8

Acknowledgements This work was funded by the European 7th Framework Program, HAMAM, ICT-2007.5.3. The authors would like to thank the Radboud University Nijmegen Medical Centre and the Charite Universitätsmedizin Berlin for the data used in this study.

References

1. S. van Engeland, P. Snoeren, J. Hendriks, and N. Karssemeijer. A comparison of methods for mammogram registration. *IEEE TMI*, 22:1436–1444, 2003.
2. S. Pereira, J. Hipwell, V. Mccormack, C. Tanner, S. Moss, L. Wilkinson, L. Khoo, C. Pagliari, P. Skippage, C. Kliger, D. Hawkes, and I. Silva. Automated registration of diagnostic to prediagnostic x-ray mammograms: Evaluation and comparison to radiologists' accuracy. *Med. Phys.*, 37:4530–4539, 2010.
3. Y. Diez, A. Oliver, X. Llado, J. Freixenet, J. Marti, J. Vilanova, and R. Marti. Revisiting intensity-based image registration applied to mammography. *IEEE TITB*, 15:716–725, 2011.
4. S. Paquerault, N. Petrick, H. Chan, B. Sahiner, and M. Helvie. Improvement of computerized mass detection on mammograms: Fusion of two-view information. *Med. Phys.*, 29:238–247, 2002.
5. S. van Engeland, S. Timp, and N. Karssemeijer. Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views. *Med. Phys.*, 33:3203–3212, 2006.
6. B. Zheng, J. Tan, M. Ganott, D. Chough, and D. Gur. Matching breast masses depicted on different views: a comparison of three methods. *Academic Radiology*, 16:1338–1347, 2009.
7. C. Tanner, G. van Schie, N. Karssemeijer, and G. Szekely. Matching regions for mammographic views: Comparison and compensation for deformations. In *BIA MICCAI Workshop*, 2011.
8. Y. Kita, R. Highnam, and M. Brady. Correspondence between different view breast X-rays using curved epipolar lines. *CVIU*, 83:38–56, 2001.
9. J. Hipwell, C. Tanner, W. Crum, J. Schnabel, and D. Hawkes. A New Validation Method for X-ray Mammogram Registration Algorithms Using a Projection Model of Breast X-ray Compression. *IEEE TMI*, 26:1190–1200, 2007.
10. G. van Schie, C. Tanner, P. Snoeren, M. Samulski, K. Leifland, M. Wallis, and N. Karssemeijer. Correlating locations in ipsilateral breast tomosynthesis views using an analytical hemispherical compression model. *PMB*, 56:4715–4730, 2011.
11. N. Ruiter, R. Stotzka, T. Muller, H. Gemmeke, J. Reichenbach, and W. Kaiser. Model-Based registration of X-ray Mammograms and MR images of the female breast. *IEEE TNS*, 53:204– 211, 2006.
12. T. Hopp, M. Dietzel, P.A. Baltzer, P. Kreisel, W.A. Kaiser, H. Gemmeke, and N.V. Ruiter. Automatic multimodal 2d/3d breast image registration using biomechanical fem models and intensity-based optimization. *MedIA*, 17(2):209–218, 2013.
13. T. Mertzanidou, J. Hipwell, L. Han, Z. Taylor, H. Huisman, U. Bick, N. Karssemeijer, and D. Hawkes. Intensity-based MRI to X-ray mammography registration with an integrated fast biomechanical transformation. In *IWDM*, pages 48–55, 2012.
14. T. Mertzanidou, J. Hipwell, M. Cardoso, X. Zhang, C. Tanner, S. Ourselin, U. Bick, H. Huisman, N. Karssemeijer, and D. Hawkes. MRI to X-ray mammography registration using a volume-preserving affine transformation. *MedIA*, 16:966–975, 2012.

ISBN: 978-87-996443-0-8

Modelling Breast Deformation Using Partial Least-Squares Regression

Duane T.K. Malcolm[1], Chun M. Goh[1], Thiranja P. Babarenda Gamage[1], BaoSheng Hou[1], Poul M.F. Nielsen[1,2], Martyn P. Nash[1,2],

[1] Auckland Bioengineering Institute, The University of Auckland, New Zealand
[2] Department of Engineering Science, The University of Auckland, New Zealand
martyn.nash@auckland.ac.nz

Abstract. Tracking tumours between breast positions is useful for clinical applications involving breast biomechanics. This study investigates the use of partial least-squares regression (PLSR) to predict breast deformation from the prone to the supine positions. A population of finite element meshes of prone breasts were generated from MRI scans of 21 individuals. Finite element methods (FEM) were used to simulate the deformation of the breast from the prone to the supine positions, and these results were used to train PLSR statistical deformation models. The PLSR models were tested by predicting the supine breast geometry from the prone mesh using a leave-one-out approach. PLSR model predictions of the supine shape took 0.8 ± 0.06 ms to compute, whereas the FEM simulations took 111 ± 28 s. The volume averaged error between the PLSR predicted supine models and the associated FEM solutions was 1.9 ± 0.7 mm. This error is expected to reduce with a larger population of breast models. The PLSR approach shows good potential as a rapid and reliable alternative to FEM for breast biomechanics applications.

Keywords: Breast deformation modelling; statistical modelling; partial least-squares regression; finite element methods.

1 Introduction

Tracking breast tumours between different imaging modalities and positions within the breast can provide important clinical information. Tracking can be aided by using predictive computational models of the deformation of the breast between different states. Typically, finite element methods (FEM) are used to simulate deformations of the breast [4], however this approach can be computationally intensive and subject to failure when dealing with large deformations and distortions such as those that occur during mammographic compression [5,10]. An alternative approach is to use a statistical model [9,3,2] to potentially provide fast and robust predictions, although this may be at the expense of geometric accuracy. A statistical model of breast deformation can be generated using partial least-squares regression (PLSR) [6] between two geometric states for a population of individuals. In this study, we investigate this approach

ISBN: 978-87-996443-0-8

by considering deformations from the prone to the supine configurations of the breast. FEM was used to simulate the prone-to-supine deformations, and the resulting pairs of prone and supine geometries were used to create PLSR statistical deformation models. The predictive ability of the PLSR approach was assessed using a leave-one-out methodology.

2 Methods

2.1 Breast Model Generation and Mechanical Simulation

Finite element models were generated using breast image data from a population of 21 individuals (43 ± 9 years of age; use of data was approved by the local ethics committees, and written informed consent was obtained from all subjects). Breast meshes were fitted to data segmented from T1-weighted MRI scans (Figure 1a) acquired with the subject lying face-down prone in a clinical scanner (either a 1.5T Siemens Magnetom Vision ($n = 16$[3]) or Avanto ($n = 5$) scanner). Each breast geometry was represented using a 16-element tricubic-Hermite mesh with 1200 degrees of freedom (Figure 1b).

Supine MRI data were not available for these subjects, so the associated supine geometries were generated using FEM mechanics simulations [1]. The FEM simulations required two solution stages: (i) deforming the prone breast geometry to its unloaded reference state; followed by (ii) simulating the deformation from the reference state to the supine configuration of the breast in a similar manner to [8]. The mechanical simulations were subject to fixed boundary constraints at the sternum, axilla, chest wall, and superior and inferior edges of the breast models. A Neo-Hookean constitutive relation, with $c_1 = 200\,\mathrm{Pa}$, was used to represent the material properties of the breast tissues. Only the prone and supine breast meshes (as shown in Figure 1b) were used to generate the PLSR models in this study.

2.2 Training the PLSR Statistical Deformation Model

Given the population of associated prone and supine models from the FEM simulations, a statistical model of the deformation was trained using partial least-squares regression (PLSR). This was implemented using the *PLSRegression* class, which is based on the NIPALS algorithm [6] from the Scikit-Learn [7] Python module. The predictive ability of a PLSR model is dependent on the number of components computed when training the model. The maximum number of components is limited by the number of subjects or parameters (degrees of freedom) in the mesh.

The dimensions of the predictor (prone) and response (supine) matrices used to train the PLSR model were $M \times N$, where $M = 21$ subjects and $N = 1200$

[3] The authors thank Dr. Nico Karssemeijer for providing some of the breast MRI datasets used in this study.

ISBN: 978-87-996443-0-8

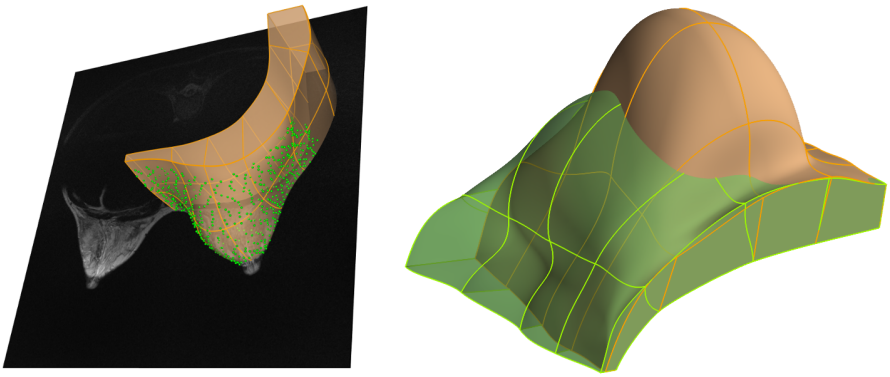


Fig. 1: a) A finite element mesh of the breast (orange surfaces) fitted to data (green dots) segmented from MRI of a participant lying in the prone position. b) Deformation from the prone (orange) to supine (green) states as predicted by the FEM mechanics simulation.

mesh parameters. The prone mesh data formed the predictor matrix. The difference between the supine and prone mesh parameters formed the data in the response matrix. Given this approach, the prediction from a PLSR model is added to the input prone mesh parameters to obtain a predicted supine geometry. Using the difference for the response matrix provided better deformation predictions compared to using the geometric parameters of the mesh, particularly in the neighbourhood of kinematic constraints, which may not be identically satisfied in the PLSR analysis.

2.3 Testing the PLSR Statistical Model

The predictive ability of the PLSR approach was tested using a leave-one-out analysis. Each subject was selected in turn to be the unseen subject, and the remaining subjects were used to train a PLSR model. This PLSR model was then used to predict the supine geometry from the prone model of the unseen subject. This PLSR-predicted supine model was then compared to the FEM supine model to quantify the error in the PLSR prediction (see Section 2.4 for error calculations). The leave-one-out approach was applied in turn to each of the 21 breast models. The mean and variance of the errors for the leave-one-out cases provide an indication of the predictive ability of the PLSR model.

The predictive ability of the PLSR model typically increases with the number of components, but with diminishing returns, and at the expense of increasing the computational time for training the PLSR models. To determine the number of components required to generate a sufficiently converged predictive model, PLSR models were also generated using a subset (1 - 19) of the components for each leave-one-out case.

ISBN: 978-87-996443-0-8

2.4 Error Calculation

The predictive ability of the PLSR approach was quantified by comparing the statistically predicted deformations to the FEM deformation simulations (used as the ground truth for this study) using two error measures. The volume averaged geometric error gives an indication of the expected absolute error when tracking features from a prone MRI to their position in the supine breast as required, for example, for surgery or tissue biopsy. The relative error gives an indication of the predictive ability of the PLSR model that is independent of the extent of deformation.

Volume Averaged Displacement (Δ). The deformation between the prone and supine states can be quantified using the volume averaged displacement (Δ), which is given by:

$$\Delta = \frac{\int_\Omega \|\mathbf{x}_{supine} - \mathbf{x}_{prone}\| \, dV}{V_T} \tag{1}$$

where locations of material point are tracked from the prone ($\mathbf{x}_{prone}$) to the supine ($\mathbf{x}_{supine}$) models, and V_T is the total volume of the model.

The **Volume Averaged Geometric Error (E_a)** is the volume averaged difference between the supine models generated using the PLSR and FEM approaches, given by:

$$E_a = \frac{\int_\Omega \|\mathbf{x}_{plsr} - \mathbf{x}_{fem}\| \, dV}{V_T} \tag{2}$$

where $\mathbf{x}$ is a material point on the supine models predicted by the PLSR and FEM approaches.

The **Relative Error (E_r)** is the volume averaged geometric error (Eq. 2) normalised by the volume averaged displacement (Eq. 1) given by $E_r = E_a/\Delta$.

3 Results

FEM simulations were used to compute the prone-to-supine deformation for 21 breast models, which had a mean volume of $1028 \pm 380\,\text{cm}^3$. The FEM solution time for the 21 cases was $111 \pm 28\,\text{s}$.

The time taken to train the PLSR models using 20 meshes (after the leave-one-out) was $19 \pm 1.5\,\text{ms}$ for models with one component, and this linearly increased to $78 \pm 4.3\,\text{ms}$ for models with all 19 components. The PLSR predictions of the 21 unseen supine breasts took $0.8 \pm 0.06\,\text{ms}$, five orders of magnitude faster than the FEM simulations. The prediction time was independent of the number of components in the PLSR model.

Figure 2 illustrates the geometric errors for the PLSR predicted prone-to-supine deformations for two example cases (medium and large breasts). Typically, the largest errors were observed towards the lateral half of the breast, however the variation in the error distributions was concentrated around the axilla, the lateral breast or evenly distributed.

ISBN: 978-87-996443-0-8

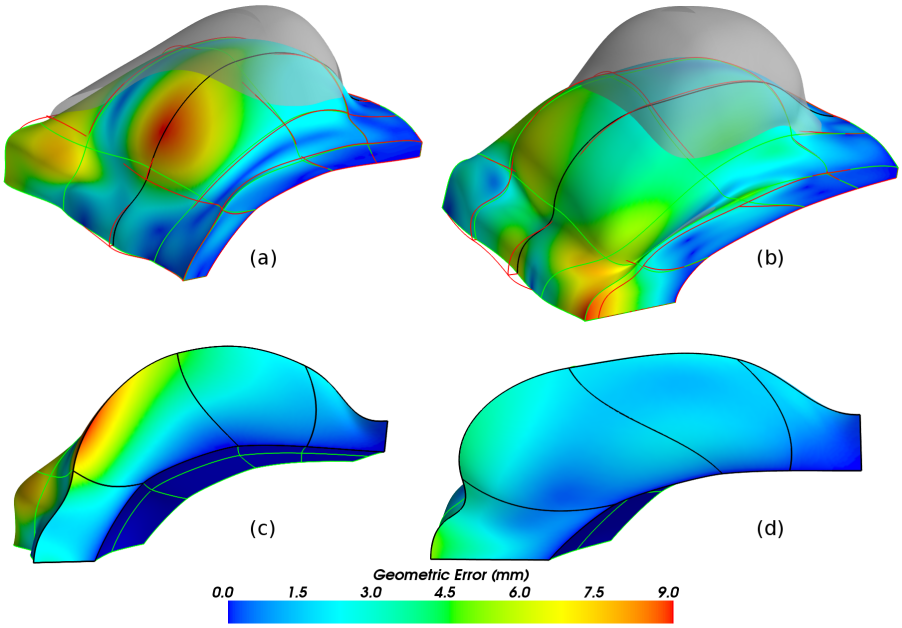


Fig. 2: Geometric errors between the PLSR and FEM predicted prone-to-supine deformations for a medium (a, c) and a large (b, d) sized breast. The volume averaged geometric errors were 2.1 mm and 1.8 mm, respectively. The axial cross-sectional views (c, d) show the internal geometric errors for the slice locations indicated by the black lines in the associated upper panels. The superimposed grey surfaces indicate the associated prone breast geometries.

The volume averaged displacement of the population was 18.5 ± 7.8 mm. The leave-one-out PLSR analysis using all 19 components was able to predict the supine configuration with a volume averaged geometric error of 1.9 ± 0.7 mm, and a relative error of $12 \pm 7\,\%$, compared to the FEM deformation simulations. The internal errors displayed on the axial cross-sections shown in Figure 2c and 2d illustrate that the maximal errors occur at the skin surface. This provides an upper limit on the tumour tracking error. However, given that tumours are located internally, the volume averaged geometric error is likely to provide a reasonable estimate of the tracking error.

The predictive ability of the PLSR approach depends on the number of components used to train the model (see Figure 3). A PLSR model with one component gave predictions with a volume averaged geometric error of 4.0 ± 1.8 mm and a relative error of $25 \pm 13\,\%$. There was a steady reduction in the errors as the number of components used was increased, and this plateaued at around 8 components, although there was a further minor reduction with 19 components. This indicates that a PLSR model with 8 or more components would be sufficient to predict the prone-to-supine deformation. Using additional components

ISBN: 978-87-996443-0-8

would further reduce the errors but with diminishing returns. Extending the population would be expected reduce the mean and the variance of the errors.

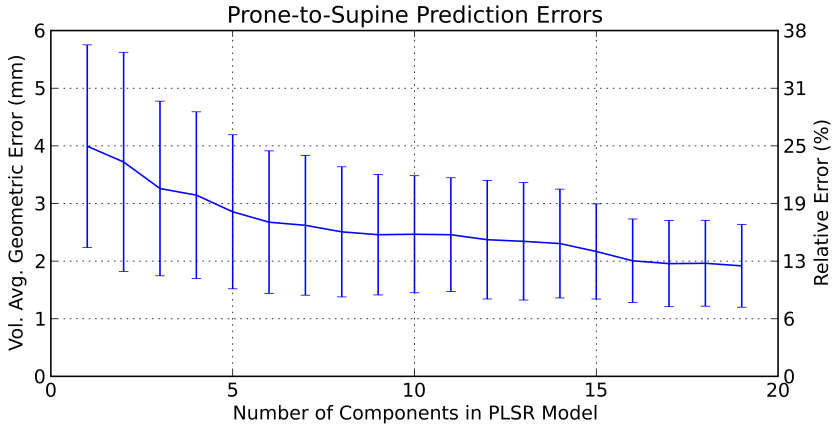


Fig. 3: Prone-to-supine prediction errors using PLSR models based on a leave-one-out analysis with 21 cases. The relative error was normalised against the mean displacement (18.5 ± 7.8 mm) for the FEM derived prone-to-supine deformation.

It is reasonable to expect that the geometric error is related to the degree of deformation, however the data for the leave-one-out analysis with 19 components showed a weak correlation due to significant variance (see Figure 4). This variance is likely due to a poor representation of the unseen breast in the trained PLSR model. It is expected that the error and the variance could be reduced by including more cases that span the population, hence improving the value of this approach in clinical applications.

4 Discussion and Conclusions

PLSR statistical models were able to predict prone-to-supine breast deformations with a volume averaged geometric error (compared to FEM predictions) of 1.9 ± 0.7 mm, and a relative error of $12 \pm 7\,\%$. The geometric errors increased with the degree of deformation. The observed variance in the prediction errors from the PLSR statistical models would most likely decrease using a larger and more representative population.

The PLSR predictions were five orders of magnitude faster than the FEM solutions, with calculations taking approximately 0.8 ms per model (FEM predictions took approximately 111 s). In addition, the PLSR approach was ideally robust (predictions were always attainable), as opposed to the FEM framework,

ISBN: 978-87-996443-0-8

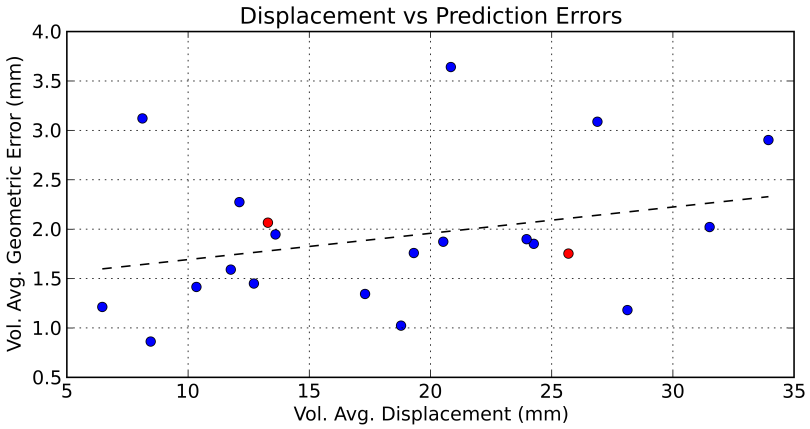


Fig. 4: The relationship between the volume averaged geometric error and the volume averaged displacement for predictions from the PLSR model with 19 components. The red dots indicate the medium (left) and large (right) sized breasts shown in Figure 2.

which was subject to solution instabilities due to large distortions or load steps. The speed of computation of the PLSR predictions can enable near-real-time simulations of deformation. This may be at the expense of prediction accuracy, however as we have demonstrated the errors are sufficiently low that the predictions from PLSR models can be used for some applications, such as visualisation and surgical planning. Furthermore, the PLSR model approach is robust, and can provide a good initial solution to the nonlinear FEM simulations, which would speed up the FEM mechanics solution. Future work for this approach includes extending the population of breast models, ideally to include both prone and supine models from scan measurements, to improve the predictions. Further investigations could also aim towards incorporating demographic data into the statistical model and test whether the inclusion of these additional parameters can improve predictions.

In summary, the PLSR statistical approach can be used to reliably and rapidly predict prone-to-supine deformations of the breast. Thus, this approach is a suitable alternative to FEM to address some of its shortcomings for clinical applications. Further research is required to improve the predictive ability of the statistical approach.

References

1. Babarenda Gamage, T.P., Rajagopal, V., Nielsen, P.M., Nash, M.P.: Patient-specific modeling of breast biomechanics with applications to breast cancer detection and treatment. In Gefen, A., ed.: Patient-Specific Modeling in Tomorrow's

ISBN: 978-87-996443-0-8

Medicine. Volume 09 of Studies in Mechanobiology, Tissue Engineering and Biomaterials. Springer Berlin Heidelberg (2012) 379–412
2. Feng, S.S.J., Patel, B., Sechopoulos, I.: Objective models of compressed breast shapes undergoing mammography. Medical Physics **40**(3) (2013) 031902
3. Gallo, G., Guarnera, G.C., Catanuto, G.: Human breast shape analysis using pca. In: BIOSIGNALS 2012 - Proceedings of the International Conference on Bio-inspired Systems and Signal Processing. (2010) 163–167
4. Han, L., Hipwell, J.H., Tanner, C., Taylor, Z., Mertzanidou, T., Cardoso, J., Ourselin, S., Hawkes, D.J.: Development of patient-specific biomechanical models for predicting large breast deformation. Physics in Medicine and Biology **57**(2) (2012) 455
5. Lee, A.W., Schnabel, J.A., Rajagopal, V., Nielsen, P.M., Nash, M.P.: Breast image registration by combining finite elements and free-form deformations. In Martí, J., Oliver, A., Freixenet, J., Martí, R., eds.: Digital Mammography. Volume 6136 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2010) 736–743
6. Martens, H., Nes, T.: Multivariate Calibration. Wiley series in probability and mathematical statistics. Wiley (1992)
7. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12** (2011) 2825–2830
8. Rajagopal, V., Chung, J., Nielsen, P.M.F., Nash, M.P.: Determining the finite elasticity reference state from a loaded configuration. International Journal for Numerical Methods in Engineering **72** (2007) 1434–1451
9. Tanner, C., Hipwell, J., Hawkes, D.: Statistical deformation models of breast compressions from biomechanical simulations. In Krupinski, E., ed.: Digital Mammography. Volume 5116 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2008) 426–432
10. Tanner, C., White, M., Guarino, S., Hall-Craggs, M.A., Douek, M., Hawkes, D.J.: Large breast compressions: Observations and evaluation of simulations. Medical Physics **38**(2) (2011) 682–690

ISBN: 978-87-996443-0-8

Evaluation of a B-spline-based breast compression simulation for correspondence analysis between MRI and mammographic image data

Julia Krüger[1], Jan Ehrhardt[1], Arpad Bischof[2] and Heinz Handels[1]

[1]Institute of Medical Informatics, University of Lübeck, Germany
[2]Department of Radiology and Nuclear Medicine, UK-SH Lübeck, Germany
krueger@imi.uni-luebeck.de

Abstract. Two-dimensional mammography is the major imaging modality in breast cancer detection. To overcome shortcomings like the projective nature of the imaging technique or poor contrast between tumorous and healthy fibro-glandular tissue, contrast-enhanced MRI can be used additionally. The combined analysis and determination of corresponding structures in the acquired mammogram and MRI data is challenging due to large breast deformations during the image acquisition. In this paper, we evaluated a previously proposed method for breast compression simulation using an ICP-based B-spline registration. The evaluation, based on 14 MRI datasets with 19 corresponding mammograms, revealed an average distance of 4.40 mm between corresponding structures.

1 Introduction

Mammography is the most commonly used breast imaging modality in breast cancer screening and diagnosis. It provides a two-dimensional projection image of the breast, where the breast is compressed between two plates. Furthermore, it is characterized by a high resolution ($[< 0.1\times < 0.1]mm^2$) and three different views are usually acquired: Medio-Lateral (ML), Medio-Lateral Oblique (MLO) and Cranio-Caudal (CC). The analysis of mammographic images can be challenging, because of the projective nature of this imaging technique and a poor contrast between tumorous and healthy fibro-glandular tissue. To overcome these shortcomings, contrast-enhanced magnetic resonance imaging (MRI) can be used additionally [1]. It provides a 3D image (resolution $\approx [1 \times 1 \times 2\text{-}3]mm^3$) of the female breast with additional functional information, which can be used for tumor staging. But in contrast to mammography, MRI is acquired without compression of the breast. During the acquisition of the MR image, breast coils are used and the woman lies on her stomach with her breast pendulous.

Before the combined information provided by both imaging techniques can be analyzed, the examining radiologist has to localize corresponding structures in the two-dimensional mammogram and the three-dimensional MR image. This task is challenging due to (1) different dimensionality and spatial resolution,

ISBN: 978-87-996443-0-8

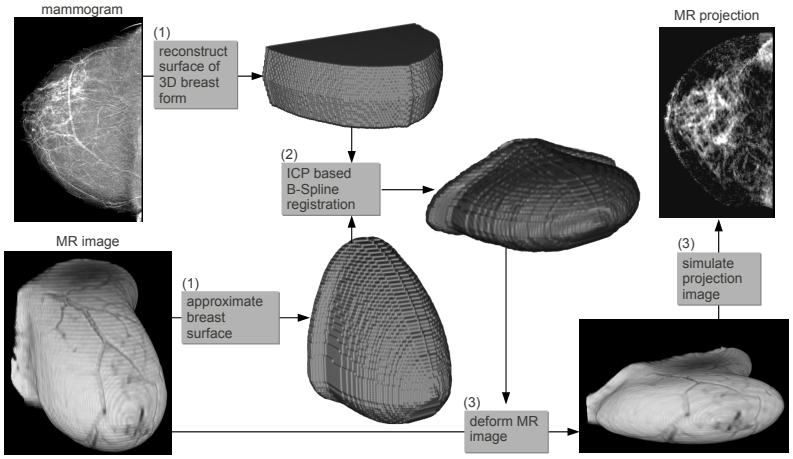


Fig. 1. *Overview of the processing steps: (1) determination of 3D breast volume surfaces: extraction of 3D MR breast surface and reconstruction of 3D mammographic breast surface, (2) ICP-based B-spline registration of the determined surfaces and (3) deformation and simulation of an MR projection image [2].*

(2) different contrasts between tissue types and (3) different deformation of the breast during acquisition. The huge breast deformation differences make a localization of corresponding structures especially challenging.

In [2], we presented an ICP-based B-spline registration to approximate the breast deformation that occurs during the mammographic acquisition. To apply the proposed surface-based registration, the MR surface information is extracted from the 3D MR image and the mammographic surface information is approximated by reconstructing the compressed mammographic volume. The resulting deformed MR image is projected onto the 2D plane to enable a comparison with the mammogram. In this work, we evaluated the registration using MRI datasets of 14 patients and 19 corresponding mammograms.

2 Methods

Previously presented approaches for correspondence analysis between MRI and mammography have often used FEM techniques to simulate the breast compression [3–5]. Mertzanidou et al. performed a 2D/3D registration between a mammogram and an MR image using a 3D affine transformation of the MR image [6]. In contrast to these methods we proposed a 3D/3D registration using the surface information of the breast in [2].

Therefore, the first step is the reconstruction of the 3D surfaces (Fig. 1, (1)). The 3D target shape of the compressed mammographic volume is automatically reconstructed by back projection of the 2D breast contour into 3D space. Information like compression plate distance and X-ray source geometry are extracted

ISBN: 978-87-996443-0-8

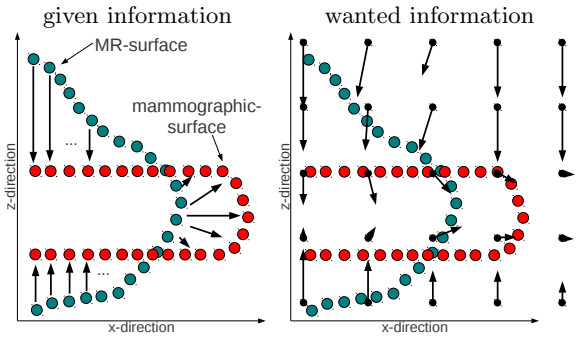


Fig. 2. *A scattered data approximation for B-splines is implemented that uses the surface distances between the given surfaces (left) as scattered data and computes a smooth deformation field that contains information of the entire breast volume (right).*

from the DICOM-header of the mammogram. The 3D source shape of the MRI dataset is approximated automatically using a slice-wise elliptical approximation of the MR breast surface to achieve a smooth convex breast surface without foldings. Then, a surface based registration of both breast volumes using an ICP-based B-spline registration is applied (Fig. 1, (2)). To evaluate the simulated deformation of the MR image, manually marked structures in the MRI are projected into the 2D plane to enable a comparison of the deformed MR data to the corresponding mammogram (Fig. 1, (3)).

2.1 B-spline deformation

The basic idea of the surface-based registration is to determine a free-form transformation of the MR breast based on cubic B-splines. To propagate the surface information throughout the entire breast volume, the control point grid of the B-spline approximation has to cover the whole breast volume and the spacing between the control points has to be rather wide (see Sec. 3.1). To manipulate a uniform control point grid (Fig. 2, right) using the local surface information (Fig. 2, left), a scattered data approximation for B-splines introduced by Lee et al. [7] is implemented.

The first step in our developed algorithm is to determine the correct correspondences between the surfaces (arrows in Fig. 2, left). Due to the large shape differences between the surfaces, an ICP with stepwise B-spline deformation of the MR surface is developed for this purpose. Following this, a one-step B-spline approximation is computed using the determined surface correspondences to describe a smooth deformation.

2.2 Alignment of the MR breast

Before the two surfaces are registered we perform an automatic initial alignment. For this purpose, the mamilla positions are automatically extracted using the

ISBN: 978-87-996443-0-8

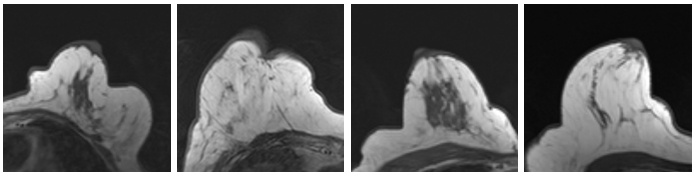


Fig. 3. *Slices of MR images that were excluded due to large deformation by breast coils or previous operations.*

curvature information of the surfaces. Furthermore, the pectoralis muscle (if visible) is extracted in the images of both modalities. In future work, the muscle determination is planned to be fully automatic, but for this evaluation it was done manually. Then, the initial alignment between the surfaces is realized by matching the mamilla positions, the pectoralis muscles and rotating the MR surface by the main projection angle (CC $\approx 0°$, MLO $\approx 45°$, ML $\approx 90°$) taken from the DICOM information.

3 Experiments and Results

MRI datasets of 24 patients were available, from which we used 14 patient datasets with 19 corresponding mammograms (CC, MLO and ML) for evaluation. We excluded MR data with breasts that were highly deformed by previous operations or deformed by non-fitting breast coils that are used during the MR acquisition (see Fig. 3). Furthermore, we excluded mammograms that showed no distinct structures that could be found in the corresponding MR image.

The simulated deformation was computed using the breast surface extracted from the T1-weighted MR data. Because, the contrast between dense healthy tissue and cysts or lesions is rather poor in T1-weighted images, we additionally used the STIR (Short-Tau Inversion Recovery) images, where the surrounding fat is suppressed, to identify interesting structures for the evaluation (see Fig. 5).

3.1 Selection of grid spacing

As mentioned in section 2.1, the grid of control points that is manipulated with respect to the surface differences has to cover the entire breast. Furthermore, to ensure that a manipulation of a control point near to the surface leads to a manipulation of its neighbors deep inside the volume, the spacing between the control points has to be rather wide. The slices in y-z-direction through the resulting deformation fields in Fig. 4 show the effect of the control point grid spacing. If the spacing is chosen to small, we only get information near the surface (high deformation between/near the surfaces, no deformation in the middle of the breast). And if the spacing is chosen to large, the resulting deformation field is smooth but not flexible enough to describe the wanted deformation. An other possibility is to use a multi scale B-spline approximation. But the problem of

ISBN: 978-87-996443-0-8

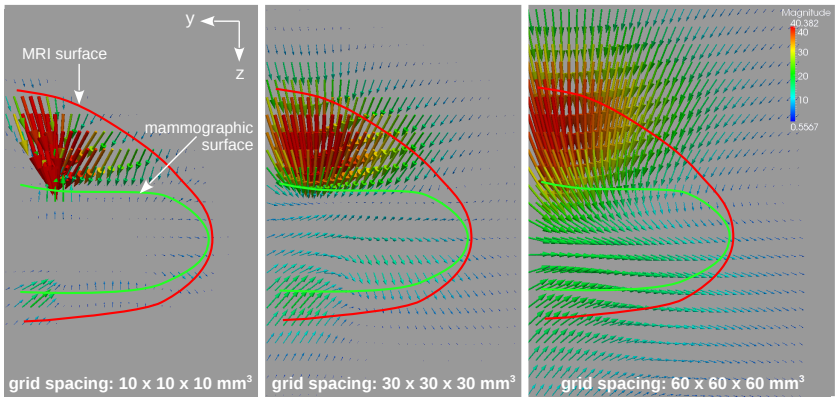


Fig. 4. *Slices of the resulting deformation fields in y-z direction using different grid spacings between the B-spline control points: very small spacings (left) lead to a high deformation near to the surface and no deformation deep inside the breast volume, whereas a very wide spacing leads to a very smooth deformation field (right).*

defining the first/widest control point spacing stays the same. Furthermore, we decided against a multi scale approach because the target surface (back projected mammographic surface) is just an approximation and the refined scales would not improve the results.

3.2 Influence of initial alignment

Due to the rather large compression during mammographic acquisition and the projective nature of the mammogram, the quality of the simulated MR deformation depends highly on the initial alignment between the breast surfaces. Therefore, we varied this initial alignment by rotating the MR surface manually by 3 degree (-3°,0°,+3°) in each direction (x-, y- and z-axis; with the mamilla position as rotation center) to determine the effect on the resulting projection of the deformed MR image. The best result (small distances between corresponding structures) for each deformation was chosen manually for further evaluation. This individual manual alignment is done to show the effect and dependence on the alignment. In future work this step should be done automatically.

3.3 Evaluation

To evaluate the resulting deformations, we marked corresponding structures like cysts, dense tissue, scared tissue, clips or lesions in the 3D MRI dataset and the 2D mammogram (see Fig. 5). After deforming the MR image with the determined deformation field, the deformed and marked MR structures are projected onto the 2D plane using the information about X-ray source geometry taken from the mammogram DICOM-header. Then, the distance between the area defined in the mammogram and the projected area from the MR image was determined

ISBN: 978-87-996443-0-8

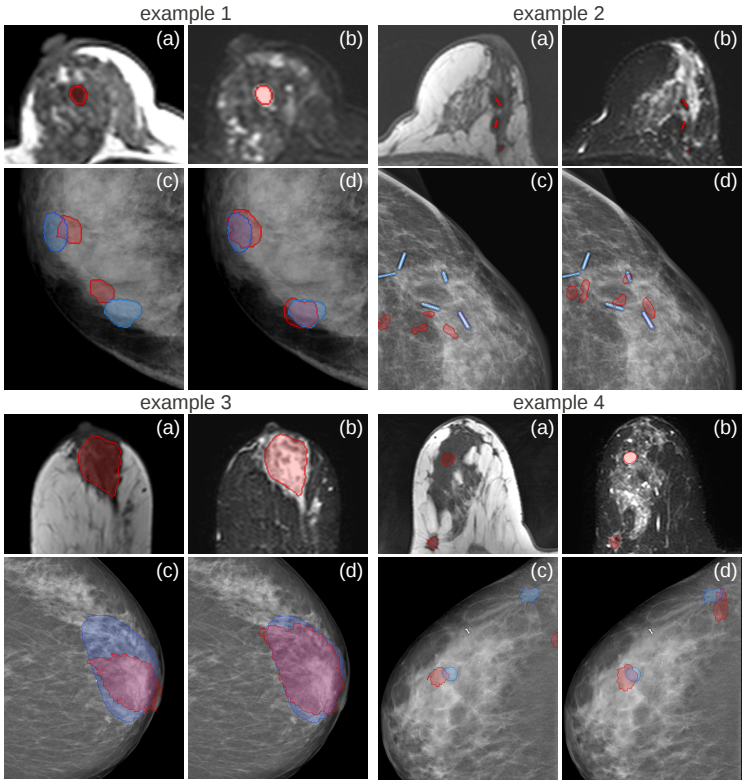


Fig. 5. *Four examples of corresponding structures: cysts (ex. 1), vessel clips (ex. 2), dense tissue (ex. 3), cysts and scared tissue (ex. 4) in mammographic and MR images. The images show the marked structure (red) in MRI slices (T1-weighted (a) and STIR (b)) and corresponding 2D marked structures (blue) in the mammogram (c,d). The images in (c) show the projected MR structures (red) before the MR image is deformed and the figures in (d) show the projected structures after the deformation.*

Table 1. *Resulting distances [mm]: minimal, maximal and mean distances [mm] for different grid spacings (left) and for a fixed grid spacing ($40 \times 40 \times 40$ mm^3) before and after deformation and with additional alignment of the MR breast (right).*

grid spacing $[mm^3]$	min. distance	max. distance	average distance
$30 \times 30 \times 30$	0.69	10.88	4.95
$40 \times 40 \times 40$	1.01	10.62	4.40
$50 \times 50 \times 50$	1.10	12.81	6.21
$60 \times 60 \times 60$	1.34	14.07	7.67

	min. distance	max. distance	average distance
undeformed	1.86	43.58	14.20
deformed	2.01	30.50	9.12
deformed and aligned	1.02	10.62	4.40

by using the centroids of the marked structures. These distances are computed before and after the deformation and after the manual variation of the alignment.

ISBN: 978-87-996443-0-8

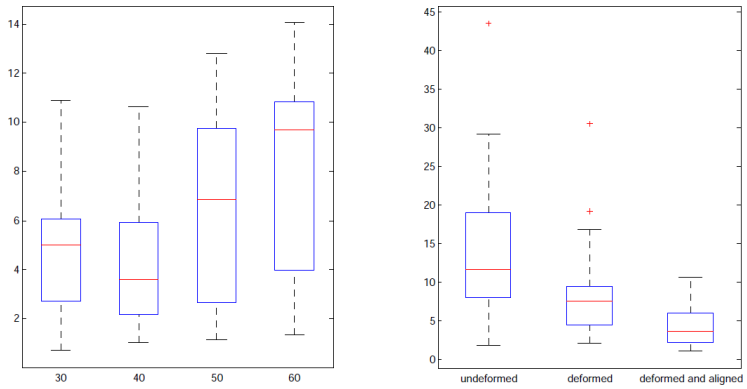


Fig. 6. *Boxplots of the determined distances: left: results (distances in [mm]) for different grid spacings (from* $30 \times 30 \times 30\ mm^3$ *to* $60 \times 60 \times 60\ mm^3$*); right: results (for grid spacing* $40 \times 40 \times 40\ mm^3$*) before deformation, after deformation and after further manual alignment and deformation of the MR breast.*

3.4 Results

The boxplot in Fig. 6 left and Tab. 1 left show the results for the different grid spacings for the B-spline control points. The best results for each MRI breast are dependent on the size of the breast or the complexity of the wanted deformation. On an average, a grid spacing of $40 \times 40 \times 40\ mm^3$ showed the best results with $4.40\ mm$. In Fig. 6 (right) and Tab. 1 (right) the results before deformation (with aligned mamilla positions), after deformation and after further manual alignment and deformation of the MR breast are shown. It is obvious that the proposed deformation can improve the distance between corresponding structures (from mean $14.20\ mm$ to $4.40\ mm$). But furthermore, the results show that the quality of the deformation depends highly on the initial alignment between the two surfaces. Without an individual optimization of the initial alignment the accuracy is $9.12\ mm$.

4 Discussion

The problem of the definition of corresponding points between mammographic and MRI data is very challenging. The compression of the breast during the mammographic acquisition is rather large and at the same time the female breast has no distinct landmarks but the mamilla. Due to the high flexibility of the female breast tissue, mammographic images of the same breast can vary by rotations angle, deformation and range of visibility (is pectoralis muscle visible or not). Furthermore, the breast in the MRI is not entirely free from deformation because of the breast coils. All these problems lead to the fact that an automatic simulation of the breast deformation without expensive and manual preprocessing will only approximate the real compression.

ISBN: 978-87-996443-0-8

We proposed a surface-based registration for the simulation of mammographic breast deformations in MRI datasets. To overcome the problem of lacking information inside the breast volume, an ICP-based B-spline registration using a scattered data approach is implemented. It is planed to embed the deformation into an automatic work-flow to facilitate the determination of correspondences between mammogram and MRI. For this purpose, automatic determination of the pectoralis muscle in both images and initial alignment and adaption of the alignment to the individual image data has to be improved.

Our resulting deformation field is a smooth approximation of the mammographic deformation. We further assume that the breast tissue is homogeneous and boundaries between different tissue types (fibro-glandular, fat, muscles) and their different mechanical properties are ignored. This can lead to an overdone compression and expansion of dense tissue.

In this work, we focused on the evaluation of the deformation itself. In the evaluation MRI datasets were excluded that were highly deformed by previous operations or deformed by non-fitting breast coils, because the proposed algorithm is not suitable for these deformations. The results of the remaining 14 patient datasets show that a surface-based smooth deformation can improve the distances between corresponding structures. Furthermore, it shows the dependence on the initial alignment between the two surfaces caused by the high flexibility of the deformation.

References

1. Yuan, Y., Giger, M.L., Li, H., Bhooshan, N., Sennett, C.A.: Multimodality computer-aided breast cancer diagnosis with FFDM and DCE-MRI. Academic Radiology **17**(9) (2010) 1158 – 1167
2. Krüger, J., Ehrhardt, J., Bischof, A., Handels, H.: Breast compression simulation using ICP-based B-spline deformation for correspondence analysis in mammography and MRI datasets. Image Processing, SPIE Medical Imaging (2013)
3. Ruiter, N., Stotzka, R., Muller, T.O., Gemmeke, H., Reichenbach, J., Kaiser, W.: Model-based registration of X-ray mammograms and MR-images of the female breast. Nuclear Science, IEEE Transactions on **53**(1) (2006) 204–211
4. Tanner, C., White, M., Guarino, S., Hall-Craggs, M.A., Douek, M., Hawkes, D.J.: Large breast compressions: Observations and evaluation of simulations. Med Phys **38**(2) (2011) 682–690
5. Hopp, T., Dietzel, M., Baltzer, P., Kreisel, P., Kaiser, W., Gemmeke, H., Ruiter, N.: Automatic multimodal 2D/3D breast image registration using biomechanical FEM models and intensity-based optimization. Med Im Analysis **17**(2) (2013) 209–218
6. Mertzanidou, T., Hipwell, J., Cardoso, M.J., Zhang, X., Tanner, C., Ourselin, S., Bick, U., Huisman, H., Karssemeijer, N., Hawkes, D.: MRI to X-ray mammography registration using a volume-preserving affine transformation. Medical Image Analysis **16**(5) (July 2012) 966–975
7. Lee, S., Wolberg, G., Shin, S.Y.: Scattered data interpolation with multilevel B-splines. IEEE TVCG **3** (1997) 228–244

ISBN: 978-87-996443-0-8

Registration of Automated 3D Breast Ultrasound Views

Tao Tan[1], Björn Eiben[2], Bram Platel[3], Jan van Zelst[1], Lianghao Han[2], Thomy Mertzanidou[2], Stian Johnsen[2], John Hipwell[2], Ritse M. Mann[1], David Hawkes[2], and Nico Karssemeijer[1]

[1] Department of Radiology, Radboud University Nijmegen Medical Centre, the Netherlands
[2] CMIC, University College London, London, UK
[3] Fraunhofer MEVIS, Bremen, Germany

Abstract. Automated 3D breast ultrasound (ABUS) has the potential to be widely used for women with dense breasts in cancer screening. During screening, more than three ABUS views, targeting different areas of the breast, are generated. Those views show partly overlapping areas. Relating findings between views is challenging due to different coverage and the large deformation that the breast undergoes. We present a registration method to determine spatial correspondence between an anterior-posterior (AP) view and a lateral (LAT) or medial (MED) view of the breast. The method incorporates a biomechanical model and a rigid registration. Our method was applied on a dataset of 19 pairs of AP and LAT/MED views with a validated lesion on the both views. The registration error was 12.99 mm±7.18 mm.

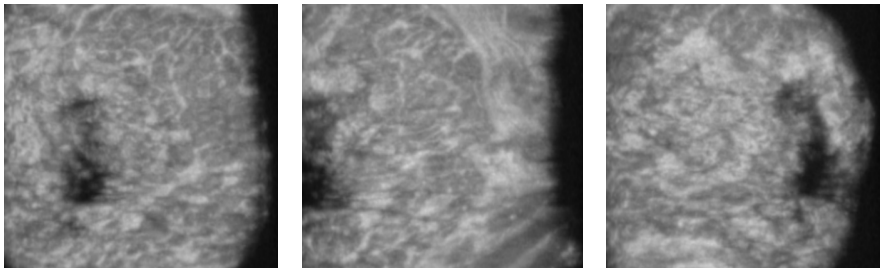
Keywords: biomechanical model, image registration, automated 3D breast ultrasound

1 Introduction

Automated 3D breast ultrasound (ABUS) is rapidly gaining popularity as a low cost adjunct screening modality for women who have dense breast tissue. Wenkel et al. [1] show that using ABUS provides a high reliability of detection of solid and cystic lesions. Giuliano and Giuliano[2] showed that using automated 3D breast ultrasound (ABUS) together with mammography also improves the detection rate, resulting in 12.3 breast cancers per 1000 screenings, compared to 4.6 per 1000 by mammography alone. In a reader study[3], using ABUS led to a significantly higher sensitivity for malignant lesions than for benign lesions.

Compared to 2D handheld ultrasound, the screening procedure using ABUS is to some extent standardized. For each scan of ABUS, a membrane which is kept flat under tension by being fixed to a frame compresses the breast, and a wide transducer (15.4 cm) is swept across the membrane to generate a volumetric representation of the breast. To cover the whole breast, usually 3-5 volumes (called views in this paper) are generated targeting different areas of the breast. Fig. 1 shows the most common three views of a breast. Although each view covers

ISBN: 978-87-996443-0-8



(a) (b) (c)

Fig. 1. Coronal slices of the three most commen views in ABUS: (a) anterior-posterior view, (b) medial view, and (c) lateral view.

a part of the breast, different views often show overlapping areas of a breast. For example, the anterior-posterior view will to some extent also be covered by the other views. During screening, radiologists have to localize the same lesions in different views to interpret the lesion thoroughly. Reading different volumetric images and searching for correspondences between different views is time-consuming. Therefore an automated registration method between different views could aid the screening procedure. Furthermore, the image registration can also provide the linkages between the suspicious regions detected by a computer-aided detection (CAD) system (i.e. [4] and [5]) in different ABUS images, which might help to improve the CAD performance.

For breast image registration, Ruiter et al.[6] used a biomechanical modeling approach to simulate the compression of the breast due to mammographic imaging. A finite element (FE) model was constructed for a breast and the mammographic compression was simulated using two contact plates which during the simulation compress the breast model. Beyond the patient specific geometry of the breast, spatially varying tissue properties can be considered with personalized biomechanicial models. For ABUS, a method was previously developed to determine the lesion correspondences between different ABUS views using a grid search in feature space [7]. However this method does not provide all point correspondences between the two views.

Each ABUS view is generated by imaging a breast under compression which results in a large deformation between different views. Each view only covers a partial area of the breast, and thus, directly using an intensity-based registration method on different two ABUS views is inappropriate. In this paper, we introduce a hybrid method which incorporates biomechanical modeling and intensity-based registration.

2 Method

In our method, we use a generic model to approximate the shape of breasts and a contact plate is used to simulate compressions of an anterior-posterior (AP) view

ISBN: 978-87-996443-0-8

and a lateral (LAT) or medial (MED) view. Next, given the original AP view and the displacement information obtained from the simulations, we generate a simulated LAT or MED view image. Subsequently we apply the intensity-based registration between the simulated view and the original LAT/MED view.

2.1 Compression Simulation using a Finite Element Model

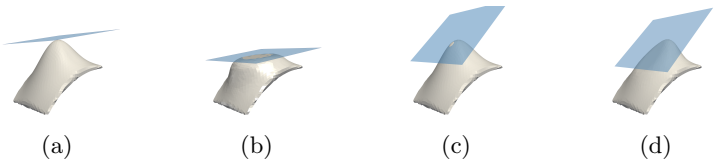


Fig. 2. Simulations of AP imaging (a) before the compression and (b) after the compression and simulations of LAT/MED imaging (c) before the compression and (d) after the compression

During the ABUS imaging, a dedicated membrane under tension compresses the breast. A finite element (FE) model was used (solver NifySim [4] [8]) to simulate the large deformation caused by compressions. At present, we are not able to obtain an uncompressed breast model from ABUS images, and therefore we use a generic breast phantom with a shape defined by a Gaussian and the rib cage is approximated by a cylinder. Although, in principle, each element can be assigned a different tissue type and appropriate material property parameters, here we use a hyper-elastic Neo-Hookean material model and assume it is homogeneous across the model domain. To mimic the compression, a frictionless contact plate is added to the FE model. The compression simulation is achieved by moving the plate towards the model until a certain displacement is achieved. The normal of the plate and the direction of plate movement during simulation can be defined separately.

During imaging, the direction and the displacement of the membrane are not recorded. For the simulation, the direction and the magnitude of the displacement of the plate are needed as inputs. To obtain a simulation close to the real imaging, for AP imaging, we conducted a number of simulations by moving the plate in anterior-posterior direction with a set of possible displacements. For LAT/MED imaging, we performed a number of simulations by moving the plate from the side to the center with a set of possible displacements and by placing the plate at a set of possible angles. Fig 2 shows an example of simulations before and after the compression for AP and LAT/MED imaging.

The displacement of each node of the model are recorded for each simulation. Therefore, the displacement of each node from the state after AP compression

[4] http://sourceforge.net/projects/niftysim

ISBN: 978-87-996443-0-8

simulation to the state after LAT/MED compression simulation can be computed. For image resampling, a dense deformation vector field is generated from the nodal displacements using a scattered data interpolation scheme based on multi-level B-splines. To generate a simulated LAT/MED view, a mapping between the image and model is required. In this work, we align the compressed model and the image by mapping the nipple position which is manually annotated.

Fig 3 shows a set of simulated MED views, using different pairs of AP and MED simulations.

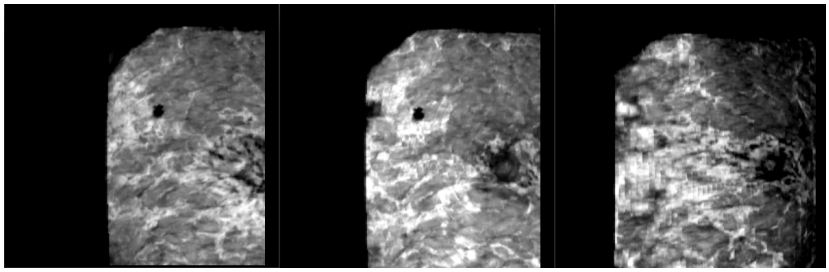


Fig. 3. Three simulated MED image generated from different pairs of AP and MED simulations

2.2 Rigid Registration

Although a simulated LAT/MED view can be generated from the simulations and the AP view, there is still misalignment between the simulated image and the reference image (the real LAT/MED view). A final registration step is performed to improve the alignment between the two images. For registration, a mask for the reference image can be provided. Since the overlapping area between different views is close to the nipple, we used a mask which includes voxels within 48 mm of the nipple in the coronal plane and exclude the voxels beyond the chest wall [9] and in the background. As the overlap area between the views can not be accurately extracted for image registration, using affine or non-rigid image registration would result in large misalignments. Therefore we use rigid image registration for the registration.

Fig 4 shows the result before and after the rigid transformation.

2.3 Selecting the Best Registration

In the deformation simulation step, for AP compression, the contact plate is moved in anterior-posterior direction with 3 different displacements and for LAT/MED compression simulation, the contact plate is oblique at a side of

ISBN: 978-87-996443-0-8

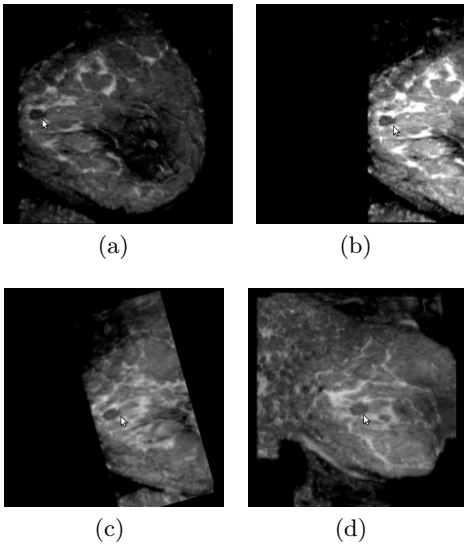


Fig. 4. (a) orginal AP view, (b) generated simulated MED view, (c) registered simulated MED view, and (d) original MED view with a corresponding lesion indicated by the cursor.

the breast and moved toward the breast with 4 different angles and 3 different displacements.

The employed set of plate displacements and compression angles involved in the simulations result in 36 pairs of AP and LAT/MED simulations and thus a set of simulated LAT/MED images. Each simulated LAT/MED image is rigidly registered to the acquired LAT/MED image. To obtain the best registration, we selected the registered image which gives the lowest absolute intensity difference between the registered image and the reference image within the mask described above. Fig. 5 shows the flowchart of our registration procedure.

With the registration transformation and displacements of each node between the AP simulation and the LAT/MED simulation, each point in the LAT/MED image can be mapped to the AP image. The following equation summarizes the transformation from LAT/MED to AP.

$$T_{MED2AP} = Tnip^{-1}_{LAT/MED}.Tmodel^{-1}_{LAT/MED}.Tmodel_{AP}.Tnip_{AP}.T_{Rigid} \quad (1)$$

where $Tnip$ is the transformation of the model to the image, $Tmodel$ is the simulation of either AP, MED or LAT views and T_{Rigid} is the rigid registration.

3 Experiments and Results

Our proposed method establishes the correspondence between an AP and a LAT or MED view. For validation we have applied our registration method on

ISBN: 978-87-996443-0-8

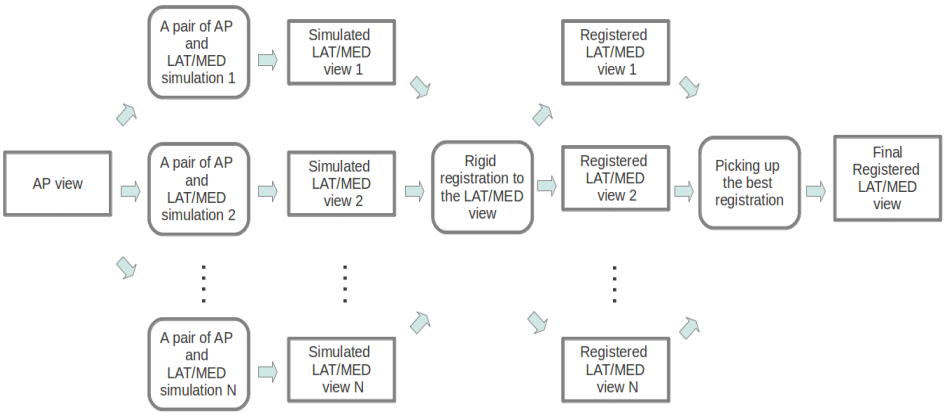
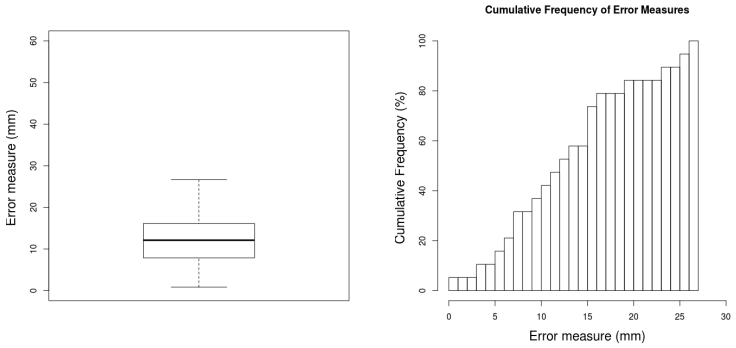


Fig. 5. The flowchart of our registration procedure.



(a) (b)

Fig. 6. (a) Box plot of the error measures. (b) Cumulative histogram for the error measures

ISBN: 978-87-996443-0-8

19 pairs of ABUS images. A corresponding lesion is visible in both images and is annotated under the supervision of radiologists.

For evaluation of the accuracy of the registrations, we measured the distance between the annotated center of the lesion in the AP view and the transformed annotated center of the lesion of the LAT or MED view.

From the 19 registrations of our dataset, we obtained a mean distance of 12.99 mm with a standard deviation of 7.18 mm. Fig 6(a) and Fig 6(b) shows the box plot of the measures and the histogram, respectively.

For comparison, we also applied the previously developed voxel-based grid search algorithm [7] which uses geometrical features only. This resulted in a error of 14.71 mm$\pm$13.29 mm. The difference between the two methods is not significant (paired t-test, $p = 0.62$).

4 Conclusion and Discussion

In this paper, we have presented a method of registration between two different ABUS views of a breast: one anterior-posterior view and one lateral or medial view. Our registration is a combination of biomechanical modeling and rigid intensity-based registration. The biomechanical simulation recovers large scale deformation between the two different views and thus compensates for the compression directions and intensity-based rigid registration further refines the registration.

About 80 percent of the registration results in a target registration error smaller than 20 mm. The preliminary results are promising in the sense that the combination of biomechanical simulations and intensity based registration should enable to assist radiologists to read ABUS images and find the correspondences more easily. The correspondences may also improve computer-aided cancer detection and diagnosis of suspicious regions.

The main application of our work is synchronization of cursors in different views to facilitate reading in a screening environment. From discussions with radiologists, we know that a registration error in the order of one cm would already be acceptable for this purpose. Note that the average cancer diameter reported in literature[10] is 15 mm in ABUS datasets. High accuracy is difficult to obtain because of the large deformations and intensity changes due to different ultrasound probe orientations. Our method achieves a mean error of 12.99 mm which is promising for a reading aid, although the error variance needs to be further suppressed.

Our method has its limitations. Firstly the model applied here is a generic model which is not able to handle the variance in breast shapes between patients. In the future, using the subject's cup size, we will build patient-specific models. Second, we assume an isotropic and homogeneous material model for the whole breast model. A per voxel tissue classification would allow us to build a more realistic biomechanical model of the breast with spatially varying material properties which should lead to more realistic simulations.

ISBN: 978-87-996443-0-8

The rigid registration used here only relies on the voxel intensity. To further improve the result, voxel-wise features describing textual tissue property can be integrated in the registration which might improve the registration.

There are few cases where the method completely fails. These failures are due to little common coverage between different views. In those cases, a voxel-based search strategy might be an alternative[7].

References

1. E. Wenkel, M. Heckmann, M. Heinrich, S. A. Schwab, M. Uder, R. Schulz-Wendtland, W. A. Bautz, and R. Janka. Automated breast ultrasound: lesion detection and bi-rads classification–a pilot study. *Fortschr Röntgenstr*, 180:804–808, 2008.
2. Vincenzo Giuliano and Concetta Giuliano. Improved breast cancer detection in asymptomatic women using 3d-automated breast ultrasound in mammographically dense breasts. *Clinical Imaging*, 2012.
3. Jung Min Chang, Woo Kyung Moon, Nariya Cho, Jeong Seon Park, and Seung Ja Kim. Radiologists' performance in the detection of benign and malignant masses with 3d automated breast ultrasound (abus). *European Journal of Radiology*, 78:99–103, 2011.
4. Tao Tan, Bram Platel, Roel Mus, and Nico Karssemeijer. Detection of breast cancer in automated 3D breast ultrasound. In *Medical Imaging*, volume 8315 of *Proceedings of the SPIE*, pages 831505–1–831505–8, 2012.
5. W. Moon, Y. Shen, M. Bae, C. Huang, J. Chen, and R. Chang. Computer-aided tumor detection based on multi-scale blob detection algorithm in automated breast ultrasound images. *IEEE Transactions on Medical Imaging*, 2012.
6. N.V Ruiter, R. Stotzka, T. Muller, H. Gemmeke, J. Reichenbach, and W. Kaiser. Model-based registration of x-ray mammograms and MR images of the female breast. *IEEE Transactions on Nuclear Science*, 53:204–211, 2006.
7. Tao Tan, Bram Platel, Michael Hicks, Ritse M. Mann, and Nico Karssemeijer. Finding lesion correspondences in different views of automated 3D breast ultrasound. In *Medical Imaging*, Proceedings of the SPIE, 2013.
8. Z. A. Taylor, O. Comas, M. Cheng, J. Passenger, D. J. Hawkes, D. Atkinson, and S. Ourselin. On modelling of anisotropic viscoelasticity for soft tissue simulation: numerical solution and gpu execution. *Med Image Anal*, 13(2):234–244, Apr 2009.
9. T. Tan, B. Platel, R. M. Mann, H. Huisman, and N. Karssemeijer. Chest wall segmentation in automated 3D breast ultrasound scans. *Medical Image Analysis*, 2013.
10. Tao Tan, Bram Platel, Roel Mus, Laszlo Tabar, Ritse Mann, and Nico Karssemeijer. Computer-aided detection of cancer in automated 3d breast ultrasound. *IEEE Transactions on Medical Imaging*, 2013.

ISBN: 978-87-996443-0-8

Contrast-Agent-Free MRI-guided Breast Biopsies Enabled by Breast Deformation Simulation

Markus T. Harz*[1], Suzan Akbey[1], Ritse M. Mann[2], Kathy Schilling[3], Joachim Georgii[1], and Horst K. Hahn[1]

[1] Fraunhofer MEVIS, Bremen, Germany
[2] Radboud University Nijmegen Medical Center, Nijmegen, NL
[3] Boca Raton Regional Hospital, Womens Center, Boca Raton, FL, U.S.A.

Abstract. Targeting lesions for biopsy in MRI-guided breast biopsy procedures relies on the administration of contrast agent and a dynamic contrast enhanced MRI scan. The index lesion will then be punctured with the biopsy needle, and a control scan reveals correct needle placement with respect to the targeted position. In the work we describe here, we propose to simulate the breast deformation from the diagnostic MRI scan to the MR biopsy preparation MRI taken before contrast agent administration. We describe the technology and setup, and provide a proof-of-concept on retrospective example data where the lesion was visible in the biopsy scan. The proposed method would allow to biopsy those 10% of lesions that would no longer enhance in the biopsy device. In addition, the procedure is faster, and by not requiring contrast agent, cheaper. We present a concept study with qualitative and quantitative evaluation on limited data.

1 Introduction

Breast biopsies are the gold standard decision making tool to determine the diagnosis and suggest the treatment of a suspicious lesion. In principle, the biopsy will be conducted using the imaging modality that displays the lesion first with a sufficient reliability, and accurately enough to determine the lesion extent. Among the most prominent biopsy techniques in breast cancer care are the following [2, 15, 5]:

Stereotactic biopsies. With two oblique mammographic acquisitions of the same breast under unchanged compression, a lesion position can be targeted with highest accuracy. This technique can be used both with core needle biopsy devices or vacuum-assisted needles, the latter being the method of choice. Mainly, small clusters of microcalcifications, indicative of early DCIS, are the target lesion class that is biopsied in this modality.

* markus.harz@mevis.fraunhofer.de

ISBN: 978-87-996443-0-8

Ultrasound-guided biopsies. Whenever a lesion is visible sonographically, hand-held ultrasound provides a preferred tool to visualize lesion and needle. The real-time control over needle position together with the monitoring of lesion displacement induced by the advancing needle tip are main factors for extremely low rates of failed procedures.

MRI-guided biopsies. There are cases where neither mammography nor ultrasound suffice to demonstrate the lesion in question. These cases arise when a suspicous finding from mammography and ultrasound remains indeterminate, and MRI is used as a decision making tool. Often, additional lesions display in MRI, or the extent of the finding is demonstrably larger than it appeared. Since in these cases, no other modality might guide the biopsy procedure, MRI is used for targeting when the patient returns for the procedure days or weeks after the MRI based diagnosis.

In this contribution, we focus on the MRI-guided breast biopsy procedure and suggest a workflow improvement enabled by deformation simulation that has the potential to save time and money, contribute to more patient comfort, and help to biopsy those 10% of lesions that currently cannot be biopsied because they don't show up in the biopsy preparation MRI. An approach similar to ours has to the best of our knowledge not been proposed before, though of course an extensive body of research has focussed on the prone-to-supine deformation simulation to help in surgery planning and open biopsies or to improve needle biopsy procedures through automation (see e.g. [6, 13, 1]). The difficulty of the approach is the largely unknown positioning change between the two visits, and the deformation of the breast between diagnostic and interventional MRI. A breast that is 87 mm wide in the diagnostic MRI is for example compressed to a width of 57 mm (measured in the axial slices), while it simultaneously expands from 112 mm to 142 mm in cranio-caudal direction.

2 Contribution

When diagnostic contrast-enhanced dynamic imaging shows a lesion that requires biopsy, the woman will be scheduled for a follow-up appointment to have the biopsy. For the MRI-guided biopsy, the woman is placed in prone positioning into the breast coil, which is equipped with a biopsy grid on the lateral side. The breast is slightly compressed to avoid as much target movement as possible during the 30 min procedure. A contrast-enhanced MRI sequence is acquired to visualize the index lesion again. Using the MRI biopsy planning software, the operator places a target marker in the three orthogonal projections. Afterwards, he can read off the grid position in the biopsy grid, as well as the depth to which the needle has to be inserted, alongside with the needle angulation in case the biopsy device allows for tilted access. Many parameters are taken into account, for example the length of the needle tip extending before the biopsy opening in the needle.

The procedure leads to reportedly about 10% of targeted lesions that are no longer visible on pre-interventional MRI [18, 4]. Although some reports state

ISBN: 978-87-996443-0-8

that those missed lesions are frequently benign and hormonal tissue alterations, they inevitably lead to a further short-term follow-up MR imaging, also inducing anxiety in the women. Also, particularly small lesions (<1 cm in diameter) are among those not recovered in pre-interventional MRI. On the other hand, such small cancers can be cured with almost certainty, so that a missed biopsy in these cases is particularly undesirable.

We explore the feasibility of a changed workflow that may help in these situations. It also speeds up the procedure by substituting the contrast-enhanced series with computational derivations of the target position. We propose to utilize deformation modeling as described above to simulate the effect of the compression induced by the breast coil on the breast tissue. The outline of this approach is the following:

Prior Preparation Segment the breast tissue and the target lesion from the contrast enhanced diagnostic MRI scan and generate the patient specific breast model breast tissue mask.

Immediate Preparation Segment the breast tissue from a contrast-free fast acquisition in the biopsy MRI using the biopsy coil. The compression by the biopsy device should be applied in these images. Note that those images are only utilized to estimate the target positions of the compression paddle of the biopsy device.

Alignment Apply a rigid registration between the diagnostic MRI and the biopsy MRI. Note that this step can easily be handled by any rigid registration algorithm.

Simulation Apply boundary conditions on the breast: Nodes on the chest are fixated, and large displacement from medial side towards lateral by the compression plate, and a light fixating displacement induced by the biopsy grid from the lateral side towards medial are modeled as dynamically updated displacement constraints. Afterwards, surface morphing accounts for the difference in surface shapes.

Biopsy Transfer target lesion position to the deformed state using a dense deformation field estimated from the vertex displacements, and visualize for biopsy planning.

If successful, this novel procedure not only helps to biopsy even those lesions going lost in the biopsy coil, but it also saves time through sparing the contrast-enhanced acquisition. It additionally saves costs by substituting the contrast agent administration by a computational method, and lastly, it contributes to patient comfort since no injection is required in the biopsy situation.

3 Methods

A prerequisite is to model the deformation of the breast from MRI in the prone position using breast coils to the compressed prone position used in MR-guided biopsy to provide the current lesion position, which can then be displayed. Our approach combines automatic segmentation and meshing methods with a fast,

ISBN: 978-87-996443-0-8

real-time capable deformation simulation. For the intended clinical setting, a compromise has to be made between fast calculations and highly realistic material behavior. Fast computations are mandatory in this scenario since in the intervention situation, the image of the compressed breast is acquired for the first time, such that while the bulk work – setting up the finite element model for the original breast MRI volume – may be computed in advance, the actual deformation simulation from the prone fixated diagnostic contrast-enhanced breast MRI to the interventional situation will be computed with the patient in the scanner. It hence has to be fast so that the benefits of the approach can be harvested.

3.1 Data and Model Generation

Standard T1-weighted non-fat-suppressed dynamic contrast enhanced bilateral breast MR images have been used of five patients who together underwent six MR-guided breast biopsy procedures. We excluded two patients from the presented results: one having large implants that are in our automated segmentation counted towards the fixed tissue types (though they deform considerably), and one patient with very small breasts where there is no real deformation visible between diagnostic and biopsy scan, and the automated setup fails to determine the fixation plates of the biopsy device.

The remaining four datasets were segmented into parts that in the model will be fixated or free. We consider the breast parenchyma and the adipose tissue to be deformable elastic, incompressible homogenous isotropic tissue. All vertices of the thorax were fixated. Our proposed system starts with MR images taken in the prone position that are automatically segmented into deformable and fixed tissues using the methods proposed by Anon [14]. A volumetric tetrahedral mesh was generated from a downsampled version of this data, tuned to deliver meshes consisting of at most 50k elements. Average downsampled voxel sizes were about 5 mm isotropic.

To define the breast compression model induced by the biopsy device, we now look at the breast tissue outer contour, with the boundary to the chest wall excluded. To avoid later possible misses in the detection of the compressing devices (the biopsy grid on the lateral side and the breast coil padding on the medial side), we crop the contour image by looking at the voxel counts in sagittal slices. Thinking of the voxel count plotted against the sagittal slice number, the resulting curve is supposed to show two peaks where the breast contour is roughly parallel to the sagittal plane. These peaks can easily be detected and used as a constraint for the search of the two planes. From this reduced mask image, we generate a point set that we thin out by iteratively removing every other point several times, until a suitable number of points remains.

These points are submitted to an implementation of the RANSAC algorithm [8] to find two parallel planes in the data that are roughly parallel to the sagittal viewing direction and have distance above a user-defined threshold (e.g. 6 cm). These planes are assumed to be infinite in the cranio-caudal direction, but limited posteriorly based on the detected size of the found plane.

ISBN: 978-87-996443-0-8

3.2 Deformation Simulation

A highly efficient FEM-based breast deformation approach is used to simulate the shape change from MRI to the current biopsy procedure positioning and compression. Our approach is based on a multigrid finite element framework developed by Georgii and Westermann [9], which efficiently simulates deformations of the breasts using the so-called co-rotated Cauchy strain formulation from Rankin and Brogan [16].

The deformation of a volumetric object is described by a displacement field $u(x)$, $u : \mathbb{R}^3 \to \mathbb{R}^3$; $x \in \mathbb{R}^3$, which maps the reference configuration Ω to the deformed configuration $\{x + u(x) \mid x \in \Omega\}$. Driven by external forces f, a deformed solid is governed by the well-known Lagrangian equation of motion, $M\ddot{u} + C\dot{u} + Ku = f$, where M, C, and K are respectively known as the mass, damping and stiffness matrices, u denotes the composition of the displacement vectors of all vertices, and f consists of the force vectors applied to these vertices. The stiffness matrix K is assembled from the so-called element stiffness matrices K^e. Typically, every element in a finite element discretization has only a very small number of neighbors, and thus the resulting stiffness matrix is very sparse. The element matrices are precomputed with a fixed elastic modulus E_0. Due to the linearity of the underlying material law, the element matrix of a particular element can then be obtained by scaling K^e by the stiffness value of this element relative to $E_0 \in \mathbb{R}$. Therefore, we can update the stiffness values within the assembling process at nearly no additional computational costs and thus achieve a fast update of stiffness values in the FE model analogously to previous approaches [7, 17]. To efficiently update the data structures of the numerical multigrid solver, we make use of a fast approach to compute sparse-sparse matrix products [10].

An implementation of a compression simulation including sliding on the plane surface governed by friction has been adapted to consider planes with finite extents. In this implementation, a collision detection probes if surface vertices of the breast volume mesh are – after one movement step of the compressing plane – behind the plane. If so, the vertex is projected onto the plane by applying a displacement constraint. After that, forces are calculated for all vertices on the plane to estimate their friction-governed sliding, which is again submitted to the vertex in the form of displacement constraints.

To align the surface resulting from the compression simulation with the known surface of the compressed breast in the biopsy device, we utilize the approach outlined in previous work by Harz et al. [12]. The resulting vertex displacements, describing a dense deformation field, are applied to the diagnostic MRI image for evaluation processes.

3.3 Evaluation

We evaluate the approach qualitatively by superposing the deformed diagnostic MRI (or a subtraction image of late and early phase) onto either the native

ISBN: 978-87-996443-0-8

intra-biopsy MRI or again a subtraction of enhanced and unenhanced intra-biopsy MRI. From this superposition, the general deformation quality was assessed visually. Next, landmark position differences are evaluated quantitatively. The landmarks are annotated by an experienced user, choosing locations that are easily recoverable even after application of the deformation in the biopsy device. The basis for the annotations were the native images of the series, where the delineation of parenchymal structures was easiest, and landmarks could be spotted in their morphology. It was attempted to place landmarks in all quadrants of the breast, and in most cases many more landmarks were explored to assure that the true accuracy is reflected by the report.

4 Results

For all cases, the simulation took about 2 min of preparation time (including automatic segmentation of the diagnostic MRI and setting up the volume mesh and matrices). These steps can be computed before the patient enters the MRI scanner for the biopsy procedure. Note that even the definition of the intended biopsy site could be done based on the diagnostic MRI already.

When the preparation scan in the scanner has been acquired, the manual registration, and the simulation including finding the compression plates, simulating the moving plates, and morphing the surfaces, took about 1 min altogether. Figure 1, top row, gives a visual impression of the shape of the deformed diagnostic MRI with respect to the true interventional MRI (right).

In two out of the four cases, a successful biopsy may potentially have been conducted based on the deformed diagnostic MRI (cases 1 and 3). In the first case, the lesion was very large, so that despite limited correspondence, the lesion was still largely corresponding between deformed diagnostic MRI and interventional scan. In case 3, the quadrant of the breast with the index lesion was registered between the scans with landmark correspondence errors averaging well below 10 mm. Two example landmark correspondences after deformation simulation of the diagnostic MRI into the biopsy MRI shape are shown in Fig. 1, bottom row. In the columns, the diagnostic MRI and the biopsy MRI are depicted, and the crosshair is automatically placed in the biopsy MRI based on the selected position in the diagnostic scan, transformed by the calculated deformation field. Table 4 shows our qualitative and quantitative assessments of the results, indicating that case 3 stands out in accuracy.

For the cases 2 and 4, if a biopsy would have been conducted based on the transformed diagnostic scan alone, there is a higher chance that the index lesion would have been missed, since even the cavity of a vacuum-assisted biopsy device is not usually more than 20 mm in diameter. We have

One of the two other cases failed due to the different nipple positioning in diagnostic and interventional scan that have not be accounted for in our approach, and the size of the breasts, causing large deformations (“Ssi”). In the other case (“GulR”), there is a systematic misalignment of about 10 mm predominantly in the transversal direction. One reason for this may again be a

ISBN: 978-87-996443-0-8

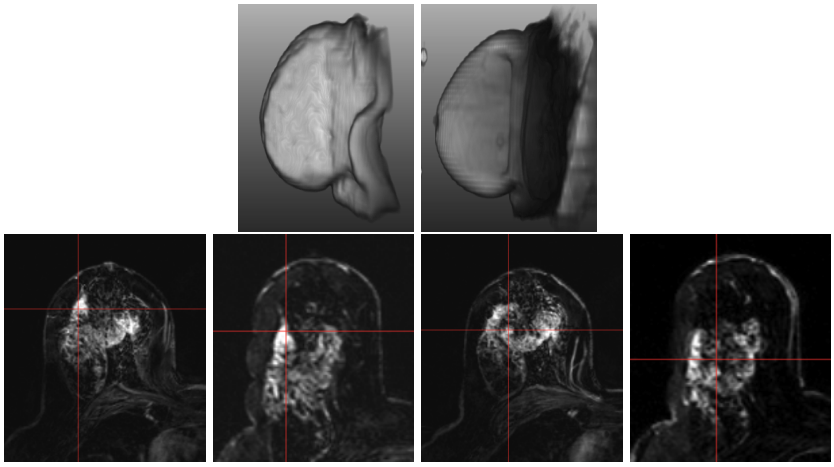


Fig. 1. Case 3: **Top row:** 3D visualizations show the surface shapes of the diagnostic MRI after deformation simulation (left), and the target shape reconstructed from the biopsy MRI. **Bottom row:** Two corresponding landmarks are shown. Left two images: first landmark, selected in diagnostic scan and correlated in biopsy scan utilizing the deformation field derived from the simulation. Right two images show the same combination for a second landmark, where the correspondence is slightly worse.

difference in the positioning of this particular breast, where the nipple is centrally placed in the medio-lateral direction, but caudal in the transversal direction.

5 Discussion and Conclusion

We acknowledge several potential improvements to setup and simulation. Firstly, in some cases the segmentation method is not optimal for the data, in particular, implants are considered not to be breast tissue, which is medically correct, but they should be modeled as a deformable tissue. Secondly, the deformation that is observed between diagnostic and pre-biopsy MRI indicated a more complex movement than what we currently model. In particular, the nipple position displacement is sometimes severe and not accounted for. This is one of the reasons why previous works that model breast deformation using the FE method reported better landmark correspondences (e.g. [11]). Also, it is evident that a linear-elastic strain formulation may for large deformations introduce unknown errors, that can to a certain extent be accounted for by the corotational formulation. On the other hand, the high performance of the implementation allows its usage in the described setting.

A second observation is that the automatic segmentation needs to be adjusted to include the pectoral muscle in the "deformable tissue" mask. A close examination of the diagnostic and biopsy MRI reveals that with the compression applied to the breast, the muscle also shifts and is compressed.

ISBN: 978-87-996443-0-8

Table 1. Qualitative and quantitative assessments. Four considered biopsies are described with respect to relevant qualitative evaluation criteria, in five subjective levels from – (poor) to ++ (excellent) with "o" indicating "indecisive". Four to seven landmarks were identified in both diagnostic and biospy MRI, and their distance (in mm) assessed after transformation in terms of average and standard deviation. **Key:** *QA*: MRI quality. *Pos*: Patient positioning. *Plane*: Plane detection performance. *Vis*: Visual inspection of results. *#L*: number of landmarks.

Case	QA	Pos	Plane	Vis	#L	Avg. dist.	Stdev	min-max
1	+	o	+	+	5	14.1	4.7	7.6 – 19.3
2	+	++	++	-	4	16.6	8.1	7.2 – 26.1
3	++	++	+	+	7	6.4	2.5	2.3 – 8.9
4	+	-	+	o	4	13.4	7.2	6.6 – 22.4

On our limited data, we currently assess the main origins of the observed misalignments. Work we are carrying out includes to utilize different non-linear material laws in the simulation to assess the influence of the material model on the results. Also, we substitute the last step of our approach (the surface morphing) by an elastic registration approach, e.g. as described in [3] to accomplish the registration on a fine level of detail. Both approaches have not formally been evaluated for this contribution, but first results indicate multiple salient future research directions.

References

1. Alderliesten, T., Loo, C., Paape, A., Muller, S., Rutgers, E., Peeters, M.J.V., Gilhuijs, K.: On the feasibility of MRI-guided navigation to demarcate breast cancer for breast-conserving surgery. Medical physics 37(6), 2617–2626 (2010)
2. Ames, V., Britton, P.D.: Stereotactically guided breast biopsy: a review. Insights into imaging 2(2), 171–176 (Apr 2011)
3. Boehler, T., Schilling, K., Bick, U., Hahn, H.: Deformable image registration of follow-up breast mr images. In: Proceedings of WBIR. Lecture Notes in Computer Science, vol. 2604, pp. 13–24. Springer, Lübeck (2010)
4. Brennan, S.B., Sung, J.S., Dershaw, D.D., Liberman, L., Morris, E.A.: Cancellation of MR Imaging-guided Breast Biopsy Due to Lesion Nonvisualization: Frequency and Follow-Up. Radiology 261(1), 92–99 (Sep 2011)
5. Bruening, W., Fontanarosa, J., Tipton, K., Treadwell, J.R., Launders, J., Schoelles, K.: Systematic review: comparative effectiveness of core-needle and open surgical biopsy to diagnose breast lesions. Annals of internal medicine 152(4), 238–246 (Feb 2010)
6. Carter, T., Tanner, C., Beechey-Newman, N., Barratt, D., Hawkes, D.: MR Navigated Breast Surgery: Method and Initial Clinical Experience. In: MICCAI '08: Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention, Part II. Springer-Verlag (Sep 2008)
7. Dick, C., Georgii, J., Burgkart, R., Westermann, R.: Computational steering for patient-specific implant planning in orthopedics. In: Proceedings of Visual Computing for Biomedicine 2008. pp. 83–92 (2008)

ISBN: 978-87-996443-0-8

8. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24(6) (Jun 1981)
9. Georgii, J., Westermann, R.: Corotated finite elements made fast and stable. In: Proceedings of the 5th Workshop On Virtual Reality Interaction and Physical Simulation. pp. 11–19 (2008)
10. Georgii, J., Westermann, R.: A streaming approach for sparse matrix products and its application in Galerkin multigrid methods. Electronic Transactions on Numerical Analysis, to appear (2010)
11. Han, L., Hipwell, J., Taylor, Z., Tanner, C., Ourselin, S., Hawkes, D.: Fast deformation simulation of breasts using gpu-based dynamic explicit finite element method. In: Mart, J., Oliver, A., Freixenet, J., Mart, R. (eds.) Digital Mammography, Lecture Notes in Computer Science, vol. 6136, pp. 728–735. Springer Berlin Heidelberg (2010)
12. Harz, M.T., Georgii, J., Wang, L., Schilling, K., Peitgen, H.O.: Efficient breast deformation simulation. In: Workshop on Virtual Reality Interaction and Physical Simulation. pp. 117–126. The Eurographics Association (2012)
13. Mallapragada, V., Sarkar, N., Podder, T.K.: Toward a Robot-Assisted Breast Intervention System. Mechatronics, IEEE/ASME Transactions on PP(99), 1–10 (2010)
14. Nonymous, A., Onymous, A.: Dummy title. In: Workshop on Anonymous Publications. pp. a–b. The Anonymous Association (2099)
15. O'Flynn, E.A.M., Wilson, A.R.M., Michell, M.J.: Image-guided breast biopsy: state-of-the-art. Clinical radiology 65(4), 259–270 (Apr 2010)
16. Rankin, C., Brogan, F.: An element-independent co-rotational procedure for the treatment of large rotations. ASME J. Pressure Vessel Tchn. 108, 165–174 (1986)
17. Schiwietz, T., Georgii, J., Westermann, R.: Freeform image. In: Proceedings of Pacific Graphics (2007)
18. Viehweg, P., Bernerth, T., Kiechle, M., Buchmann, J., Heinig, A., Koelbl, H., Laniado, M., Heywang-Kobrunner, S.H.: MR-guided intervention in women with a family history of breast cancer. European journal of radiology 57(1), 81–89 (Jan 2006)

ISBN: 978-87-996443-0-8

Standard Attenuation Rate and Volpara™ Volumetric Density Maps

Faraz Janan[1,1] , Sir Michael Brady[1], Christopher Tromans[2], Ralph Highnam[2]

[1] University of Oxford, UK (farazjanan@some.ox.ac.uk, mike.brady@oncology.ox.ac.uk)
[2] Matakina Ltd
(chris.tromans@matakina.com, ralph.highnam@matakina.com)

Abstract. We compare two quantitative volumetric measures that have been proposed to estimate the amount of dense breast tissue in mammograms, namely Standard Attenuation Rate (SAR) and Volpara®. Corresponding regions for a range of tissue types were analyzed to determine regression and linearity between them. In total, 156 regions from 7 pairs of density maps were used in this study. The regions were obtained using Simple Linear Iterative Clustering (SLIC), which decomposes an image into feature-based, visually homogeneous, k-means clusters, known as Superpixels. The means of corresponding Superpixels are then compared to test linearity. It is found that both volumetric measures are strongly correlated (R^2 = 0.9618). An average SAR per millimeter of Volpara® density gives an even higher regression coefficient (R^2 = 0.9875). This strong relationship reinforces the reliability of these volumetric measures for estimating breast density.

Keywords: Breast Density, Standard Attenuation Rate, Volpara®, Breast Cancer, Digital Mammography, x rays

1. Introduction

X-ray mammograms are used routinely in screening to detect abnormalities in the breast for an asymptomatic population of post-menopausal women, and to assess the qualitative density of breast regions. However, the study of mammograms currently relies on subjective assessment, which leads to large inter- and intra-observer variability. The appearance of a mammogram, considered as an image, varies very considerably among women and depends upon the imaging parameters, tissue characteristics, and the response of different breast tissues to x-ray attenuation. The need for reproducible breast measurements is evident from the fact that parenchymal patterns are subject to change over time, for a number of reasons. Therefore,

[1] *Note that the authors other than FJ are intimately involved with Matakina. Though these authors contributed to the paper, they* ***were not in any way involved in analyzing the results, which appear in FJ's forthcoming DPhil thesis.***

adfa, p. 1, 2011.

quantitative estimation of breast density which is independent of imaging conditions is required, not only to assess the risk of developing cancer but also to help in diagnosis and prognosis of the disease.

The Wolfe and BI-RADS classification models provide a basis for the assessment of automatic breast density measurement methods. These reporting standards, especially the latter, are widely used, but are highly subjective and numerous studies have shown that they have poor repeatability and inter-observer variation. Cohort studies exist that have attempted to describe parenchymal breast density quantitatively [1], [2]. However, all such methods have severe limitations. For example, Boyd et al [2]presented a computer-based interactive thresholding method that finds the percentage of segmented breast tissue relative to the segmented breast area. The method was semi-automatic and requires the user to provide a threshold value to segment the breast parenchyma. Petroudi et al. [3] normalized mammograms using Standard Mammographic Form (SMF) [4], [5] and used the volume of adipose tissue in the breast for quantification. Texture representations of breast tissue have also been used for quantification. For example, Byng et at attempted to analyze the breast automatically using fractal dimension and skewness [6]. Miller et al. classified breast tissue by applying Laws' texture features and granulometric techniques to discriminate between fatty and dense breast types. Bovis et al. also used a texture feature based model using artificial neural networks to categorize breast classes. Others have used mathematical morphology [7], density segmentation using Kittler's method [8] and variance histogram discriminant analysis of segmented regions [9]. Van Gils et al [10], [11] investigated whether a reduction in breast density lowers the risk of breast cancer, whereas Karssemeijer [12] developed an automated determination of breast parenchymal patterns based on a distance transform, which is applied along with segmentation between breast tissue and the skin line. Readily available tools such as CUMULUS[13] and MDEST [14] are more objective, subject to the constraint of being semi-automated; for this reason they have had limited use in clinical practice. This is because such area-based thresholding methods require additional decision time and suffer from inter-and-intra-observer variability along with the extensive training to use the software [15], [16].

In this paper, we compare two volumetric density estimation techniques for a mammographic screening population: Standard Attenuation Rate (SAR) [17] and Volpara®, commercial software provided to us by Matakina Ltd. Our primary goal is to evaluate consistency in volumetric measures by investigating the nature of their relationship, and to this end we show a strong linear association between their outputs. The results further reinforces the case for using such volumetric measures to assess risk of breast cancer with respect to mammographic density in clinical practice.

2. Methods

Digital Mammographic images of seven women aged 45-71 were processed using both the SAR and Volpara® systems. The mammograms were made available by Matakina Ltd, imaged using a GE DS FFDM scanner acquired during routine

ISBN: 978-87-996443-0-8

screening. At least two of the mammograms were suspected to contain abnormalities. SAR and Volpara® (V 1.4.3) generated density maps, though they have substantially different spatial resolutions and cover slightly different regions of the breast.

2.1. Standard Attenuation Rate

SAR [17], [18] is a normalized measure of the tissue radiodensity traversed by the primary beam incident upon each pixel of a mammogram, enabling estimation of breast density. It is an "absolute physics" model [19] which calculates the x-ray spectrum incident upon breast, the energy exiting the breast, and subtracts the estimated scattered radiation. SAR is estimated by scaling the primary attenuation of the breast to a suitable reference material (eg. 50:50 fat/fibroglandular, though the choice is made by the user). The radiodensity SAR is obtained by,

$$E_{recorded} = \sum_{\varepsilon} D_{absorbed}^{\left(e^{-SAR_{\mu}\, ref(\varepsilon)\, H}\, I_{incident}(\varepsilon),\ \varepsilon\right)}$$

where $E_{recorded}$ is the energy arising from the primary fluence, $D_{absorbed}$ describes the detector absorption for a photon fluence of magnitude x; ε is the photon fluence incident on the upper surface of the breast from the x-ray tube; $\mu\, ref\ (\varepsilon)$ is the linear attenuation coefficient of the reference material; and H is the thickness of the compressed breast tissue traversed. The multiplicative factor $SAR_{x,\,y}$ at each pixel is,

$$\begin{aligned} SAR_{x,y} &= m_{x,y} \ln\left(\frac{D^{-1}(I_{x,y}) - scat_{x,y}}{I_{incident,\ x,y}}\right) + C_{x,y} \\ &= m_{x,y} \ln\left(\frac{E_{recorded,\ primary,\ x,y}}{I_{incident,\ x,y}}\right) + C_{x,y} \end{aligned}$$

where $scat_{x,y}$ is the scatter fluence recorded by the detector; $m_{x,y}$ and $C_{x,y}$ depend on the spatial location (x,y); and D is the detector calibration transfer function relating a recorded pixel intensity i, to the total photon energy absorbed by the pixel detector from which the pixel intensity resulted. The coefficients of the linear mapping from $\ln(primary_{x,y})$ to $SAR_{x,y}$ are calculated from a number of simulated images using the image formation model.

2.2. Volpara ®

Volpara® embodies a "relative physics" [19] model and is commercially available software that is a core component of the European Union (EU) collaborative Adapting Breast Cancer Screening Strategy Using Personalized Risk Estimation (ASSURE) project. It is based on SMF and the volumetric breast density estimation model [20] by Van England et al. The latter method maps dense tissue by using a physical model of image acquisition based on the assumption that the breast is either composed of fat or parenchyma (dense tissue). The Volpara® model substantially reduces the dependence of breast quantification on imaging physics data. The fundamental

ISBN: 978-87-996443-0-8

assumption in the version we used is that the observed pixel image intensity in a mammogram has a linear relationship to the energy imparted to the x-ray detector.

Volpara® finds the linear attenuation coefficients μ for fat and dense tissue at a specific filter, tube voltage, and recorded breast thickness, area of the breast that is entirely fat P_{fat} and defines it as a reference intensity value to measure the thickness of dense tissue h_d at each pixel (x, y). The thickness of the dense tissue is computed as:

$$h_d\,(x,y) = \frac{\ln(\frac{P(x,y)}{P_{fat}})}{\mu_{fat} - \mu_{dense}}$$

The method uses phase congruency [21] and an approach to find the uncompressed breast edge [20] as well as the relative breast edge [5], which helps to locate an accurate breast edge and to calculate P_{fat}. The method is robust to errors in detector gain, multiplicative variations and exposure time as these parameters cancel out. To deal with different tissue types, the scatter removal process works in a relative manner. In the first published version, a fixed compression plate slant estimation is used that works for most imaging scanners that have parallel compression plates, however, in more recent versions it does not rely on the DICOM header and accommodates almost all non-parallel compression plate imaging systems. The breast volume is the product of recorded breast thickness and breast area, whereas their ratio gives the breast density. The volume of the dense tissue is the summation of all $h_d\,(x,y)$ over the image. Over 1.2M mammograms have been processed using Volpara® over the past year, identifying 3600 early cancers.

2.3. Simple Linear Iterative Clustering (SLIC)

To compare the two techniques, a direct pixel/pixel comparison cannot be applied because of the high variability in density values of neighboring pixels, particularly in bright regions; making it difficult to estimate a tissue value from a single pixel.

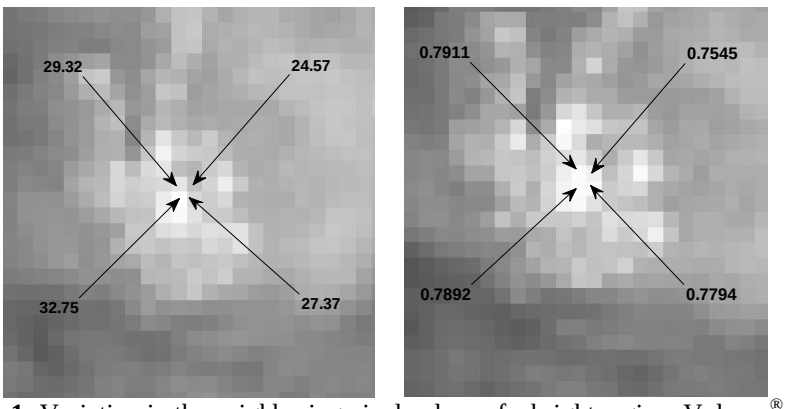


Figure 1: Variation in the neighboring pixel values of a bright region. Volpara® density in mm (left) and SAR attenuation (right)

ISBN: 978-87-996443-0-8

To address this issue, we applied Simple Linear Iterative Clustering (SLIC) [22], [23] to decompose the image into visually homogeneous regions. The method is based on a spatially localized version of k-means clustering where each pixel is associated with a feature vector and k-means clustering is then run on those vectors. The image is first divided into a grid of a certain region size and the center of each grid tile is used to initialize a corresponding k-means. Finally, the k-means centers and clusters are refined by using the Lloyd algorithm [24]. The standard Lloyd algorithm assigns pixels to the closest clusters and iteratively updates their centers. After k-means has converged, SLIC eliminates any connected region whose area is less than a minimum region size. This is done by greedily merging regions to neighboring ones.

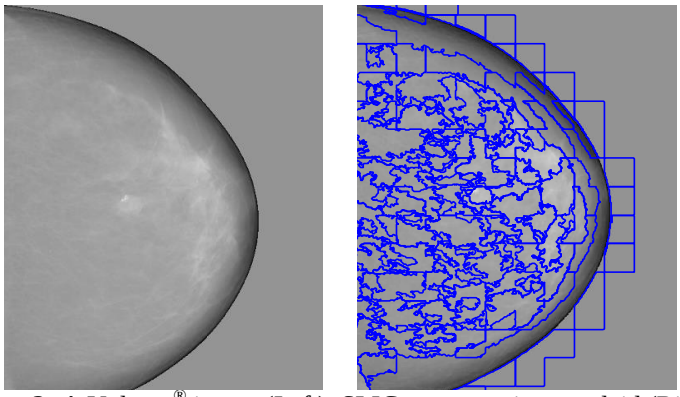


Figure 2: A Volpara® image (Left), SLIC segmentation overlaid (Right).

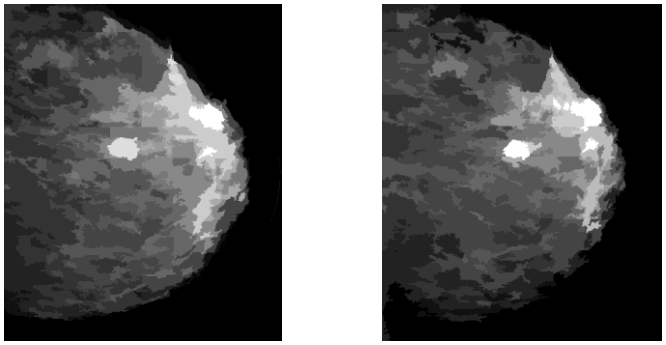


Figure 3: Mean density region map of Volpara® (left) and SAR (right) in Figure 1

We applied SLIC to both Volpara® and SAR images, yielding the segmentation overlay, while dividing the image into homogeneous regions or super-pixels. We then compute a mean region density map, where each superpixel now represents the mean density of that region. The spatially corresponding Superpixels with similar sizes related manually for regression analysis. Figure 2 shows a Volpara® image segmented

ISBN: 978-87-996443-0-8

using SLIC, while the mean region density maps of a SAR\Volpara image pair are given in Figure 3. Note that Volpara® and SAR generate different SLIC segmentations because of their intrinsic variations.

3. Results

Using the above procedure, 156 super-pixels from the 7 pairs of Volpara® and SAR mean regions density maps were manually placed in correspondence and compared to find a relationship between the two. A good linear relationship was found for the complete population with a regression coefficient R^2 = 0.9618. The mean SAR per Volpara® millimeter was calculated for a sample size of 1mm to 43mm, where an average for each complete mm was found, gives R^2 = 0.9875 (Figure 4 & 5). A SAR value of 0.78 is approximately what Volpara® assumes is fat (SAR_{Fat}), and a SAR value of 1.116 is approximately what Volpara® considers to be fibroglandular (SAR_{FG}). Both SAR_{Fat} and SAR_{FG} are empirical estimates. However, SAR is calculated from an exact carefully calibrated physics model, whereas Volpara® is based solely on estimates from the image.

$$SAR_{Weighted} = \frac{(SAR - SAR_{Fat})}{(SAR_{FG} - SAR_{Fat})} = \%\ \text{fibroglandular in Volpara's output}$$

It is the percentage of fibroglandular that is of greater importance to clinicians and amounts to breast density. From the regression model,

$$Vol_{Estimated} = 35.608 * SAR_{Weighted} + 1.4251$$

$Vol_{Estimated}$ is an approximation of Volpara® image from linearly weighted SAR.

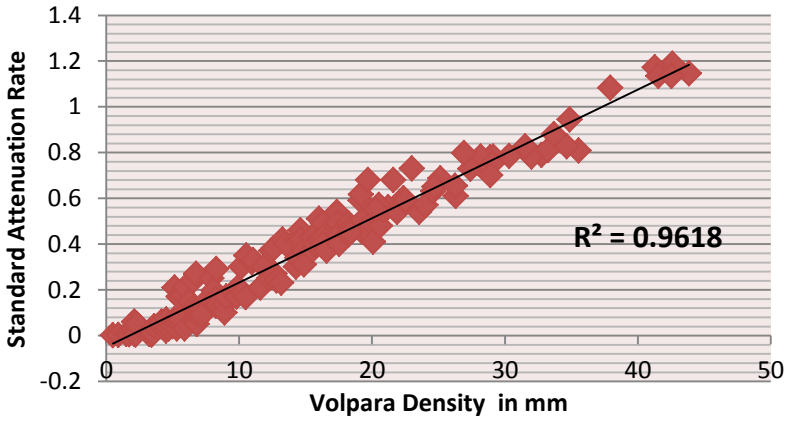


Figure 4: SAR-Volpara®- scatter graph for absolute sample values

ISBN: 978-87-996443-0-8

4. Conclusions

We have established a strong relationship between the commercially-available, (FDA-approved) fully-automated relative physics software Volpara® and the recently introduced absolute physics breast density metric SAR. The high correlations reported above argue that it is possible to compute volumetric estimates of breast density with high accuracy and high robustness. Further work is underway to validate these findings with a larger dataset and samples from tissue types. There is a need now to further investigate this relationship and to fully understand the relation between the breast cancer risk and density estimating measures.

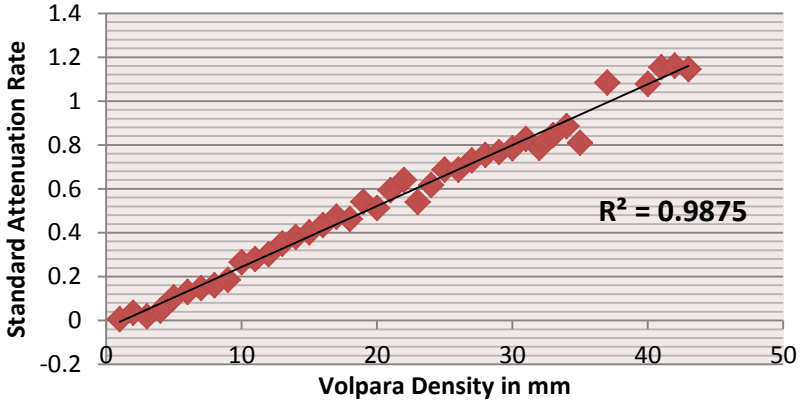


Figure 5: Mean SAR per millimeter of Volpara® density

References

[1] A. F. Saftlas, R. N. Hoover, L. A. Brinton, M. Szklo, D. R. Olson, M. Salane, and J. N. Wolfe, "Mammographic densities and risk of breast cancer," *Cancer*, vol. 67, pp. 2833–2838, 1991.

[2] C. Byrne, C. Schairer, J. Wolfe, N. Parekh, M. Salane, L. A. Brinton, R. Hoover, and R. Haile, "Mammographic features and breast cancer risk: effects with time, age, and menopause status," *J. Natl. Cancer Inst.*, vol. 87, pp. 1622–1629, 1995.

[3] S. Petroudi, K. Marias, R. English, R. Adams, and M. Brady, "Classification of mammogram patterns using area measurements and the standard mammogram form (SMF)," *Med. Imaging Underst. Anal.*, pp. 197–200, 2002.

[4] R. Highnam, X. Pan, R. Warren, M. Jeffreys, G. D. Smith, and M. Brady, "Breast composition measurements using retrospective standard mammogram form (SMF)," *Phys. Med. Biol.*, vol. 51, 2006.

[5] R. Highnam and J. M. Brady, *Mammographic image analysis*, vol. 14. Springer, 1999.

[6] J. W. Byng, M. J. Yaffe, G. A. Lockwood, L. E. Little, D. L. Tritchler, and N. F. Boyd, "Automated analysis of mammographic densities and breast carcinoma risk," *Cancer*, vol. 80, pp. 66–74, 1997.

[7] Y.-H. Chang, X.-H. Wang, L. A. Hardesty, T. S. Chang, W. R. Poller, W. F. Good, and D. Gur, "Computerized assessment of tissue composition on digitized mammograms,"

ISBN: 978-87-996443-0-8

Acad. Radiol., vol. 9, pp. 899–905, 2002.

[8] T. Matsubara, D. Yamazaki, M. Kato, T. Hara, H. Fujita, T. Iwase, and T. Endo, "An automated classification scheme for mammograms based on amount and distribution of fibroglandular breast tissue density," in *International Congress Series*, 2001, vol. 1230, pp. 545–552.

[9] R. Sivaramakrishna, N. A. Obuchowski, W. A. Chilcote, and K. A. Powell, "Automatic segmentation of mammographic density," *Acad. Radiol.*, vol. 8, pp. 250–256, 2001.

[10] C. H. Van Gils, J. H. C. L. Hendriks, R. Holland, N. Karssemeijer, J. D. M. Otten, H. Straatman, and A. L. M. Verbeek, "Changes in mammographic breast density and concomitant changes in breast cancer risk," *Eur. J. Cancer Prev.*, vol. 8, pp. 509–516, 1999.

[11] C. H. Van Gils, J. H. C. L. Hendriks, J. D. M. Otten, R. Holland, and A. L. M. Verbeek, "Parity and mammographic breast density in relation to breast cancer risk: indication of interaction," *Eur. J. Cancer Prev.*, vol. 9, pp. 105–112, 2000.

[12] N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms," *Phys. Med. Biol.*, vol. 43, 1998.

[13] J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe, "The quantitative analysis of mammographic densities," *Phys. Med. Biol.*, vol. 39, 1994.

[14] N. F. Boyd, L. J. Martin, M. Bronskill, M. J. Yaffe, N. Duric, and S. Minkin, "Breast tissue composition and susceptibility to breast cancer," *J. Natl. Cancer Inst.*, vol. 102, pp. 1224–1237, 2010.

[15] B. T. Nicholson, A. P. LoRusso, M. Smolkin, V. E. Bovbjerg, G. R. Petroni, and J. A. Harvey, "Accuracy of assigned BI-RADS breast density category definitions," *Acad. Radiol.*, vol. 13, pp. 1143–1149, 2006.

[16] K. E. Martin, M. A. Helvie, C. Zhou, M. A. Roubidoux, J. E. Bailey, C. Paramagul, C. E. Blane, K. A. Klein, S. S. Sonnad, and H.-P. Chan, "Mammographic Density Measured with Quantitative Computer-aided Method: Comparison with Radiologists' Estimates and BI-RADS Categories1," *Radiology*, vol. 240, pp. 656–665, 2006.

[17] C. E. Tromans, M. R. Cocker, and M. Brady, "Quantification and normalization of x-ray mammograms," *Phys. Med. Biol.*, vol. 57, 2012.

[18] C. Tromans and M. Brady, "An alternative approach to measuring volumetric mammographic breast density," in *Digital Mammography*, Springer, 2006, pp. 26–33.

[19] R. Highnam, M. Brady, M. J. Yaffe, N. Karssemeijer, and J. Harvey, "Robust breast composition measurement-volparaTM," in *Digital Mammography*, Springer, 2010, pp. 342–349.

[20] S. van Engeland, P. R. Snoeren, H. Huisman, C. Boetes, and N. Karssemeijer, "Volumetric breast density estimation from full-field digital mammograms," *Med. Imaging, IEEE Trans.*, vol. 25, pp. 273–282, 2006.

[21] P. Kovesi, "Image features from phase congruency," *VIDERE: J. Comput. Vis. Res.*, vol. 1, pp. 1–26, 1999.

[22] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels," *École Polytech. Fédéral de Lausssanne*, vol. 149300, 2010.

[23] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in *Proceedings of the international conference on Multimedia*, 2010, pp. 1469–1472.

[24] S. Lloyd, "Least squares quantization in PCM," *Inf. Theory, IEEE Trans.*, vol. 28, pp. 129–137, 1982.

ISBN: 978-87-996443-0-8

Effect of Reference Image Retrieval on Breast Mass Classification Performance: ROC Analysis

Chisako Muramatsu[1], Tokiko Endo[2,3], Mikinao Oiwa[2], Misaki Shiraiwa[3], Kunio Doi[4,5], Hiroshi Fujita[1]

[1] Department of Intelligent Image Information, Gifu University, Gifu, Japan
{chisa, fujita}@fjt.info.gifu-u.ac.jp
[2] Department of Radiology, Nagoya Medical Center, Nagoya, Japan
endot@nnh.hosp.go.jp
[3] Department of Advanced Diagnosis, Nagoya Medical Center, Nagoya, Japan
[4] Gunma Prefectural Collage of Health Sciences, Maehashi, Japan
[5] Department of Radiology, The University of Chicago, Chicago, USA
k-doi@uchicago.edu

Abstract. Retrieval of similar cases with the diagnostic and therapeutic results as a reference may be useful in differential diagnosis of abnormalities. Image retrieval method for breast masses on mammograms has been investigated in our previous study, and the result indicated the potential advantage of a machine learning technique using sample cases with experts‘ subjective similarity data. In this study, the effect of presenting reference images to observers‘ ablity to distinguish between benign and malignant masses was investigated. Eleven physicians and 11 radiological technologists evaluated 98 masses and recorded their confidence of a lesion being malignant without and with reference images. The areas under the receiver operating characteristic curves improved from 0.926 to 0.938 (p=0.17) and from 0.895 to 0.928 (p=0.004) for the physician and technologist groups, respectively. The results indicate that reference images may be useful for diagnosis of breast masses.

Keywords: similar image retrieval; breast masses; digital mammograms; differential diagnosis; observer study

1 Introduction

Breast cancer is the most frequently diagnosed cancer in women in the U.S., some European countries, and Japan [1-3]. To reduce the number of death from breast cancer and to improve patients‘ quality of life, early detection and proper treatment are important. Periodic screening with mammography is considered effective for the early detection for women with normal risk [4-6].

When a new lesion is found on mammograms, it is generally evaluated with other image modalities. Even in such situation, it is beneficial to thoroughly evaluate mammographic findings to compare with those in other modalities. However, it is not easy to make differential diagnosis of lesions on mammograms. It has been suggested

ISBN: 978-87-996443-0-8

that computer-aided diagnosis (CAD) that provide the likelihood of malignancy of lesions may be useful in improving radiologists‘ diagnostic accuracies in the observer performance studies [7-9]. On the other hand, numeric guide may not be sufficient to some radiologists. Compared to computer-aided detection in which radiologists are prompted the suspicious areas on mammograms, the likelihood generally lacks in providing evidence to radiologists. Since radiologists‘ diagnostic ablity is based on experience, retrieval of similar images as a reference may be beneficial in providing supplemental information.

Several research groups have investigated the methods for automatic selection of similar images on mammograms and breast ultrasound images [10-19]. In earlier studies, the selection was based on the simple distance measures in the feature space [11], [13-14], [17]. In more recent studies, machine learning methods using samples with subjective similarity data were investigated, and the results were evaluated subjectively [15-16]. Other groups proposed specific decision making methods such as a 2-step selection method [18] and decision tree methods [19].

We have previouly investigated a similarity determination method using an artificial neural network (ANN), in which feature vectors of pairs of masses and the corresponding average subjective similarity ratings by experts were employed as input data and teacher, respectively [15]. In our recent study, a new method using multidimensional scaling (MDS) was introduced for understanding subjective similarity relationship between masses with different pathologies and for visualizing the subjective similarity space [20]. The objective similarity measure was determined on the basis of the distance in reconstructed subjective space using linear regression model. By leave-one-out cross validation test, the result indicated the usefulness of the MDS-based method.

In the present study, instead of applying ANN directly to estimate subjective similarity rating, MDS was first employed to map each image in subjective space, and the configured space was modeled by ANN [21]. After proper weights were determined, test cases were mapped in the modeled space, and the similarity between masses were determined by the distance in the similarity space. Using the proposed method, the effect of retrieved images on the readers‘ abilities to distinguish between benign and malignant masses was investigated in an observer performance study.

2 Materials and Methods

2.1 Mass Database

Digital mammograms used in this study were obtained at the National Hospital Organization, Nagoya Medical Center, Nagoya, Japan. The study protocol was approved by the institutional review board. The images were obtained with three digital sytems, including phase contrast mammography (PCM) system (Mermaid or Pureview, Konica Minolta Holdings, Inc.), direct conversion digital mammography system (Amulet, Fujifilm Corporation), and computed radiography systems (Mammomat 3000, Siemens, with C-Plate, Knoica Minolta, or Profect, Fujifilm). The original images have pixel sizes of 25 (PCM), 43.75 (C-Plate), or 50 (Amulet and

ISBN: 978-87-996443-0-8

Profect) μm and grayscales of 10 (Profect), 12 (PCM and C-Plate), or 14 (Amulet) bits. For computational purposes, the pixel size and grayscale were unified to 50 μm and 10 bits, respectively.

Two radiologists reviewed the images and identified the masses by placing square regions of interest (ROIs) on the basis of the radiologic and pathologic reports. The ROIs were extracted from both craniocaudal (CC) and mediolateral oblique (MLO) views. When a lesion was partially cut off by the field of view, the ROI was excluded in this study. The size of the ROIs varied from 168 x 168 to 1888 x 1888. In this study, masses with 9 pathologic types were included: ductal carcinoma in situ (DCIS), invasive lobular carcinoma (ILC), mucinous carcinoma (MC), papillo-tubular carcinoma (PTC), scirrhous carcinoma (SC), solid-tubular carcinoma (STC), cyst, fibroadenoma (FA), and benign phyllodes tumor (BPT). PTC, SC, and STC are the subtypes of invasive ductal carcinomas, and invasive ductal carcinomas with unknown subcategories were not included in this study. The numbers of ROIs and lesions, and their mean effective diameters for the 9 types are listed in Table 1. The fractions of images obtained by different mammographic systems were 39% (Amulet), 23% (Profect), 21% (PCM), and 17% (C-Plate). All the malignant masses were confirmed by biopsy and/or surgery and benign masses were confirmed by biopsy or follow-up by mammography and ultrasonography.

Table 1. The numbers of ROIs and lesions and their mean effective diameters for 9 pathologic types used in this study

Pathologic type	Number of ROIs	Number of lesions	Mean effective diameter (mm)
Ductal carinoma in situ (DCIS)	14	10	24±13
Invasive lobular carcinoma (ILC)	12	7	34±9
Mucinous carcinoma (MC)	9	6	32±13
Papillotubular carcinoma (PTC)	39	21	27±12
Scirrhous carcinoma (SC)	69	40	35±13
Solid-tubular carcinoma	38	22	47±20
Cyst	99	63	25±18
Fibroadenoma (FA)	90	58	29±11
Benign phyllodes tumor	8	6	55±32
Total	378	233	28±16

2.2 Method for Similarity Determination

Determination of Subjective Similarity. For employing the MDS, in general, similarity (dissimilarity) data for all paired combinations of subjects must be obtained. To include a large variety of cases, but also retaining the number of comparisons by experts reasonably small, three masses from each of the 9 pathologic groups, thus a total of 27 masses, were sampled. As the result, subjective similarity ratings for all possible 351 pairs of masses were obtained independently by 8 physicians who have been

ISBN: 978-87-996443-0-8

certified for reading mammography by the Central Committee on Quality Control of Mammographic Screening in Japan. The details of the method and the analysis of the results have been described elsewhere [22]. Briefly, 2 ROIs for comparison was displayed in one monitor, and their entire views of the breast were displayed in another monitor. Each physician was asked to rate the similarity of the pair on a continuous scale from dissimilar (0.0) to similar (1.0) based on the overall impression for the shape, density and margin by taking into account the predicted pathology types. We asked them not to weigh on the size of the lesions, the surrounding normal tissue, and unrelated calcifications. The average subjective ratings were considered as the gold standard of similarity and used in MDS analysis.

Determination of Objective Similarity. Kruskal's nonmetric MDS in R programming language was employed. The configuration dimension was set to 3 in this study to reduce a risk of overtraining. For determination of ANN parameters, i.e., numbers of hidden units and iterations, a leave-one-out cross-validation was employed in a series of MDS analysis and configuration modeling. In this process, one ROI was removed, and MDS was applied to remaining 26 ROIs. Once subjective space was constructed, each dimension was modeled by ANN with 13 image features. These features were defined elsewhere [21]. The test ROI was then mapped to the reconstructed space by trained ANN. This process was repeated for all 27 ROIs. After all 27 ROIs were mapped, distances between all pairs of ROIs were determined, and they were converted to similarity measures by use of an exponential function. The adequacy of the model was evaluated by the correlation between the gold standard and the similarity measures.

2.3 Selection of Similar Images for Test Dataset

Usefulness of the similarity measure for selection of reference images was evaluated by precision, *P*, which is defined as

$$P = \frac{Number\ of\ images\ with\ matched\ pathology}{Number\ of\ retrieved\ images} \quad (1)$$

For the testing, MDS was applied to 27 ROIs, and the configured similarity space was modeled by ANN with the parameters selected by the leave-one-out regime. After excluding all the ROIs belonging to the same cases as the 27 ROIs, 324 ROIs were used for evaluation. For each test ROI, the most similar cases were selected from the 324 ROIs excluding the ones of the same case. If a query mass and the selected mass are both benign or both malignant, it was counted for matched pathology. The number of retrieved images was varied from 1 to 10 images, and the average precision was determined.

ISBN: 978-87-996443-0-8

2.4 Observer Study

The effect of providing reference images in the diagnosis of breast masses was evaluated in the observer performance study. Ninety-eight cases, including 48 benign masses and 50 malignant masses, were randomly selected and included in the study. Eleven physicians who have been certified for reading mammograms and eleven radiological technologists who have been certified for mammography imaging with training of reading participated. In the reading session, a test ROI was displayed in one monitor, and the corresponding bilateral mammograms were displayed in another monitor. By clicking a button, images were switched between CC and MLO views on both monitors. After reviewing both views, observers were asked to mark their ratings on a continuous rating scale from definitely benign to definitely malignant. Subsequently, 5 reference images with their known pathologic types and the similarity map were shown, and the observers were asked again to mark their ratings. The results were evaluated by the multi-reader multi-case (MRMC) receiver operating characteristic (ROC) analysis [23] (MRMC software, the University of Chicago [24]). Figure 1 shows an observer study interface when reference images and similarity map were presented.

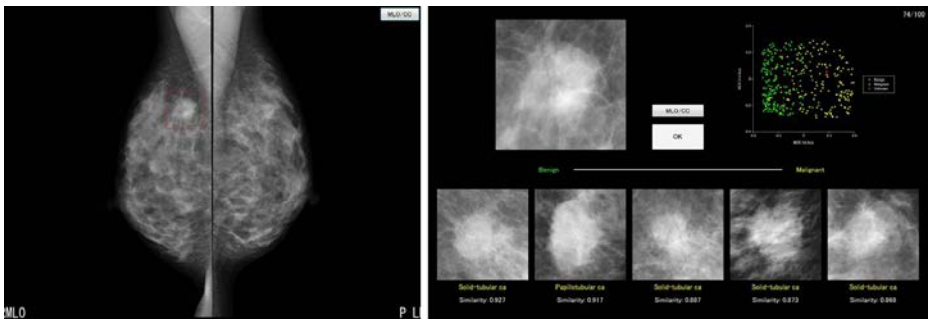


Fig. 1. Observer study interface with reference images. MLO views of the test case are shown in the left monitor, and the test ROI is shown in the right monitor with five reference images and their known pathologic types below and the similarity map on the top right.

3 Results

By the leave-one-out cross validation, similarity measures using MDS were determined for 351 pairs. The correlation coefficient between the gold standard and MDS-based similarity measures was 0.76, when the MDS dimension was 3. For comparison, the correlation coefficient between the gold standard and our previous ANN-based similarity measures for the 351 pairs was 0.68.

The method was applied to 324 test cases, and average precisions for the benign and malignant query images in retrieval of 1 to 10 images were determined and shown in Fig. 2. The average precisions were about 80% for both benign and malignant query images when 1 to 10 images were retrieved by the MDS-based measures. The result is relatively good as four of five reference images, on average, would be

ISBN: 978-87-996443-0-8

retrieved from the same benignity/malignancy groups as the test case. For comparison, the average precisions by the previous method were about 70% when 4 to 10 images were retrieved. The precision was slightly lower for the benign query images when small numbers of cases were retrieved. The proposed method was superior all across more than 100 retrieved images, although such large volume is impractical.

The observers' ability to distinguish between benign and malignant masses without and with reference images was evaluated in the observer performance study. By MRMC-ROC analysis, the average areas under the curves (AUCs) were 0.926 and 0.938 without and with reference images, respectively, for the physicians and 0.895 and 0.928, respectively, for the technologists. On average, AUCs for both groups were slightly improved. The difference was found to be statistically significant for the technologists ($p=0.004$), although we failed to find the statistically significant difference for the physicians ($p=0.17$).

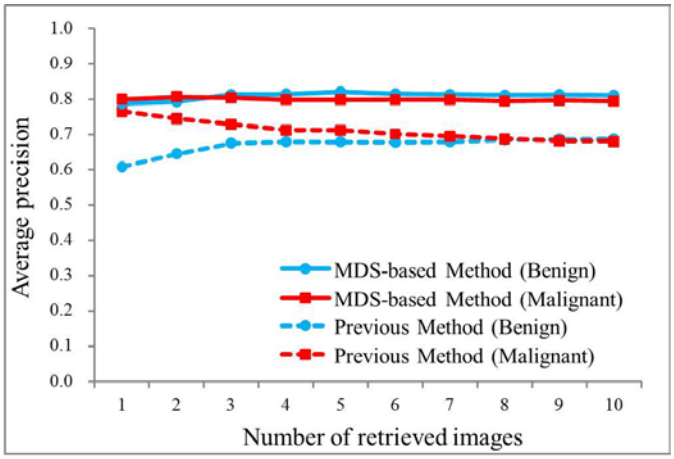


Fig. 2. Average precisions in retrieving pathology-matched reference images by the MDS-based similarity measures and previous ANN-based measures

4 Discussion and Conclusion

We have been investigating an effective image retrieval method to select images that are visually similar and useful in the point of view of diagnosis. Our new similarity measures are determined by applying MDS to the subjective similarity ratings obtained by experts for constructing a subjective similarity space and employing ANN to estimate the space with the image features. In this study, the proposed method was applied to the test cases, and the result was evaluated by precision in selecting pathology-matched reference images. When 1 to 10 images were retrieved, the majority of the cases (80%) were from the same pathologic group as the query images.

The effect of presenting reference images was evaluated in the observer study, in which observers' ability in distinguishing between benign and malignant masses was tested without and with the reference images. The average AUCs for both physician

ISBN: 978-87-996443-0-8

and technologist groups were slightly improved by showing reference images. Because the cases in this study were selected randomly and the observers in both groups were well trained, the AUCs were very high and the improvement was rather small. The technologists had the tendency to be slightly less confident at the initial reading and more likely influenced by the reference images. With the reference images, their average AUC was comparable to that of the physicians without reference images.

In conclusion, presentation of reference images may be useful in the diagnosis of breast masses on mammograms, especially for less experienced readers and slightly difficult cases. Our new similarity measures based on MDS may be effective in selecting useful reference images.

Acknowledgment

This study was partly supported by the Grants-in-Aid for Scientific Research for Young Scientists (No. 21791179) by Japan Society for the Promotion of Science and by a Grant-in-Aid for Scientific Research on Innovative Areas (No. 21103004), MEXT, Japan. Authors are grateful to following people for their contribution in this study and for participation in the observer study: K. Nishimura, T. Takahashi, T. Horiba, MD, M. Kato, MD, N. Kobayashi, MD., T. Morita, MD, M. Nishikawa MD, T. Niwa, MD, M. Sassa, MD, H. Yamamoto, MD, R. Suzuki, MD, Y. Hayashi, MD, N. Shinohara, J. Azuma, A. Imoto, J. Kikukawa, M. Yamada, Y. Suwabe, I. Morishita, A. Adachi, R. Kasugai, E. Katagiri, M. Tanaka, and E. Tukasaki.

References

1. Americal Cancer Society: Cancer Facts & Figures 2012. American Cancer Society, Atlanta (2012)
2. Ferlay, J., Autier, P., Boniol, M., Heanue, M., Colombet, M., Boyle, P.: Estimates of the cancer incidence and mortality in Europe in 2006. Ann. Oncol. 18, 581-592 (2007)
3. Matsuda, T., Marugame, T., Kamo, K.I., Katanoda, K., Ajiki, W., Sobue, T.: The Japan Cancer Surveillance Research Group. Cancer incidence and incidence rates in Japan in 2006: Based on data from 15 population-based cancer registries in the Monitoring of Cancer Incidence in Japan (MCIJ) Project. Jpn. J. Clin. Oncol. 42, 139-147 (2001)
4. Tabar, L., Fagerberg, G., Duffy, S.W., Day, N.E., Gad, A., Grontoft, O.: Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol. Clin. North Am. 30, 187-210 (1992)
5. Shapiro, S., Venet, W., Strax, P., Venet, L., Roeser, R.: Selection, follow-up, and analysis in the health insurance plan study: A randomized trial with breast cancer screening. J. Natl. Cancer Inst. Monogr. 67, 65-74 (1985)
6. Humphrey, L.L., Helfand, M., Chan, B.K.S., Woolf, S.H.: Breast cancer screening: A summary of the evidence for the U.S. preventive services task force. Annals. Internal Medicine. 137, E-347-367 (2002)
7. Chan, H.P., Sahiner, B., Roubidoux, M.A., Wilson, T.E., Adler, D.D., Paramagul, C., Newman, J.S., Sanjay-Gopal, S.: Improvement of radiologists' characterization of

mammographic masses by using computer-aided diagnosis: an ROC study. Radiology. 212, 817-827 (1999)
8. Huo, Z., Giger, M.L., Vyborny, C.J., Metz, C.E.: Breast cancer: effectiveness of computer-aided diagnosis – observer study with independent database of mammograms. Radiology 224, 560-568 (2002)
9. Jiang, Y., Nishikawa, R.M., Schmidt, R.A., Metz, C.E., Giger, M.L., Doi, K.: Improving breast cancer diagnosis with computer-aided diagnosis. Acad. Radiol. 6, 22-33 (1999)
10. Swett, H.A., Mutalik, P.G., Neklesa, V.P., Horvath, L., Lee, C., Richter, J., Tocino, I., Fisher, P.: Voice-activated retrieval of mammography reference images. J. Digit. Imaging. 11, 65-73 (1998)
11. Qi, H., Snyder, W.E.: Content-based image retrieval in picture archiving and communications systems. J. Digit. Imaging. 12, 81-83 (1999)
12. Sklansky, J., Tao, E.Y., Bazargan, M., Ornes, C.J., Murchison, R.C., Teklehaimanot, S.: Computer-aided, case-based diagnosis of mammographic regions of interest containing microcalcifications. Acad. Radiol. 7, 395-405 (2000)
13. Alto, H., Rangayyan R.M., Desautels, J.E.L.: Content-based retrieval and analysis of mammographic masses. J. Electron. Imaging. 14, 1-17 (2005)
14. Horsch, K., Giger, M.L., Vyborny, C.J., Lan, L., Mendelson, E.B., Hendrick, E.R.: Classification of breast lesions with multimodality computer-aided diagnosis: observer study results on an independent clinical data set. Radiology. 240, 357-368 (2006)
15. Muramatsu, C., Li, Q., Schmidt, R.A., Shiraishi, J., Doi, K.: Determination of similarity measures for pairs of mass lesions on mammograms by use of BI-RADS lesion descriptors and image features. Acad. Radiol. 16, 443-449 (2009)
16. Oh, J.H., Yang, Y., El-Naqa, I.: Adaptive learning for relevance feedback: application to digital mammography. Med. Phys 37, 4432-4444 (2010)
17. Xu, J., Faruque, J., Beaulieu, C.F., Rubin, D., Napel, S.: A comprehensive descriptor of shape: method and application to content-based retrieval of similar appearing lesions in medical images. J. Digit. Imaging. 25, 121-128 (2012)
18. Wang, X., Li, L., Liu, W. Xu, W., Lederman, D., Zheng, B.: An interactive system for computer-aided diagnosis of breast masses. J. Digit. Imaging. 25, 570-579 (2012)
19. Cho, H., Hadjiiski, L., Sahiner, B., Chan, H.P., Helvie, M., Paramagul, C., Nees, A.V.: A similarity study of content-based image retrieval system for breast cancer using decision tree. Med. Phys. 40, 012901-1-13 (2013)
20. Muramatsu, C., Nishimura, K., endo, T., Oiwa, M., Shiraiwa, M., Doi, K., Fujita, H.: Represenattion of lesion similarity by use of multidimensional scaling for breast masses on mammograms. J. Digit. Imaging. DOI: 10.1007/s10278-012-9569-0 (2013)
21. Nishimura, K., Muramatsu, C., Oiwa, M., Shiraiwa, M., Endo, T., Doi, K., Fujita, H.: Psychophysical similarity measure based on multi-dimensional scaling for retrieval of similar images of breast masses on mammograms. Proc. of SPIE Med. Imaging. (2013)
22. Muramatsu, C., Nishimura, K., Endo, T., Oiwa M., Shiraiwa, M., Doi, K., Fujita, H.: Correspondence among subjective and objective similarities and pathologic types of breast masses on digital mammography. In: Maidment, A.D.A., Bakic, P.R., Gavenonis, S. (eds.) IWDM 2012. LNCS, vol. 7361, pp. 450-457. Springer, Heidelberg (2012)
23. Dorfman, D.D., Berbaum. K.S., Metz. C.E.: Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. Invest. Radiol. 27, 723-731 (1992)
24. Available at http://metz-roc.uchicago.edu/

ISBN: 978-87-996443-0-8

Predicting False-Positive Biopsy Risk from Digital Mammography Using Locally-Adaptive Parenchymal Texture Analysis

Jae Young Choi, Brad M. Keller, Emily F. Conant, and Despina Kontos

University of Pennsylvania, Department of Radiology, Philadelphia, PA 19104, USA
Despina.Kontos@uphs.upenn.edu

Abstract. We evaluate a prediction model based on computer-extracted parenchymal tissue complexity features for estimating the risk of false-positive (FP) biopsy after diagnostic evaluation from screening digital mammography. We proposed a novel locally-adaptive parenchymal texture analysis algorithm to identify and extract mammographic features of local parenchymal tissue complexity potentially relevant for the prediction of false-positive biopsy. Our algorithm has the following innovative aspects: (1) the adaptive nature of automatically determining the optimal number of region-of-interests (ROIs) and the size of each ROI, based on the parenchymal tissue portion and distribution of the whole breast region and (2) the advantage of characterizing both the local and global mammographic appearance of the parenchymal tissue, aiming to provide more discriminative information for FP biopsy risk prediction. Results show that our locally-adaptive texture analysis algorithm in conjunction with a classification model can predict the likelihood of false-positive biopsy, with an ROC AUC up to 0.95 when considering a setting of ensemble classifiers.

Keywords: Localized and adaptive texture analysis, breast parenchymal tissue complexity, false-positive biopsy risk, prediction model, digital mammography

1 Introduction

In the process of mammography screening and diagnosis, false-positive (FP) mammography results, including FP recall or biopsy recommendation, are common. In the study [1], it has been found that after a decade of annual screening, a majority of women receive at least one false-positive result, and 7% to 9% receive a false-positive biopsy recommendation. In particular, false-positive biopsies can lead to unnecessary physical pain and scarring, as well as psychological stress. In addition, a large portion of the cost of any breast cancer screening program is associated with false-positive biopsies [3]. Reducing unnecessary false-positive biopsies is therefore of a paramount importance in the breast cancer screening and diagnosis process.

Women with false-positive biopsy results are more likely to manifest suspicious and complex mammographic patterns in their breast parenchymal tissue including tumor-like masses, suspicious microcalcifications, skin thickening or retraction,

ISBN: 978-87-996443-0-8

distortions, and suspicious axillary lymph nodes [2-3]. In light of this fact, analyzing mammographic patterns and characteristics of parenchymal tissue could potentially offer a promising approach to identify women at high-risk for false-positive biopsy. Such an approach could provide objective and reproducible thresholds for assessing the risk of false-positive biopsy from screening or diagnostic mammography.

Substantial research effort to date has been made to develop computerized analysis methods for mammographic parenchymal tissues [4-6]. Some of the previous methods have been focused specifically on the application of breast cancer risk assessment. Such previously developed methods typically fall into two broad categories: (1) manual selection of region-of-interests (ROIs) from the central breast region behind the nipple [4] to investigate parenchymal texture properties and (2) spatial lattice based approach [5-6] that first forms a surrounding window with fixed size for each lattice point and then applies computerized analysis within a given particular window.

In this paper, we develop a novel locally-adaptive texture analysis algorithm to quantify parenchymal tissue complexity, with the intent to provide objective and reproducible thresholds in predicting the risk of false-positive biopsy for a given digital mammogram (DM). Different from previous approaches [4-6], our texture analysis algorithm allows for adaptive ROI selection, depending upon the spatial distribution and patterns of the parenchymal tissue characteristics of the specific mammogram at hand. In addition, in the proposed method, texture features are first locally extracted to capture the more subtle, localized parenchymal texture pattern and then these locally extracted features are combined to represent the global pattern characteristics of entire parenchymal tissue. By adaptively focusing on local tissue characteristics, our approach could provide more discriminating, patient-specific, information for FP biopsy risk prediction. While this paper represents preliminary analysis, ultimate such an approach could be used to better guide tailored screening recommendations, and potentially reduce the number of unnecessary biopsies.

2 Methods and Materials

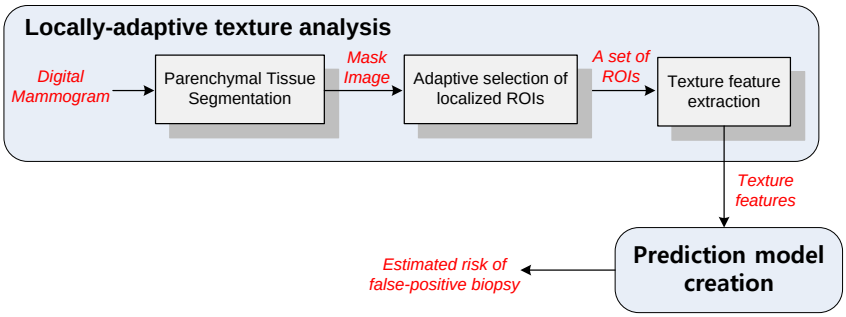


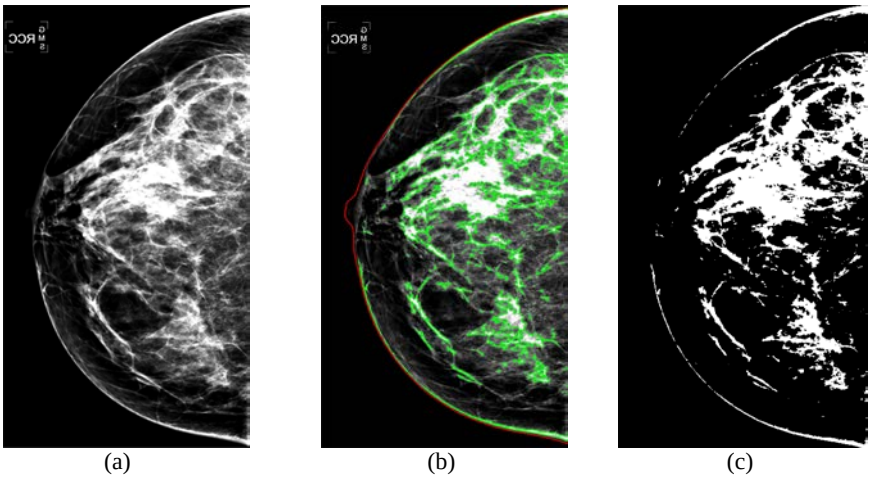
Fig. 1. The proposed framework for predicting the risk of false-positive biopsy from digital mammograms using our proposed locally-adaptive parenchymal patter analysis.

ISBN: 978-87-996443-0-8

Figure 1 provides our proposed framework for the prediction of false-positive biopsy risk from digital mammograms. The method largely consists of two components: (1) the proposed locally-adaptive texture analysis and (2) the classifier prediction model. Also, as shown in Fig. 1, there are three main pipeline modules in the proposed locally-adaptive texture analysis, each of which is described in the following sections.

2.1 Automatic Parenchymal Tissue Segmentation

In general, parenchymal tissue patterns on digital mammograms are a major indicator of the complexity of the breast tissue. In order to perform texture analysis within only breast parenchymal tissue, we first need to delineate the breast area in a mammogram. We therefore first apply an automated segmentation algorithm [7] to all DM images. This segmentation algorithm involves the following main steps: (1) breast region and pectoral muscle segmentation; (2) fuzzy c-means (FCM) clustering [10] based on the gray-level intensities of the breast region; (3) adaptive histogram-based determination of optimal number of clusters for parenchymal tissue segmentation; (4) parenchymal tissue merging based on extracted features and (5) a decision classifier. An example digital mammogram with segmented parenchymal tissue is shown in Figure 2. Note that a mask image, in which the segmented parenchymal tissue regions are represented as the set of foreground pixels (with values '1'), is subsequently applied to the adaptive selection of localized ROIs to be described in the next subsection.



(a) (b) (c)

Fig. 2. Examples of (a) original digital mammogram; (b) segmented parenchymal tissue regions using an automated segmentation algorithm [7]; (c) mask image in which the segmented parenchymal tissue regions are represented as the set of foreground pixels (with values of '1').

ISBN: 978-87-996443-0-8

2.2 Adaptive Selection of Localized ROIs

In this section, we describe the proposed locally-adaptive texture analysis algorithm. Note that the mask image shown in Fig. 1(c) is assumed to be represented as a binary image that has only two possible values (*i.e.*, 1 or 0) for each pixel. Here, the segmented parenchymal tissue regions consist of the foreground pixels (with values of '1'), while the other regions consist of background pixels (with values of '0'). Unless stated otherwise, we will consider the foreground pixels to be the set of interest.

For removing any very small parenchymal tissue regions being segmented, we first apply morphological erosion operation [8] to the given mask image. The rationale behind the use of erosion operation is that segmented parenchymal regions with very small size would not contribute much to characterizing mammographic patterns through using our texture analysis. They would rather act on noise in the process of performing texture analysis. Let $\Re^2$ be a Euclidean space (or an integer grid in the 2-dimensional plane) and let $\mathbf{I}_{\text{mask}}$ be a mask image, where $\mathbf{I}_{\text{mask}}$ is viewed as a subset in $\Re^2$. The erosion operation on the $\mathbf{I}_{\text{mask}}$ by structuring element $\mathbf{S}$ is computed as [8]:

$$\mathbf{I}_{\text{mask}}^{(\text{erosion})} = \{z \in \Re^2 \mid \mathbf{S}_z \subseteq \mathbf{I}_{\text{mask}}\} \tag{1}$$

where $\mathbf{S}_z$ is the translation of $\mathbf{S}$ by the vector $z = (x, y)$, *i.e.*, $\mathbf{S}_z = \{b + z \mid b \in \mathbf{S}\}, \forall z \in \Re^2$. Note that $\mathbf{I}_{\text{mask}}^{(\text{erosion})}$ has a value of 1 at each of the locations where the structuring element fits entirely within the foreground, otherwise has a value of 0. In this study, a disk-shaped structuring element with specific radius equal to 15 was used. We chose this parameter value of the structuring element small enough to fit within relatively large parenchymal tissue regions, while large enough to entirely cover any too small parenchymal tissue regions. Subsequently, we apply the morphological "closing" operation [8] to $\mathbf{I}_{\text{mask}}^{(\text{erosion})}$ as follows:

$$\mathbf{I}_{\text{mask}}^{(\text{erosion+closing})} = \left((\mathbf{I}_{\text{mask}}^{(\text{erosion})})^c \circ \mathbf{S}_z\right)^c \tag{2}$$

where c denotes the complement of a set and $(\mathbf{I}_{\text{mask}}^{(\text{erosion})})^c \circ \mathbf{S}_z = \bigcup_{\mathbf{S}_z \subseteq \left(\mathbf{I}_{\text{mask}}^{(\text{erosion})}\right)^c} \mathbf{S}_z$.

The purpose of using the closing operation is to (a) fill small holes and join narrow breaks and to (b) smooth the contours of the parenchymal tissue objects. This enables us to reliably locate reasonably sized ROIs to be useful for localized texture analysis.

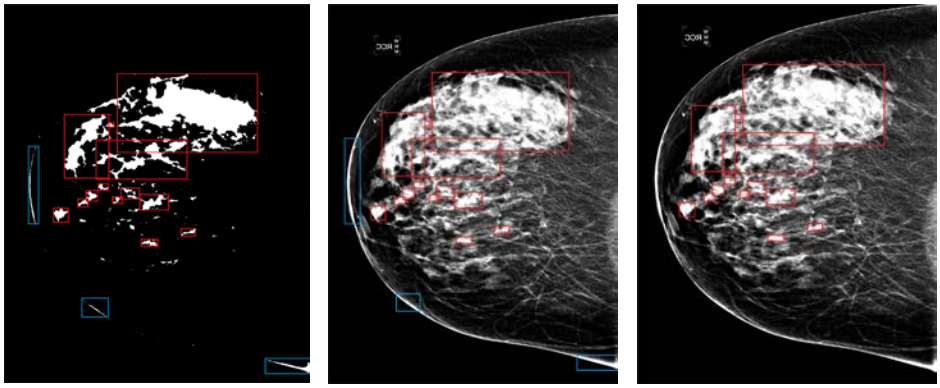
To automatically locate a group of ROIs, we apply a connected component algorithm [8] to the $\mathbf{I}_{\text{mask}}^{(\text{erosion+closing})}$. To this end, we employ 8-connected adjacency to find the set (where the set is defined as the connected component) of all foreground pixels connected to any foreground pixel, z, such that $z \in \mathbf{I}_{\text{mask}}^{(\text{erosion+closing})}$. Based on the connected components, we compute a sequence of path to identify the location of different ROI regions. Examples of a digital mammogram image superimposed with the generated ROIs are shown in Figure 3. Note that as shown in Fig. 3(a), some erroneous ROI regions (*i.e..,* the regions contained in these ROIs are closer to the breast margin and do not really contain true parenchymal tissue) can be observed. To

ISBN: 978-87-996443-0-8

remove these ROIs, the value of overlapping ratio between the size of each ROI and the size of corresponding foreground pixels is computed as follows:

$$r_i = \frac{|\mathrm{R}_i^{\text{foreground}}|}{|\mathrm{R}_i|}, \tag{3}$$

where the R_i is the *i*-th ROI rectangular box, the $\mathrm{R}_i^{\text{foreground}}$ is the region corresponding to foreground pixels of R_i, and $|\cdot|$ returns the region (or area) of a given input argument. If the ratio r_i is less than the prefixed threshold, then we remove the corresponding ROI from the final set of the ROI regions to be used for localized texture analysis. In our study, a good empirical compromise was found by setting $r_i = 0.15$ (*i.e.*, 15% overlap between ROI box and its foreground pixels). As illustrated in Figures 3(b) and (c), the incorrectly located ROIs are successfully removed.



(a) (b) (c)

Fig. 3. Examples of (a) mask image and (b-c) digital mammogram each superimposed with a set of selected ROI regions (rectangular boxes). (b) Before the deletion of incorrect ROI regions (blue boxes). (c) After the deletion of incorrect ROI regions.

2.3 Texture Feature Extraction and Prediction Model Training

Texture features for a given ROI defined in (3) as extracted from a Spatial Gray Level Dependence (SGLD) matrix [9]. Once the SGLD matrix is calculated for each pixel of a given ROI, seven Haralick texture features (descriptors), namely, “entropy”, “energy”, “correlation”, “contrast”, “inverse difference entropy”, “sum average”, and “sum variance” are extracted. Details on the mathematical definition of these seven texture features are found in [9]. These seven different texture features are extracted from each SGLD matrix (of a corresponding ROI) at 12 different distances in the

ISBN: 978-87-996443-0-8

range of [1,12] and in four directions ($\theta = 0^o, 45^o, 90^o, 135^o$), yielding 48 SGLD matrices; this results in a total of 336 features for each ROI. These 336 features are then averaged over the 12 difference distances and four different angles; as a result, the averaged seven texture features are obtained from texture analysis in each ROI.

It should be noted that for a given case, the two standard mammographic views from the breast side that underwent the biopsy were used for our texture analysis. Assuming that from the two mammographic views, K ROI regions, $\mathrm{R}_i\ (i = 1,\ldots,K)$, are finally selected for texture feature extraction (via the way described in Section 2.1), then let $\mathrm{T}_i^{(n)}$ be the n-th type texture feature ($n = 1,\ldots,7$) of all seven texture features for the i-th ROI image ($i = 1,\ldots,K$). To characterize the mammographic patterns of the entire parenchymal breast tissue, as well as to capture the inherent local mammographic appearance at the same time, the mean feature for each corresponding n-th type texture feature is computed over the K selected ROIs such that $\overline{\mathrm{T}}^{(n)} = \frac{1}{K}\sum_{i=1}^{K}\mathrm{T}_i^{(n)}$. Finally, the seven mean texture features, $\overline{\mathrm{T}}^{(n)}$ ($n = 1,\ldots,7$), are used to train a prediction model that estimates the likelihood of false-positive biopsy.

In order to evaluate the proposed locally-adaptive texture analysis, we used three different classifiers as prediction models: a classifier ensemble [10], a Support Vector Machine (SVM) [11], and Linear Discriminant Analysis (LDA) [10]. For the classifier ensemble, we used the adaptive boosting algorithm (AdaBoost) [12]. For the implementation of AdaBoost, we used a decision tree classifier [10], (as base classifier), which is the most widely used in the areas of ensemble classification [12]. To implement tree base classifiers, the 'Gini impurity' [10] criterion was used. In addition, an SVM classifier with a radial basis function (as kernel), was constructed.

2.4 Dataset

This study was in compliance with the Health Insurance Portability and Accountability Act (HIPAA) and received institutional review board (IRB) approval. A retrospective study was performed in a total of 10,187 women who underwent screening digital mammography between 01/2009-06/2010. Bilateral CC and MLO digital mammogram imaging was performed using a full-field digital mammography unit (LORAD Selenia, Hologic Inc.). In this screening cohort, 255 women underwent biopsy after callback for diagnostic evaluation, yielding 66 malignancies (true-positives) and 189 benign or high risk lesions (false-positives). For each of the 255 women that received biopsy, we obtained information regarding known demographic risk factors for breast cancer, such as age. The raw (*i.e.*, "FOR PROCESSING") digital mammograms were acquired with a 70 *um* resolution using a 12-bit gray-level depth. Two standard mammographic views for each breast containing lesions (from 255 women) were available. For the purposes of our preliminary evaluation, 33 malignant cases and 92 benign or high-risk cases (*i.e.*, FP biopsy cases) were randomly selected from the above cohort and were used for the analysis preformed in this study. The post-processed (*i.e.*, "FOR PRESENTATION) CC-view and MLO mammograms were used for our locally-adaptive parenchymal texture analysis.

ISBN: 978-87-996443-0-8

3 Results and Discussion

To achieve the least-biased estimates of classification, our evaluation was carried out using 5x2-Fold cross-validation (cv) [10] for testing all the classifier models considered. In each cv run, we divided the dataset into training and testing halves. The roles were swapped at each fold to generate 10 training and 10 testing sets. Training sets were used to train the classifier models. The resulting 10 classifier models were tested on corresponding 10 testing sets and these 10 testing results were then averaged to compute classification accuracy. Area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate prediction performance. The confidence outputs of classifier models were used as decision variable in ROC analysis.

Figure 4 shows ROC curves and the corresponding AUC values obtained using the classifier ensemble, SVM, and LDA to assess our locally-adaptive texture analysis for predicting false versus true positive biopsy. As shown in Figure 4, the AUC of the classifier ensemble, SVM, and LDA are 0.95 ($p<0.001$), 0.88 ($p<0.001$), and 0.85 ($p<0.001$), respectively. Especially, the highest value of AUC=0.95 (with 95% confidence interval [0.91, 0.99]) was achieved by the classifier ensemble-based prediction model in conjunction with our developed locally-adaptive texture analysis.

In addition, it is worth noting that at some interesting operating points (thresholds) of the ROC curve obtained for the classifier ensemble, we can observe that sensitivity for false-positive biopsies is about 22%, 65%, and 93% respectively, at the 100%, 95%, and 90% cancer detection rate. This suggests that, based on our preliminary results, the number of unnecessary FP biopsies could be reduced up to about 22% at 100% cancer detection rate. These encouraging results suggest that the proposed locally-adaptive texture analysis combined with classifier models may have clinically important value in predicting the risk of false-positive biopsies.

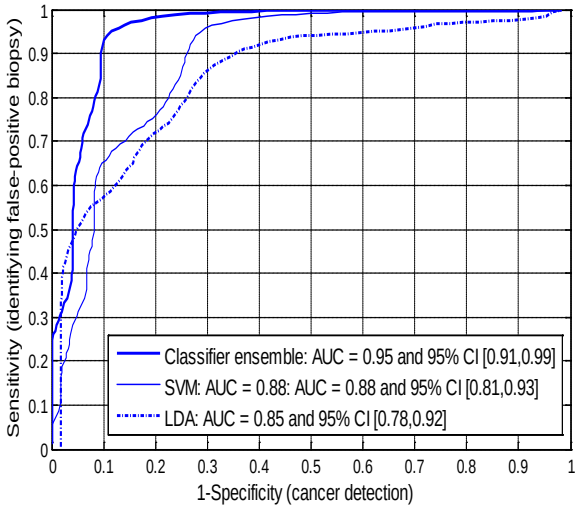


Fig. 4. ROC curves and AUC values using classifier ensemble, SVM, and LDA in conjunction with locally-adaptive texture analysis for predicting false versus true positive biopsies.

ISBN: 978-87-996443-0-8

4 Conclusions

The preliminary work presented in this paper is a first step towards developing new prediction models based on breast parenchymal tissue texture analysis for predicting the risk of false-positive biopsy from digital mammography. It was shown that the combined use of parenchymal tissue complexity texture features and classifier models could result in prediction of false-positive biopsy risk with reasonable accuracy. While this work is considered as preliminary evaluation, the concept of the prediction models proposed in this paper could ultimately have important clinical implication in providing stratification of false-positive biopsy mammograms in personalized breast cancer screening. Larger studies are needed to prospectively validate our findings with additional clinical factors and in conjunction with cancer detection. Future work should also consider the optimization of the texture feature extraction parameters, and combination of the parenchymal pattern features with lesion-specific characteristics.

References

1. McCann, J. et. al., Impact of false-positive mammography on subsequent screening attendance and risk of cancer. Breast Cancer Reserch, 2002. **4**(5): p. 1–9.
2. Euler-Chelpin, M.V. et. al., Risk of Breast Cancer After False-Positive Test Results in Screening Mammography. Journal of the National Cancer Institute, 2012. **104**: p. 1–8.
3. Kopans, D.B. et. al., Breast Imaging (3rd Edition). 2007, Philadelphia: Lippincott Williams & Wilkins.
4. Huo, Z. et. al., Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: Feature selection. Med. Phys., 2000. **27**(1): p. 4–12.
5. Wang, Y. et. al., Texture feature standardization in digital mammography for improving generalizability across devices. SPIE Medical Imaging, 2013. **8670**: p. 1-7.
6. Zheng, Y. et. al., A fully-automated software pipeline for intergrating breast density and parenchymal texture analysis for digital mammograms: Parameter optimization in a case-control breast cancer risk assessment study. SPIE Medical Imaging, 2013. **8670**: p. 1-7
7. Keller, B.M. et. al., Estimation of Breast Percent Density in Raw and Processed Full Field Digital Mammography Images via Adaptive Fuzzy C-means Clustering and Support Vector Machine Segmentation. Med. Phys., 2012. **39**(8): p. 4903-17.
8. Gonzales, R.C. et. al., Digital Image Processing (3rd edition). 2007, New York: Prentice Hall.
9. Haralick, R.M. et. al., Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, 1973. **3**(6): p. 610-21.
10. Kuncheva, L.I., Combining Pattern Classifiers: Methods and Algorithms. 2004, New York: Wiley.
11. Vapnik, V.N., Statistical Learning Theory. 1998, New York: Wiley.
12. Freund, Y. et. al., A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J. Comput. Syst. Sci., 1997. **55**(1): p. 119-39.

ISBN: 978-87-996443-0-8

Characterising and Quantifying the Variation in Adipose and Fibroglandular Tissue between Women when Measuring Breast Density

Christopher E Tromans and Sir Michael Brady

Matakina Technology Ltd, PO Box 24404, Manners Street Central, Wellington, 6142, NZ.
chris.tromans@matakina.com

Abstract. We investigate the effect of the variation in the x-ray and MR characteristics, and thereby composition, of adipose and fibroglandular tissue between women in the general population. To date, these two tissue types have formed the basis for risk assessment using volumetric breast density. In the vast majority of previous work, they have been assumed to be distinct entities, with negligible inter-woman variations between the two types. Using the Standard Attenuate Rate (SAR) for x-ray quantification, and correlating with the 3D volume data available from MR scans, we measure the validity of this approximation. Our findings suggest there might be variation between subjects, in both the adipose and fibroglandular categories and that the variation in tissue characteristics is a potential biomarker for cancer and for breast cancer risk. Variations in the composition identified in MR also appear to correspond to varying textural appearance in the mammographic image. Finally, we identify several studies in the literature which may provide a possible biological basis to our findings.

Keywords: volumetric breast density, tissue characterisation, quantitative x-ray mammography, magnetic resonance imaging.

1 Introduction

Over the past decades much has been contributed to the study of mammographic breast density: the assessment of cancer risk by inferring tissue characteristics from the appearance of a mammogram. This line of work was pioneered in 1969 by Wolfe [1] who proposed a four category classification for visual assessment of mammograms and who presented findings showing that each of the four groups, from lowest to highest density, had progressively increased incidence of developing malignancies. Since then, much has been contributed by Boyd [2], as well as by the Cumulus threshold method [3], to aid quantification accuracy.

There are two main limitations of such visual measurement techniques: a lack of consideration of the substantial effect that the image acquisition parameters have on the appearance of the anatomy within the image; and a lack of consideration of the three dimensional nature of the breast, relying instead on a two dimensional projec-

adfa, p. 1, 2011.

ISBN: 978-87-996443-0-8

tion radiograph which shows only projected areas and which is affected by the extent of compression and tissue overlap (which is why the breasts are tightly compressed in mammograms). Volumetric breast density measurement techniques have been developed to address these issues [4-6]. Broadly speaking, these follow the same underlying methodology of using models of the physics of mammographic image formation, to separate the observed attenuations recorded within the pixel intensities of the image, into volumes of adipose and fibroglandular tissue. This highlights a continuing enigma surrounding breast density: what is the biological and genetic basis underlying the observed phenomena? In the case of volumetric methods, it remains unexplained how the ratio of adipose to fibroglandular tissue drives risk as compared to the volume of fibroglandular tissue. Perhaps an improved understanding of this would improve performance.

Inspection of MR images and specimen radiographs clearly indicates two broad tissue categories, one of higher, and one of lower, density, However, an analysis by Tromans and Brady [7] found considerable variation between published studies reporting the x-ray attenuation properties of adipose and fibroglandular tissue. This raises two questions: (1) what effect does such variation have on the accuracy of quantifying and comparing the adipose/fibroglandular ratio across the population at large; and (2) more interestingly, does the variation in the characteristics of the tissues themselves, rather than simply their relative volumes, carry information about breast cancer risk? These observations led Tromans and Brady to develop the Standard Attenuation Rate (SAR) [7-8], which enables the x-ray attenuation properties of the tissue above each pixel in a mammogram to be quantified, relative to a reference material (this is exactly analogous to Hounsfield units in CT), independently of the image acquisition parameters and with compensation for the effect of scatter. An empirical validation of the technique's normalisation performance using tissue equivalent materials [7] gave upper error bounds of the order of 5%. Note that this makes no assumptions about categorical models which tissues should fit, and thus provides a basis for tissue characterisation and analysis of variation with the population.

The aim of this paper is to present a method and preliminary results which begin to investigate the variation in breast tissue between women as quantified from their x-ray mammograms using SAR. The SAR values are correlated with the voxel intensities observed in T1 weighted MR scans which provide volume information unavailable directly from projection images.

2 Method

The patient images were selected randomly from our image library from those cases for which both a digital mammogram and an MR scan were available. The time between the two examinations was less than one year for all cases, and two months on average. All examinations were performed between December 2000 and September 2003. The mammograms were acquired using standard clinical configurations and settings on a GE Senographe 2000D. Only cranio-caudal (CC) mammograms were

ISBN: 978-87-996443-0-8

used. The SAR software was calibrated for the 2000D using the procedure described in [8].

The MRI examinations were performed on a 1.5T Siemens Magnetom Vision, with a dedicated Siemens CP Breast Array coil. Patients were scanned in the prone position using a dynamic contrast enhanced T1-weighted FLASH-3-D sequence, (repetition time 8.1 ms, echo time 4 ms, and flip angle 20 degrees). The pixel spacing was 1.25 mm by 1.25 mm, and the slice thickness 1.5 mm, with 108 slices acquired, with no interslice gap. The pre-contrast series was used in this analysis. The observable bias field was corrected using the N4ITK algorithm [9] implementation in the ITK 4.3.2 library. The N4 algorithm was applied with a mask defining only the breast volume of interest, using the default parameters in the ITK implementation.

The breast volume within the MRI scan which corresponds to that projected within the mammogram was segmented using a semi-automatic tool developed specially for the purpose (**Fig. 1**). Vertical boundary planes (in the axial view) are placed by the user to remove the arms (if present) and to separate the two breasts at the cleavage. Volumes are then marked as background by the user using a flood fill tool. The breast air boundary is identified by the user using an interactive threshold. A straight boundary plane is also placed at the chest wall which models the back of the mammographic compression paddle, and at an angle which mimics the angle at which the patient stood at the machine. The position of this plane is optimised to give an approximately equal chest wall length between MR and x-ray, and yields a volume of tissue that falls in the range observed for the CC mammogram (the area multiplied by the compressed breast thickness recorded) which has an upper bound given by the breast air boundary assuming the periphery occupies the entire space between the compression plates, and a lower bound which ignores the periphery entirely. In cases where the pectoral muscle was slightly visible in the CC, it was found that this must also be intruded into in the MR scan, in order that the volumes matched as described.

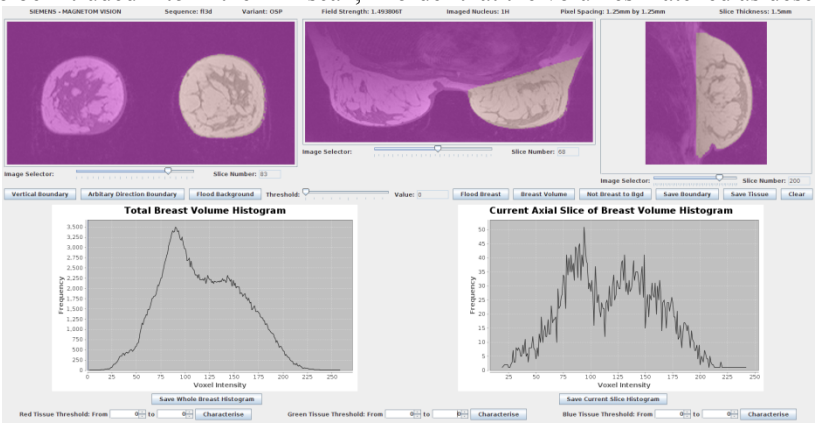


Fig. 1. The tool used for segmenting the MRI volume corresponding to the mammogram.

The tissue volume within the breast in the MRI was segmented by the user using an interactive threshold to separate by eye, as closely as possible, light from dark. The thresholds were marked on the histogram plots (**Fig. 1**) to aid this.

ISBN: 978-87-996443-0-8

The breast area excluding the periphery of the SAR images of the x-ray mammograms were segmented by the user using an interactive threshold.

3 Results

The L-CC SAR images of five women are shown in **Fig. 2**, which are displayed using identical display parameters in terms of window and level, and are arranged so that the SAR pixel intensity which corresponds to the attenuation of adipose according to the composition reported by Hammerstein [10] is black, and the attenuation of fibroglandular, again as reported by Hammerstein [10], is white. Cases 2 and 3 saturate the display range, that is they contain a multitude of pixels for which the x-ray beam apparently encountered an attenuation greater than that of Hammerstein reported fibroglandular. **Fig. 3** shows cases 2 and 3 on a different display range, which has the same window width as **Fig. 2**, but is now centred on Hammerstein reported fibroglandular, rather than this being the maximum extremity as was the case in **Fig. 2**.

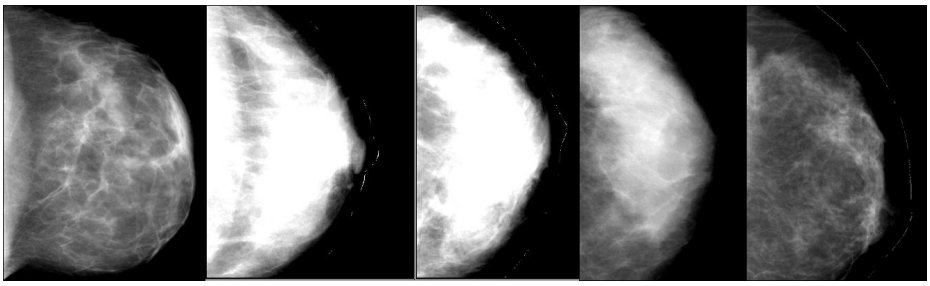


Fig. 2. The SAR image of the L-CC mammogram of five women, cases 1 to 5 from left to right.

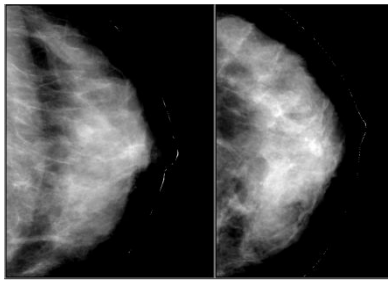


Fig. 3. The SAR images of cases 2 and 3 re-windowed to show denser tissue.

Histograms of the SAR images for all five cases are shown in **Fig. 4**, and the corresponding histograms of the voxel intensities in the MR volume are shown in **Fig. 5**. It may be observed qualitatively that the increased SAR values from higher radiodensities in the dense cases, correspond with lower MR voxel intensities.

ISBN: 978-87-996443-0-8

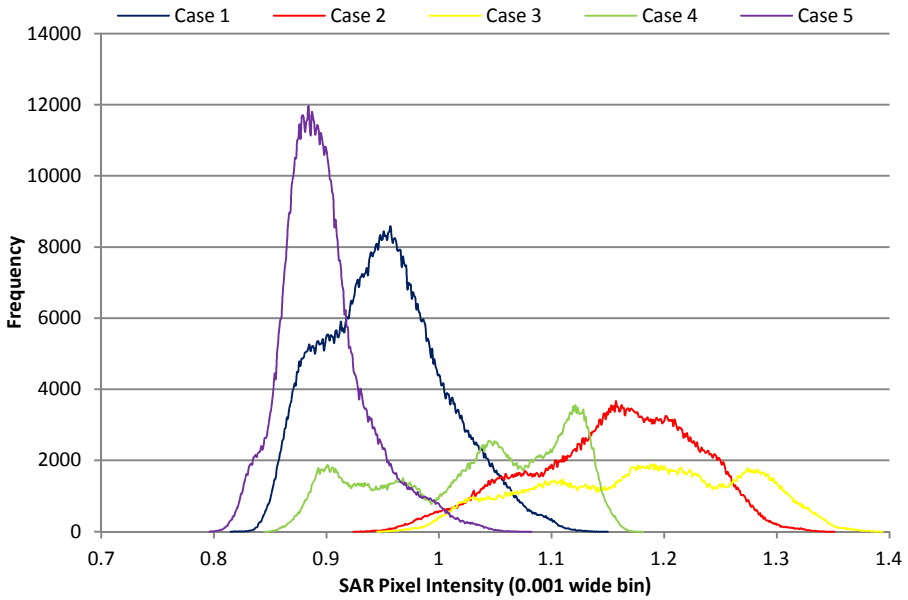


Fig. 4. The histogram of the breast areas of the SAR images for the cases in **Fig. 2**.

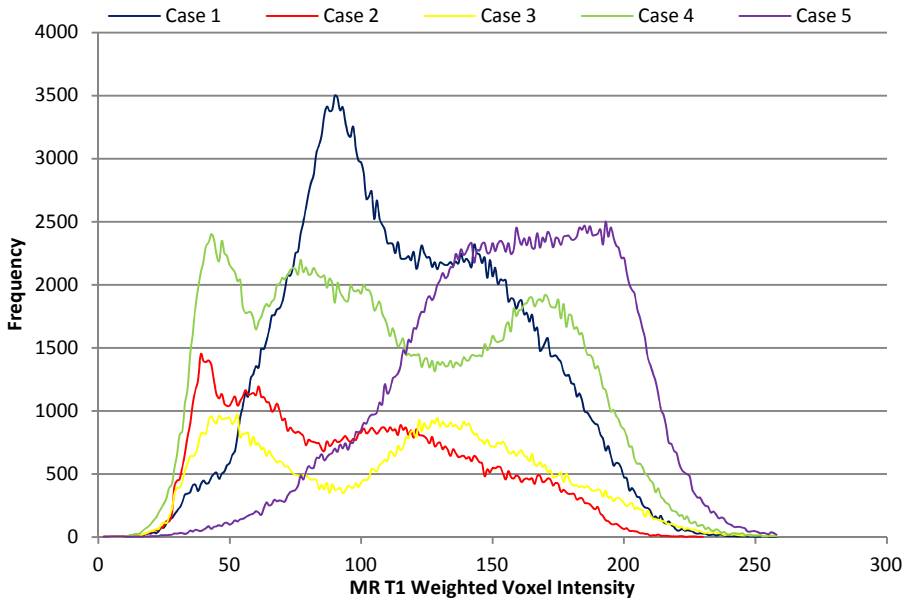


Fig. 5. The histogram of the breast volume in the MR data for the cases in **Fig. 2**.

The mean MR and SAR intensities, for both the adipose and fibroglandular segmentations, as well as the % of fibroglandular tissue in the MR volume, are given in **Table 1**.

ISBN: 978-87-996443-0-8

Table 1. Adipose and fibroglanular statistics for SAR and MR for the cases in **Fig. 2**.

Case	Mean SAR	Mean Fibro MR	Mean Adipose MR	%Fibro
1	0.952	77	143	40.0
2*	1.154	58	133	43.8
3	1.178	66	155	39.3
4	1.032	45	130	19.8
5	0.897	101	181	11.9

*Includes significant pectoral muscle to match mammogram.

The lowest mean SAR value (case 5) is associated with the highest values of MR voxel intensity for both fibroglandular and adipose tissue, with the volume of fibroglandular tissue also occupying the smallest fraction. The second lowest mean SAR, case 1, has a large fibroglandular fraction, but high fibroglandular and adipose MR intensities. The two very dense breasts (cases 2 and 3) have low fibroglandular MR intensity, and high fibroglandular fraction. Case 4 is of interest since it has a relatively low fibroglandular fraction, but also the lowest fibroglandular and adipose MR intensity, and the mean SAR shows the case to be the third most dense. For this case in particular, the high density results from what appears to be higher radiodensity tissue, rather than an increased proportion of fibroglandular tissue.

4 Discussion

Comparing case 1 with case 4: the former has a lower x-ray density than the latter, while it has a larger proportion of fibroglandular tissue, but with the average voxel intensity of both adipose and fibroglandular being lower. The pertinent question is whether or not any difference in cancer risk exists between these two permutations. The difference in texture observed between case 1 and case 4 is also of interest, since the fibroglandular exhibiting the very low MR signal in case 4 is heterogeneously dense (despite there being only 19.8% of it), whilst in case 1, where the fibroglandular MR signal is higher (and there being 40% of it), the density is scattered with a wispy cloud like pattern.

The literature contains possible explanations of the biological processes that may underlie some of the observed results. The adipose tissue in the breast is primarily composed of adipocytes, which specialise in storing energy as fat. The energy is stored in the form of lipids. Petreas et al [11] reported that in a study of 161 subjects "the lipid content in breast adipose was even more variable ranging from <10% to almost 100% with a mean of 76.4%". They conclude "younger women had a broader range of lipid content. Given that age is a known risk factor for breast cancer, use of non-lipid-adjusted adipose tissue concentrations may lead to misclassifications and distort the ORs for disease." A possible explanation for the variations observed here may therefore be a varying lipid concentration within the adipocytes between the various subjects. In the case of fibroglandular tissue, a study by Alowami et al [12] concluded that in areas of high mammographic density that whilst no significant difference in the density of ductal and lobular units exists, a significantly higher collagen

ISBN: 978-87-996443-0-8

density and extent of fibrosis exists within the stroma. This may provide a possible explanation for the variation observed in dense tissue, the collagen density being higher in that with the higher radiodensities quantified in the higher SAR values.

The N4 algorithm [9] using the ITK default parameters appeared to correct only to a partial extent the bias field in the MR images. The magnetic permeability of the thoracic cavity results in the field strength at the chest wall being considerably smaller than that at the nipple, and thus biases the T1 readings which are heavily dependent on the field strength. Had the bias issues not been present, we would have used Otsu's algorithm [13] for MR tissue segmentation. However, the remnants of the bias field gave poor results when assessed visually. This is likely to have also affected the tissue quantification. We intend to investigate parametric corrections in our future work.

The characterisation of tissue in x-ray mammography is complicated by the projection of the 3D volume into 2D, thus requiring consideration to be given to the tissue in the surroundings which are traversed by any given ray. This is expressed more succinctly in (1):

$$I = e^{-(\mu_{background}(H - t_{density}) + \mu_{density} t_{density})} I_0 \quad (1)$$

where $\mu_{density}$ is the value of interest (expressed using SAR in a normalised form), but one must establish $\mu_{background}$ and the thickness of the density $t_{density}$ in order to isolate it. For primarily dense tissue, depending on the breast surroundings, its observed characteristics in a mammogram may be "diluted" by its surroundings. One possible solution for resolving this is to work across the scale of sizes from macro to micro, identifying regions of interest in a mammogram, and working down the scale using specimen radiographs and histology slides to isolate the tissue characteristics of interest.

5 Conclusion

Significant variation was observed between the women investigated in the x-ray and MR characteristics of both the widely adopted adipose and fibrgolandular tissue classes. An increase in the radiodensity as measured by the SAR values, had a corresponding fall in the associated T1 weighted MR voxels. In light of these results it seems that adipose and fibroglandular tissues may need to be treated as broad classifications. The variation observed may carry useful risk information in itself. This pilot study suggests further investigation of the effect of variation in tissue characteristics across the population is worthy of investigation. The use of specimen radiographs and histology slides offers the opportunity to ascertain the biological processes governing the tissue variation, and will be the subject of our future work, so as to explore a possible biological basis for breast density.

ISBN: 978-87-996443-0-8

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme FP7 under grant agreement no 306088 ASSURE.

References

[1] J. N. Wolfe, "The prominent duct pattern as an indicator of cancer risk," *Oncology,* vol. 23, pp. 149-58, 1969.

[2] N. F. Boyd*, et al.*, "Mammographic density and the risk and detection of breast cancer," *N Engl J Med,* vol. 356, pp. 227-36, Jan 18 2007.

[3] J. W. Byng*, et al.*, "The quantitative analysis of mammographic densities," *Phys Med Biol,* vol. 39, pp. 1629-38, Oct 1994.

[4] O. Alonzo-Proulx*, et al.*, "Validation of a method for measuring the volumetric breast density from digital mammograms," *Phys Med Biol,* vol. 55, pp. 3027-44, Jun 7 2010.

[5] R. Highnam and M. Brady, *Mammographic image analysis*. Dordrecht; London: Kluwer Academic, 1999.

[6] S. van Engeland*, et al.*, "Volumetric breast density estimation from full-field digital mammograms," *IEEE Trans Med Imaging,* vol. 25, pp. 273-82, Mar 2006.

[7] C. E. Tromans*, et al.*, "Quantification and normalization of x-ray mammograms," *Phys Med Biol,* vol. 57, pp. 6519-40, Oct 21 2012.

[8] C. E. Tromans*, et al.*, "A model of primary and scattered photon fluence for mammographic x-ray image quantification," *Phys Med Biol,* vol. 57, pp. 6541-70, Oct 21 2012.

[9] N. J. Tustison*, et al.*, "N4ITK: improved N3 bias correction," *IEEE Trans Med Imaging,* vol. 29, pp. 1310-20, Jun 2010.

[10] G. R. Hammerstein*, et al.*, "Absorbed radiation dose in mammography," *Radiology,* vol. 130, pp. 485-91, Feb 1979.

[11] M. Petreas*, et al.*, "Distribution of persistent, lipid-soluble chemicals in breast and abdominal adipose tissues: lessons learned from a breast cancer study," *Cancer Epidemiol Biomarkers Prev,* vol. 13, pp. 416-24, Mar 2004.

[12] S. Alowami*, et al.*, "Mammographic density is related to stroma and stromal proteoglycan expression," *Breast Cancer Res,* vol. 5, pp. R129-35, 2003.

[13] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics,* vol. 9, pp. 62-66, 1979.

ISBN: 978-87-996443-0-8

Atlas-Based Segmentation of Breast MR Images

Farzad Khalvati and Anne Martel

Sunnybrook Research Institute
Toronto, Ontario, Canada
www.sunnybrook.ca/research

Abstract. This paper presents a new pipeline for the segmentation of breast in 3D MR images which can be used to calculate breast density; an important risk factor in determining the possibility of breast cancer. We propose an efficient atlas-based segmentation algorithm that classifies atlas images based on local phase information and uses groupwise registration to create centroid for each class. To segment a target image, it is first determined which class it belongs to. Next, a nonrigid registration is used to align the target image and centroid of the class in order to generate the segmentation result. The proposed algorithm was applied to 400 MRI datasets (350 for training and 50 for testing) and an average Dice coefficient (DSC) of 0.93 was achieved.

Keywords: Breast MRI segmentation, image registration, groupwise registration, phase congruency, image classification

1 Introduction

Magnetic resonance imaging (MRI) of breast is increasingly becoming a common approach for monitoring and detection of breast cancer mainly due to higher sensitivity and no ionizing radiation compared to conventional x-ray mammography. The segmentation of breast tissue in MR images is a crucial task in the analysis of patient data because it makes the 3D visualization possible without the clutter of unimportant structures such as heart obscuring the breast tissue. This is useful in monitoring breast cancer where MR imaging is now also recommended for screening women who are known to be at a higher risk of breast cancer [1]. In addition, it is necessary to segment breast images prior to breast density calculation which is a significant risk factor and an important biomarker in determining the possibility of breast cancer [2]. Moreover, Computer-Aided Diagnosis (CAD) needs breast segmentation prior to performing an efficient auto-detection task and finally, breast segmentation is necessary prior to therapy planning where multiple segmentations of breast images such as CTV (Clinical Tumor Volume) and GTV (Gross Tumor Volume) are required. Atlas-based segmentation (ABS) is a well established and widely used technique for extracting contours from medical images. In this method, processed images are stored in a database or an atlas along with their optimal segmentation results (i.e., manual segmentation or label). A target image is usually registered to the atlas and the label of

ISBN: 978-87-996443-0-8

the atlas is deformed using the registration transformation. In general, there are two approaches to design an ABS algorithm: probabilistic atlas [3] and multi-atlas [4] approaches, where in the former, a probability map of images in the atlas is created and registered to the target image and in the latter, the labels of multiple images in the atlas contribute to generating the segmentation result for the target image. The Multi-atlas approach is usually computationally expensive limiting its practicality. A multi-atlas approach for segmenting pectoral muscle in breast MRI was proposed in [5] where all images in the atlas were registered and compared to the target image and the deformed labels of the best-match images were selected and fused using the method proposed in [4] to yield the final segmentation result.

In this paper, we present a new multi-atlas approach for fully automatic segmentation of breast MR images. It combines image classification and mean image building to create an efficient atlas that while producing highly accurate results, it incurs a reasonable computational cost.

2 Methods

In a multi-atlas approach, the goal is to increase the accuracy of the results by diversifying the atlas (i.e., atlas selection). However, bigger atlases[1] make ABS computationally expensive. We use a pre-processing stage to create local phase maps of images in the atlas (section 2.1), based on which the images are clustered. We use groupwise registration (section 2.2) to create a centroid image[2] for each class which is used for registering to the target image. We present the proposed algorithm in section 2.3 followed by results and discussion in sections 3 and 4, respectively. Section 5 concludes the paper.

2.1 Phase Congruency Map (PCM)

The information about the local phase of an image can be used to detect structural characteristics of the image in a way that is invariant to image intensity. The main idea behind phase congruency is that the Fourier components of an image are all in phase (congruent) where there is a meaningful edge in the image.

Equation 1 calculates the phase congruency of an image at location x ($PC(x)$) where $E(x)$ is local energy of the image, T is a threshold to suppress the effect of noise on the local energy of the image at that location, A_n represents the amplitude of the n^{th} Fourier component, and ϵ is set to a small number to avoid division by 0 [6]. In order to implement phase congruency, a bank of quadrature filters with different spatial frequencies (e.g., Log Gabor filters) are used [6] [7].

$$PC(x) = \frac{\lfloor E(x) - T \rfloor}{\Sigma_n A_n(x) + \epsilon} \tag{1}$$

[1] We refer to images used in training as *atlas* and hence, an atlas may contain one image or multiple images.

[2] Centroid image of each class has minimum dissimilarity from all images in the class.

ISBN: 978-87-996443-0-8

2.2 Groupwise Image Registration

In contrast to pairwise registration where every image in the population is registered to a reference image, groupwise registration transforms a group of images into a reference image such that the dissimilarity between the reference image and each image in the group is minimal. A reference image which has already been manually segmented, plays a crucial role in ABS methods in which it is registered to the target image. In practice, however, generating a reference image is not straight forward; even with careful atlas selection, the selected image may not represent the population well enough and hence, the registration may not produce acceptable results. The advantage of using groupwise registration versus pairwise registration for atlas generation is that a sophisticated mean image that represents all images in the group is automatically generated without being biased toward a specific image in the population. Groupwise registration aligns a set of images to a virtual reference image by generating a set of transformations that map the reference image to each of the images in the group. Balci et al. [8] proposed a framework for groupwise registration of images where it uses the stack entropy cost function and a multi-resolution B-spline non-rigid deformation to generate the set of image transformations. The motivation behind using stack entropy as cost function was that by aligning the images accurately, the intensity values of pixels in the corresponding locations in the stacked images should not vary significantly, which means that the stack entropy should be low [8]. The proposed approach uses a combination of global and local transformations where the former is an affine transformation and the latter is a nonrigid deformation based on B-splines [9].

2.3 Proposed Algorithm

The proposed algorithm in this paper consists of two stages namely training (or atlas building) and testing. The algorithm is applied to 3D breast MR images and therefore, all the intermediate stages (e.g., creating PCMs, clustering, and registration etc.) are performed on 3D images.

Training The training data consists of images with the manual segmentation for whole breast. In order to have a diverse atlas with a reasonable size to manage computationally, we cluster the training images into different classes based on the similarity of the corresponding PCMs. The PCMs of images in the training data are compared (using correlation coefficient) pairwise to create a (dis)similarity matrix. The similarity matrix is fed to a multidimensional scaling (MDS) algorithm [10] to create a 2D distance map of PCMs of all images. The K-means algorithm [11] is used to cluster images into different classes based on the distance map of the corresponding PCMs.

At this point, each class consists of several images and the corresponding PCMs along with the labels. First, we use groupwise registration (section 2.2) to register all the PCMs in each class together to create the representative PCM for each class. Afterward, we apply groupwise registration to the original images in

ISBN: 978-87-996443-0-8

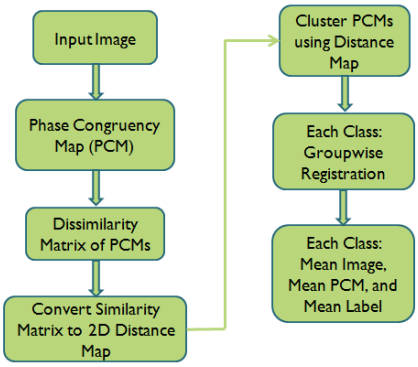


Fig. 1. Training

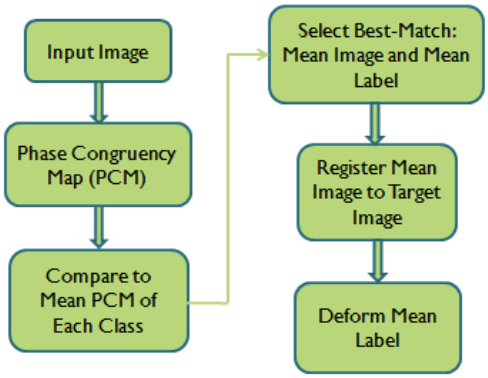


Fig. 2. Testing

each class to create the mean image and the corresponding mean label. For each class of images in the atlas, this gives a representative PCM, a mean image (or centroid), and mean label to be used for segmenting the target image (Figure 1).

Test When an unseen image (target image) arrives, first, its PCM is created and compared to the representative PCMs of all classes. Once the best-match class is found, the corresponding mean image is registered to the target image via a nonrigid registration algorithm using elastix [12]. The registration transformation is then applied to the mean label of the best-match class to create the segmentation result for the target image (Figure 2).

2.4 Materials

The training and test data, used from a previous study [13], contained 400 breast MRI datasets ($94 \times 94 \times 44$ pixels) of Dixon imaging sequence (used for water and fat separation) manually segmented by an expert to mark the boundaries of breast. The population consisted of two age groups; 320 women aged 15-30 and 80 women aged 40-60. Out of 400 MRI datasets, 350 datasets were used to create the atlas (280 datasets from the younger and 70 datasets from the older group) and 50 datasets were used to generate breast volume using the proposed algorithm to evaluate performance (40 datasets from the younger and 10 datasets from the older group).

3 Results

To evaluate the performance of the proposed algorithm, the segmentation results for the entire volume were compared to the ground-truth results (i.e., manual

ISBN: 978-87-996443-0-8

segmentation) using Dice Similarity Coefficient (DSC) and Jaccard index[3]. The best configuration of the algorithm (i.e., 20 classes of images in the atlas) yielded mean DSC and Jaccard index of 93%±5% and 87±8%, respectively. The median values for the DSC and the Jaccard index were 94% and 89%, respectively. Figure 3 shows a sample test image, the corresponding best-match atlas, and the segmentation result for a single slice. Figure 4 shows the distribution of the segmentation accuracy for all 50 patients test data. To segment a target MR image volume, it took 2 min on Intel(R) Core(TM)2 i5 CPU 3.33GHz.

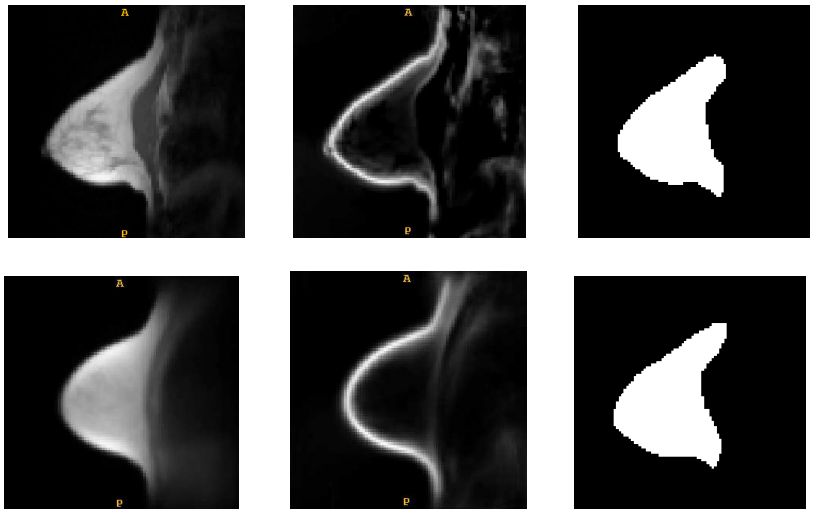


Fig. 3. Top - Target image (left to right): original image, PCM, ground-truth result. Bottom - (left to right): best-match centroid, best-match PCM, segmentation result.

4 Discussion

The proposed algorithm classifies the training images, based on the similarity of the corresponding PCMs, using K-means algorithm. The target image's PCM is compared to the representative PCMs of all classes and the best-match centroid (or mean image) is selected to be registered to the target image. With respect to image registration terminology, the target image is in fact the *fixed* image $F(x)$ and the best-match centroid is the *moving* image $M(x)$ with mean label $M'(x)$. By registering $M(x)$ to $F(x)$, we obtain a transformation $y(x)$ which yields:

$$M(y(x)) \approx F(x) \tag{2}$$

[3] For two sets A and B, DSC and Jaccard index are defined as $\frac{2|A \cap B|}{|A|+|B|}$ and $\frac{|A \cap B|}{|A \cup B|}$, respectively.

ISBN: 978-87-996443-0-8

Applying the transformation $y(x)$ to the mean label of the best-match centroid $M'(x)$ produces $M'(y(x))$. The closer $M'(y(x))$ to the actual label of $F(x)$ (i.e., $F'(x)$), the higher the accuracy of the segmentation results will be. The number of classes is expected to influence the overall performance of the algorithm. To investigate this, we ran experiments for different numbers of classes in the atlas (i.e., 1 class to 75 classes) with the same training and test data as described in section 2.4.

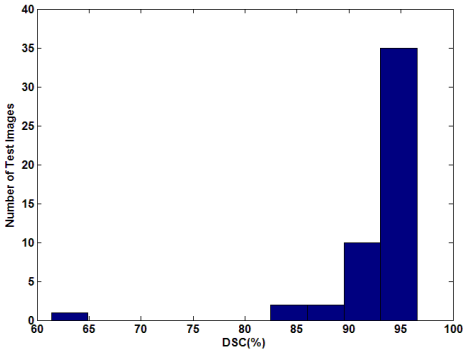


Fig. 4. DSC value comparison between the segmentation and ground-truth results

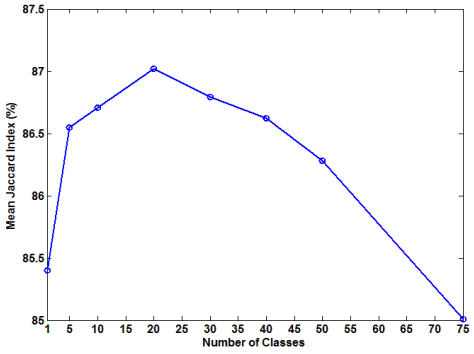


Fig. 5. Algorithm performance with respect to the number of classes in the atlas

Figure 5 shows the mean Jaccard index values for the test data for different numbers of classes used to classify the training images. It is interesting to observe that there is an optimal number of classes of images[4] (i.e., 20) that yields the best results in terms of the accuracy of the proposed algorithm. The first phenomenon that affects the performance of the proposed algorithm is the fact that when registering the best-match centroid $M(x)$ to the target image $F(x)$, for a fixed number of iterations of the cost function's optimizer, the more similar $M(x)$ and $F(x)$, the higher the similarity of the transformed $M(x)$ and $F(x)$ will be (i.e., higher $sim(M(y(x)), F(x))$). This will lead to a more accurate segmentation result (i.e., higher $sim(M'(y(x)), F'(x))$).

From K-means classification it is known that the more the number of classes, the higher the similarity of a new random variable to the best-match centroid will be. This is, in general, true because as the number of classes increases, the error of clustering (i.e., dissimilarity between target image and centroid of best-match class) should decrease because clusters are smaller. We ran an experiment

[4] The optimal number of classes may be different for different training/test images.

ISBN: 978-87-996443-0-8

for different numbers of classes and it was confirmed that the higher the number of classes of atlas images, the more similar the best-match centroid image was to the target image (Figure 6), which means higher $sim(M(x), F(x))$.

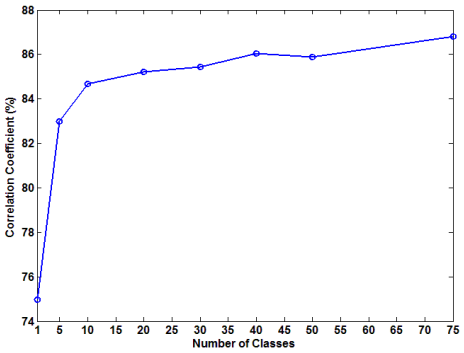


Fig. 6. Similarity of target image vs. best-match centroid

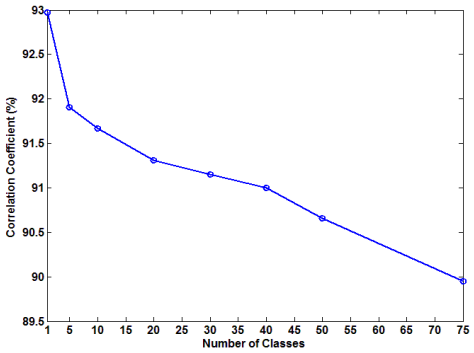


Fig. 7. Similarity of best-match centroid vs. its label

Another phenomenon that affects performance of the algorithm with respect to the number of classes (Figure 5) is the amount of similarity between the best-match centroid and its label. When registering $M(x)$ to $F(x)$, all pixels in $M(x)$ are considered. In order to obtain the segmentation result, the transformation obtained by registering $M(x)$ to $F(x)$, $y(x)$, is applied to a fraction of pixels in $M(x)$ which has been binarized (i.e., the label or $M'(x)$). Thus, the more similar $M(x)$ and $M'(x)$, the more accurate the result of applying $y(x)$ to $M'(x)$ will be (i.e., higher $sim(M'(y(x)), F'(x))$).

For a fixed number of training images (e.g., 350), increasing the number of classes decreases the average number of images per class. When registering a larger population of images in one class using groupwise registration, there is a higher chance that the mean image (i.e., centroid) converges toward the region(s) of interest (i.e., label). This may be due to the fact that the region(s) of interest in the image contain(s) less randomness in their texture in comparison to the background. We ran an experiment for different numbers of classes and it was confirmed that the higher the number of classes (i.e., fewer images per class), the less similar the centroid image was to its corresponding label (Figure 7)[5].

The two phenomena discussed above contribute to the behaviour of the proposed algorithm in which there is an optimal number of classes that yields the best accuracy for the segmentation results (Figure 5).

[5] The similarity of two images was calculated by correlation coefficients.

ISBN: 978-87-996443-0-8

5 Conclusion

A new multi-atlas-based segmentation algorithm was presented where atlas images are clustered based on local phase maps and each class centroid is created using groupwise registration. The proposed algorithm generates highly accurate results for segmentation of breast MR Dixon images with a reasonable computational cost. As future work, local phase maps of training images will be directly used to create an atlas that is intensity invariant. An ABS algorithm will be developed in which the intensity-invariant atlas created from Dixon images will be used to segment other MR image sequences such as T1 and T2.

References

1. Warner E.: The role of magnetic resonance imaging in screening women at high risk of breast cancer. Top Magn Reson Imaging. 19(3):163-9 (2008)
2. Boyd, N., Guo, H., Martin, L., Sun, L., Stone, J., Fishell, E., Jong, A., Hislop, G., Chiarelli, A., Minkin, S., Yaffe, M.: Mammographic density and the risk and detection of breast cancer. N. Engl. J. Med. 356, 227236 (2007)
3. Martin, S., Troccaz, J., Daanen, V.: Automated segmentation of the prostate in 3D MR images using a probabilistic atlas and a spatially constrained deformable model, Med. Phys., Vol. 37, pp. 1579-1590, 2010.
4. Klein, S., van der Heide, U., Lips, I., van Vulpen, M., Staring, M., Pluim, J.P.W.: Automatic segmentation of the prostate in 3d MR images by atlas matching using localized mutual information. Med. Phys. 35(4), 14071417 (2008)
5. Gubern-Mrida A., Kallenberg M., Mart R., Karssemeijer N.: Segmentation of the pectoral muscle in breast MRI using atlas-based approaches, Med. Image Comput. Comput. Assist. Interv. 15(Pt 2), 371-8 (2012)
6. Kovesi, P.: Phase congruency: a low-level image invariant. Psychological Research 64 (2), 136148 (2000)
7. Gallego Ortiz, C., Martel, A. L.: Automatic atlas-based segmentation of the breast in MRI for 3D breast volume computation. Med. Phys. 39 (10), 5835–5848. (2012)
8. Balci, S. K., Golland, P., Shenton, M., Wells, W.M.: Free-Form B-spline Deformation Model for Groupwise Registration. Int. Conf. Med. Image Comput. Comput. Assist. Interv. 10, 23–30.(2007)
9. Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., Hawkes, D. J.: Nonrigid registration using free-form deformations: Application to breast mr images. IEEE Tans. Med. Imaging 18 (9),712-721. (1999)
10. Steyvers, M.: Multidimensional Scaling. Multidimensional Scaling. Encyclopedia of Cognitive Science. (2006)
11. Lloyd, S. P.: Least squares quantization in PCM. IEEE Trans. Inf. Theory 28(2), 129136 (1982)
12. Klein., S., Staring, M., Murphy, K., Viergever, M. A., Pluim, J. P. W.: elastix: a toolbox for intensity based medical image registration. IEEE Trans. Med. Imaging, 29 (1), 196–205. (2010)
13. Boyd, N., Martin, L., Chavez, S., Gunasekara, A., Salleh, A., Melnichouk, O., Yaffe, M., Friedenreich, C., Minkin, S., Bronskill, M.: Breast-tissue composition and other risk factors for breast cancer in young women: A cross-sectional study, Lancet Oncol. 10, 569580 (2009)

ISBN: 978-87-996443-0-8

Automated localization of malignant lesions in breast DCE-MRI

Albert Gubern-Mérida[1,2], Bram Platel[2], Ritse M. Mann[2], Robert Martí[1], and Nico Karssemeijer[2]

[1] University of Girona, Spain
{agubern,marly}@eia.udg.edu
[2] Radboud University Nijmegen Medical Centre, The Netherlands
{a.gubernmerida,b.platel,r.mann,n.karssemeijer}@rad.umcn.nl

Abstract. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is increasingly used for breast cancer assessment. Compared to mammography DCE-MRI provides higher sensitivity, however the specificity of DCE-MRI is variable. Continued efforts are focused on identifying distinguishing characteristics of malignant and benign lesions. DCE-MRI data analysis is time consuming and presents high inter- and intra-observer variability. The aim of this work is to propose an automated breast lesion localization system for DCE-MRI. Such a system can be used to support radiologists during DCE-MRI analysis, to facilitate pre-calculation of very computationally demanding features and to form the basis of a standalone computer aided diagnosis application. The proposed method initially segments the breast and uses a gentle adaboost classifier and features extracted from the relative signal enhancement to detect malignant lesions. Evaluation was performed on a dataset of 212 DCE-MRI studies from 126 patients with no sign of breast cancer and 86 patients with biopsy-proven annotated malignant lesions. The results obtained by our method are promising for clinical applications: 96% of the lesions of our study dataset were correctly detected at 10.4 false positives per patient without cancer.

Keywords: breast cancer, lesion localization, breast DCE-MRI, breast segmentation, computer-aided detection

1 Introduction

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) of the breast is increasingly employed for screening of breast cancer in high risk patients, evaluation of tumor extent in specific groups of patients with breast cancer (e.g. patients with invasive lobular carcinoma), dense breast diagnosis, evaluation of tumor response to chemotherapy treatment and as an additional diagnosis tool in case of inconclusive findings from other modalities [8]. Compared to mammography, which is the image modality used in screening, breast DCE-MRI has a higher sensitivity [10]. However, DCE-MRI analysis requires interpretation of four-dimensional data and is therefore more time-consuming. Furthermore, specificity varies between 67% and 72% [10].

ISBN: 978-87-996443-0-8

In order to overcome these limitations, computer aided detection (CAD) systems are currently used to interactively assist the radiologist in detection and classification of breast lesions in DCE-MRI data. These systems suffer from inter- and intra-observer variability because human interaction is required to identify and characterize suspicious areas [9]. Other more sophisticated methods for lesion characterization [1, 9, 6] are also found in the literature. Clinical implementation of some of these complex methods has been hampered by the heavy computations required that can only be started after manual localization of a possible lesion. This can be alleviated by an automatic lesion detection system that indicates suspicious areas. These detections allow algorithms to compute required parameters without user interaction. Furthermore an automated detection system can form the basis for a fully automatic lesion classification algorithm. Both scenarios aid in the reduction of subjectivity, false positive findings and analysis time.

To our knowledge, automatic lesion detection in breast DCE-MRI is still an open problem and only a few authors presented algorithms aiming to automate the task. Vignati et al. [12] and Ertas et al. [2] focused on automatic lesion detection, and Renz et al. [11] presented an approach to automatically detect and classify breast lesions. The authors reported satisfactory results but their studies have some limitations since the performance of the algorithms was not tested on dataset with normal DCE-MRI studies [12, 2, 11], evaluation was performed in small datasets [2] or the proposed algorithms were only applied to mass-like lesions [12, 11].

The purpose of this study is to obtain automated localization of breast lesions in DCE-MRI with a high sensitivity and an acceptable number of false positives, allowing offline analysis of suspicious regions before the actual reading takes place. Evaluation was performed on a dataset of 212 DCE-MRI studies from different patients (86 with biopsy-proven malignant lesions and 126 confirmed normal patients).

2 Material and Methods

2.1 Study dataset

The dataset used in this study was composed of 212 T1-weighted coronal DCE-MRI studies from different patients (age: 22-81 years, mean age: 48.20 years). At least one malignant lesion was visible in 86 DCE-MRI studies (total number of malignant lesions: 109). All lesions were histopathologically proven and retrospectively annotated by an observer with experience in breast MRI. A radiologist supervised all the annotations. The annotations were performed with an in-house developed dedicated breast DCE-MRI annotation workstation that provided the visualization of DCE-MRI time sequences, subtraction images and relative enhancement images. Each annotation was represented by a sphere centered in the lesion with radius corresponding to lesion size. Multiple overlapping spheres were used to annotate non-mass-like lesions. The distribution of malignant lesion types in the dataset we used was as follows: 24 ductal carcinoma in

ISBN: 978-87-996443-0-8

situ (DCIS), 25 invasive lobular carcinoma (ILC), 11 adenocarcinoma, 48 invasive ductal carcinoma (IDC) and 1 lobular carcinoma in situ (LCIS).

The remaining 126 DCE-MRI studies of the complete dataset were obtained from different patients with no sign of breast cancer. These patients and their scans are referred to as normal patients and normal DCE-MRI in the paper. The women included in the subset of normal patients participated in a high-risk screening program, obtained BI-RADS diagnostic outcome of either 1 or 2 in all their annual DCE-MRI studies, their DCE-MRI studies included in the dataset had at least two follow up scans, and no previous history of breast cancer or breast surgery was reported.

Breast DCE-MRI image acquisitions were performed on either a 1.5 or 3 Tesla Siemens scanner (Magnetom Vision, Magnetom Avanto and Magnetom Trio), with a dedicated breast coil (CP Breast Array, Siemens, Erlangen). All acquisitions were collected from 2003 to 2010. Pixel spacing (from 0.586 mm to 1.25 mm), coronal slice thickness (from 1 mm to 1.5 mm) and 4D volume size (left-to-right: [256 - 512] voxels, superior-to-inferior: [128 - 256] voxels, posterior-to-anterior [72 - 160] voxels, and time points: [5 - 6]) differed between acquisitions. The first time point (T_0) was acquired before intravenous agent injection. The acquisition of the following time points (T_1 - T_n) started within the first 2 minutes after intravenous agent injection as indicated by the breast MRI guidelines from the European Society of Breast Imaging [8]. The contrast medium was administered at a dose of 0.1 mmol/kg (Medrad, Warrendale, PA) using a power injector (Medrad, Warrendale, PA) at a flow rate of 2.5 mL/s, followed by a saline flush.

2.2 Fully automatic detection framework

A general overview of the automatic detection process is shown in Fig. 1 and is explained below.

Fig. 1. Automatic lesion detection framework in breast DCE-MRI.

Motion correction Motion correction is initially applied to remove artifacts from the DCE-MRI data. These artifacts are caused by movement of the patient during the image acquisition process and could influence the lesion kinetic characteristics. Images at post-contrast time points (T_1 - T_n) are mapped to the pre-contrast acquisition T_0 by using a registration algorithm that combines rigid and non-rigid B-Splines transform. Mutual information is used as similarity measure because signal intensity varies along time. Elastix [5] was used for the implementation.

ISBN: 978-87-996443-0-8

Breast volume segmentation The breast is automatically segmented to focus the detection only on the breast and therefore reduce false-positive findings in other areas. A breast MRI segmentation method that captures anatomic variation of the pectoral muscle and chest wall region is applied to the pre-contrast DCE-MRI volume T_0. The method is an extension of previously reported work [4]. It uses spatial information provided by probabilistic atlases. A probabilistic atlas is a volume that contains the complete spatial distribution of probabilities for a voxel to belong to one or more organs. In this work, three probabilistic atlases are used to have a good representation of differences in women with small, medium and large breasts. The size of the breast is measured as the distance along the anterior-posterior axis between the anterior of the breast and the coronal plane through the sternum. This distance is noted as d_{AP} (see Fig. 2(a)). The large probabilistic atlas is chosen to segment patients with d_{AP} larger than 95 mm, the medium for patients with d_{AP} between 66 and 95 mm, and the small for patients with d_{AP} smaller than 66 mm. Figure 2(b) shows an example of the segmentation obtained by the automatic breast segmentation approach.

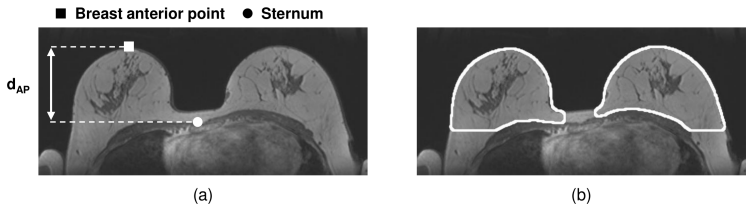


Fig. 2. Automatic breast segmentation: axial slice of pre-contrast DCE-MRI volume T_0 with (a) breast size measured as d_{AP} and (b) its breast segmentation.

Feature extraction The voxel features used in the lesion detection framework are computed from the relative signal enhancement at the first post-contrast time point (T_1). The relative signal enhancement at T_1 represents the concentration of contrast agent at the first time point after contrast agent injection and is defined as

$$R_1 = \frac{T_1 - T_0}{T_0}. \tag{1}$$

We chose this first time point because, considering the image acquisition protocol used in our study dataset, malignant lesions reach their maximum enhancement at this time point [8].

Seven features are used. The first feature is the smoothed R_1 defined as $\overline{R}_{1\sigma_r} = G_{\sigma_r}(R_1)$, where G_{σ_r} is a Gaussian smoothing filter at scale σ_r. $\overline{R}_{1\sigma_r}$ is computed to reduce artifacts produced by noise and local deformation not corrected in the motion correction step. In our experiments, σ_r was set to 1.25 mm. The remaining six features are used to characterize bright blob-like structures

ISBN: 978-87-996443-0-8

because breast lesions mostly appear as bright blobs in relative signal enhancement volumes. Two blob detection filters are used in this work: a blob detection filter based on the Laplacian filter and a blob detection filter based on the eigenvalues of the Hessian matrix as defined by Li et al. [7]. Both blobness measures are based on the second derivate of R_1 at certain Gaussian scale parametrized by σ_b to detect blobs of different size. In this work, blobness features at $\sigma_b = 3$, 10 and 17 mm are calculated for each blob detection filter. The σ_b values were selected to cover the maximum range of lesion sizes.

Initial voxel candidate detection An initial detection step is applied to reduce the number of voxels processed by the classifier of the final candidate detection stage. All voxels with $\overline{R}_{1\sigma_r}$ lower than a suspiciousness threshold s_t are marked as non-lesion voxels. The s_t can be seen as the value that discriminates between a suspicious and non-suspicious voxel in terms of contrast agent concentration measured with $\overline{R}_{1\sigma_r}$. In this work s_t was 0.51. This value was computed as the minimum $\overline{R}_{1\sigma_r}$ value found on a subset of voxels that contained the 20 voxels with highest $\overline{R}_{1\sigma_r}$ of each lesion annotated in our dataset.

Final candidate detection In the final detection stage, first a likelihood of abnormality is computed at each voxel with $\overline{R}_{1\sigma_r} \geq s_t$ using the seven images features described above. The gentle adaboost classifier [3] with regression stumps as weak learners was chosen because it is considered a state-of-the-art classifier. The number of regression stumps was set to 50. To obtain the final lesion candidates, local maxima of the voxel likelihood map are determined, by means of a spherical kernel (radius = 5 mm). Finally, the highest maxima within a radius of 20 mm are taken as the final candidates. We chose for this two-stage approach and not a single sweep with a 20 mm spherical kernel to reduce computation time.

The training stage was designed to minimize the number of false-positive results at high sensitivity detection. The training set of cases contained DCE-MRI studies with and without lesion annotations (normal patients). The process of selecting voxels to build the training dataset is defined as follows:

for *each DCE-MRI scan of the training set with a malignant lesion* **do**
  Select the $n_{malignant}$ voxels with the highest $\overline{R}_{1\sigma_r}$
  Randomly select n_{normal} voxels from normal cases included in the training set that satisfy:
    1. The voxels belong to the breast.
    2. $\overline{R}_{1\sigma_r} \geq s_t$
end

By selecting normal samples with high $\overline{R}_{1\sigma_r}$, we provide the classifier with normal samples that are likely to be detected as false positive because they present significant contrast enhancement. Moreover, we ensure that only true non-lesion voxels are included in the training stage since only voxels from normal cases are selected. In our experiments, $n_{malignant}$ and n_{normal} were set to 20 and 300 voxels, respectively.

ISBN: 978-87-996443-0-8

2.3 Experiments and evaluation

Several parameters are defined in the proposed lesion detection algorithm. We empirically experimented with different values. The parameter values described above are the ones which obtained the best performance measured as the trade-off between high sensitivity and low false-positive detections.

We applied our lesion detection method to all the 212 DCE-MRI (normal patients and patients with lesions) using a leave-one-out strategy on patient level. Reported results were obtained using the best combination of parameters values described previously. Results without $\overline{R}_{1\sigma_r}$ and using the relative signal enhancement R_1 without smoothing instead, and, independently, results without using blobness features in the final candidate detection stage are also reported to show the importance of these features.

We used free-response receiver operating characteristic (FROC) analysis to determine the performance of our system on the suspiciousness level obtained per lesion. As mentioned before, each annotation was a sphere or a combination of spheres covering the lesion. A lesion is considered a true positive when a lesion candidate hits the annotation. Each lesion can only obtain one hit. When multiple candidates hit the annotation, the one with the highest suspiciousness is chosen and the rest is discarded. False positive detections are computed on normal patients.

3 Results

FROC curves of Fig. 3 show the performances for the detection of malignant lesions. The lesion detection system presented in this paper achieved a lesion-based sensitivity of 59% and 73% at 1 and 2 false positives per DCE-MRI volume of a normal patient (see Fig. 3 (1)). Maximum sensitivity (96%) is reached at 10.4 false positives per normal DCE-MRI volume. Figure 3 also shows the performance of (2) the detection algorithm using the relative signal enhancement R_1 without smoothing and (3) the lesion detection algorithm without incorporating blobness features in the final detection stage. Figure 4 shows examples with three different types of lesions that were detected by our algorithm.

4 Discussion

We have developed a lesion detection system for breast cancer in DCE-MRI. The presented framework provides the automated localization of breast lesions. The FROC curves demonstrate that our approach obtains promising results for clinical application. The sensitivity at lesion level was 73% at 2 false positive per DCE-MRI volume from a normal patient. A maximum sensitivity (96% per lesion) is reached at 10.4 false positive per normal DCE-MRI. Only normal patients were used to compute false positives in order to evaluate the performance of the presented approach when applied to a screening population.

Seven voxel features are used in the detection framework. We showed that smoothed relative signal enhancement $\overline{R}_{1\sigma_r}$ and the blobness features provide meaningful information. The smoothed relative signal enhancement is used to

ISBN: 978-87-996443-0-8

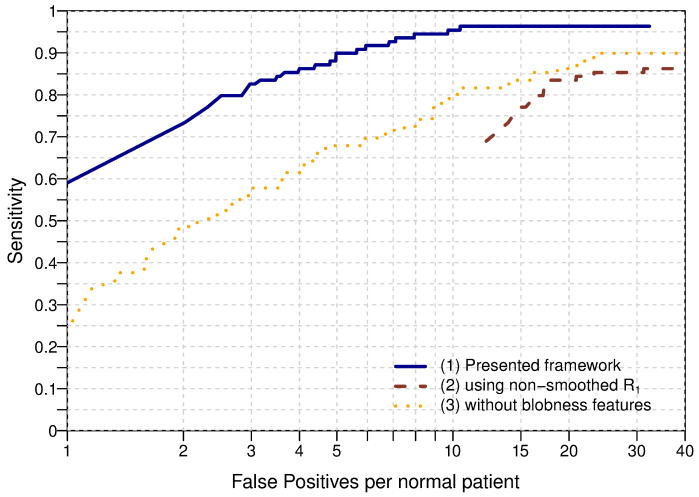


Fig. 3. FROC curves of the proposed lesion detection system using (1) the parameters values that obtained the best performance, (2) using R_1 instead of smoothed $\overline{R}_{1\sigma_r}$, and (3) without incorporating blobness features in the final stage.

initially select suspicious voxels and is incorporated as a feature in the last lesion detection stage. The number of false positives considerably increases when smoothing is not applied to the relative signal enhancement. In this case small artifacts cause many false positive detections. The number of malignant samples selected in the training stage of the final detection step had also a strong influence in the performance of our algorithm. We observed that, by including 20 malignant voxels per case, the most representative training malignant samples of each case were considered for training. Sensitivity dropped when $n_{malignant}$ was set with smaller values. Differences were not observed when $n_{malignant} > 20$.

It is noted that four breast malignant lesions were missed (3.6% of the total number of lesions). An adenocarcinoma was missed by the local maxima selection

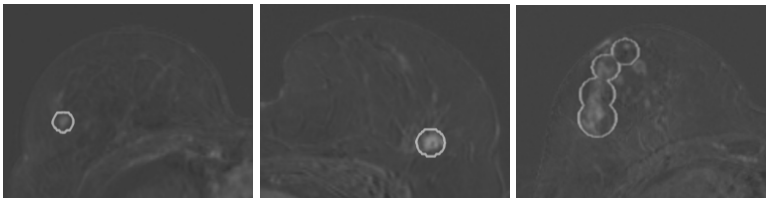


Fig. 4. Axial slices of subtraction images ($T_1 - T_0$) from different patients with three different types of malignant lesions and their manual annotations. These lesions were automatically detected by our algorithm.

ISBN: 978-87-996443-0-8

in a patient with another annotated lesion within 20 mm distance. Both annotations obtained high likelihood of abnormality but only one was chosen as a final candidate. A DCIS was missed because blob-like shape enhancement was not visible. The other 2 lesions were not detected due to breast segmentation issues. One of these lesions was missed because the algorithm incorrectly excluded the area close to the chest wall where the lesion was located. This patient had dense breast and the separation of the breast from the body was difficult. The other lesion was located near the axila and the breast segmentation algorithm used by our detection framework does not consider this area as breast. Voxel features to characterize lesions with less blob-like shape and other breast segmentation techniques will be studied in the future to solve these problems and improve the performance. Moreover, a more sophisticated optimization procedure will be implemented to study and choose the best parameters values.

References

1. Chen, W., Giger, M.L., Bick, U., Newstead, G.M.: Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI. Med Phys 33, 2878–2887 (2006)
2. Ertaş, G., Gülçür, H.O., Osman, O., Uçan, O.N., Tunaci, M., Dursun, M.: Breast MR segmentation and lesion detection with cellular neural networks and 3d template matching. Comput Biol Med 38, 116–126 (2008)
3. Friedman, J., Hastie, T., Tibshirani, R.: Special invited paper. additive logistic regression: A statistical view of boosting. Ann Stat 28, pp. 337–374 (2000)
4. Gubern-Mérida, A., Kallenberg, M., Martí, R., Karssemeijer, N.: Fully automatic fibroglandular tissue segmentation in breast MRI: atlas-based approach. In: MICCAI Workshop: Breast Image Analysis (2011)
5. Klein, S., Staring, M.: Elastix. `http://www.isi.uu.nl/Elastix/` (Accessed March 30, 2009)
6. Levman, J., Leung, T., Causer, P., Plewes, D., Martel, A.L.: Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines. IEEE Trans Med Imaging 27(5), 688–696 (May 2008)
7. Li, Q., Sone, S., Doi, K.: Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. Med Phys 30, 2040–2051 (2003)
8. Mann, R.M., Kuhl, C.K., Kinkel, K., Boetes, C.: Breast MRI: guidelines from the european society of breast imaging. Eur Radiol 18, 1307–1318 (2008)
9. Newell, D., Nie, K., Chen, J.H., et al.: Selection of diagnostic features on breast MRI to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement. Eur Radiol 20, 771–781 (2010)
10. Peters, N.H.G.M., Rinkes, I.H.M.B., Zuithoff, N.P.A., et al.: Meta-analysis of MR imaging in the diagnosis of breast lesions. Radiology 246, 116–124 (2008)
11. Renz, D.M., Bttcher, J., Diekmann, F., et al.: Detection and classification of contrast-enhancing masses by a fully automatic computer-assisted diagnosis system for breast MRI. J Magn Reson Imaging 35(5), 1077–1088 (May 2012)
12. Vignati, A., Giannini, V., De Luca, M., et al.: Performance of a fully automatic lesion detection system for breast DCE-MRI. J Magn Reson Imaging 34(6), 1341–1351 (Dec 2011)

ISBN: 978-87-996443-0-8

A Texture Based Approach to Automated Detection of Diagnostically Relevant Regions in Breast Digital Pathology

Mohammad Peikari[†], Judit Zubovits[‡], Gina Clarcke[§], Anne Martel[†§]

[†]Medical Biophysics, University of Toronto, Toronto, Canada
[‡]Faculty of Medicine, University of Toronto, Toronto, Canada
[§]Physical Sciences, Sunnybrook Research Institute, Toronto, Canada

Abstract. Purpose: Digital whole-slide images (WSIs) are important sources of clinical information. They could be utilized to a much greater extent than they are being used today. However, there are still challenges to overcome, including the size of these images and the amount of information to be processed. In this work, we present an automated method to facilitate and accelerate the analysis of breast digital pathology images by reducing the size of the regions to be analyzed. Method: Our proposed automated classification method is to use texture features to identify relevant regions within tissue. The 'Bag of visual Words' method was employed to reduce the dimensionality of the feature vectors and abstract them into more compact forms. Finally, a linear support vector machine was trained to identify diagnostically relevant image patches. Result: The proposed method with the mean ROC area under the curve of 0.81($\pm$0.14) at 64 x 64 pixels tile size had a superior performance compared to a similar colour based method. Reducing the size of the regions to be analyzed had the largest impact on the performance of the method. Conclusion: Texture features can be used to automatically identify diagnostically relevant regions in digital WSIs.

1 Introduction

With the development of whole-slide scanners, digital microscopy images are produced in ever-increasing quantities. Biomedical image analysis focuses on retrieving quantitative clinically relevant knowledge contained within these images that might be beneficial for the patients' care.

Hematoxylin and eosin (H&E) stained pathology slides can provide knowledge for clinical decision and biomedical research but are currently being underutilized. Digital WSIs are usually several GB in size. Manual annotations of a dataset of this kind for diagnosis or research purposes could be laborious for a pathologist. Making use of the full potential is possible only when advanced computer methods and machine learning techniques are employed to automatically extract the most useful information from this vast amount of data.

In this work, we present an automated method to facilitate and accelerate the analysis of breast digital pathology images by automatically detecting diagnostically important patches of data and discarding the rest. This is done by

ISBN: 978-87-996443-0-8

finding appropriate representations of images, based on the texture variations in different tissue components and employing a suitable machine learning technique to understand the differences between the extracted visual patterns.

One of the specific applications of our method would be improving breast diagnosis systems, when the grade and stage of the cancer is to be determined from WSIs. Currently, pathologists manually find the areas where important diagnostic markers are present. For example, in breast cancer grading, diagnostic markers may include tubule formation, nuclear pleomorphism, and mitotic count [1]. Existing automated methods of assessing these features are very computationally intensive [2, 3]. Triaging the images first ensures that all areas without relevant information are excluded from analysis. Additional benefits of this approach include, speeding up digital pathology diagnosis systems, increasing throughput and accuracy of the semi-automated pathology cancer grading software, and improving the response time for interactive digital pathology workstations through managing their priorization and compression algorithms [5].

2 Related Works

Earlier studies tried to identify diagnostically relevant regions by looking at the intensity or colour values [4, 5]. Petushi *et al.* [4] aimed to identify different tissue components from within the images. The identified tissue components were fat cells, stroma, and three morphologically distinct cell types used in grading of cancer. This was done by segmenting grayscale images and extracting multidimensional representative features. Bahlmann's *et al.* [5] approach was to convert the H&E three-channel (RGB) patches to only two channels corresponding to H (purple/blue) and E (pink/red) colours. A 22 dimensional feature descriptor corresponding to 11 uniformly distributed percentiles of the cumulative histogram generated from each of H and E channels were generated for each image patch. A linear classifier was trained to classify the test image patches into diagnostically relevant or irrelevant regions. These methods maybe sensitive to the colour variations that may occur between image slides due to using different H&E colour staining protocols or different scanning devices which produce different qualities of images. We will compare our results with the method described in [5].

3 Methodology

In this study, diagnostically relevant regions are distinguished by a high density of epithelial or lymphocytic cells in any given image patch. This may include both malignant and benign tissue components corresponding to ducts or lobules. On the other hand, the diagnostically irrelevant regions are distinguished by the amount of fat-like or stroma tissues. The goal is to achieve the highest possible sensitivity while maintaining a relatively high specificity.

Our proposed automated identification aims to find the distribution of nuclei, fat-like and stroma components based on analyzing the structures each of these components present in pathology slides. Pathologists normally start with visual

ISBN: 978-87-996443-0-8

examination of digital slides at *low resolution* to identify the most diagnostically relevant regions, a process which is prone to missing small important parts of the tissue. In our automated method, we avoid this by subdividing the slides into small square patches at *full resolution*. We concentrate on the texture information to examine the possibility of developing a colour independent algorithm for pathology image analysis. Our proposed method proceeds with the following multi-step process: 1) image patches are converted to smaller tiles and texture features are extracted from them forming feature vectors, 2) Bag of visual Words method is employed to reduce the dimensionality of feature vectors and form histograms of words for each image patch, and 3) image patches are assigned with the correct class labels using a linear support vector machine classifier. Details of each step is explained in the subsequent sections.

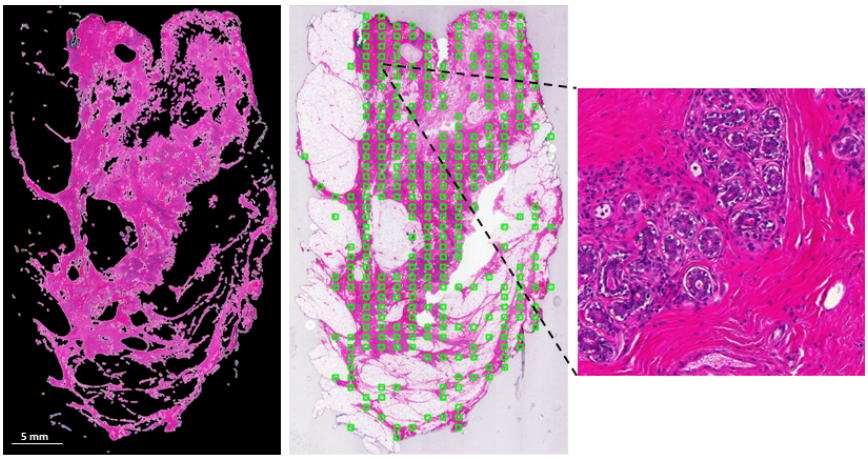


Fig. 1. Adaptive thresholding was used to remove large areas of fat and paraffin before patch selection (left), 512 x 512 pixel green box patches randomly placed but uniformly spaced on tissue images (middle), and a 512 x 512 pixel image patch (right).

3.1 Image Dataset and Data Collection

For this study, we have used wholemount H&E stained digital pathology slides from breast lumpectomy surgical specimens from 10 patients (that included 5 DCIS only and 5 mixtures of DCIS and IDC patients). The slides were prepared and scanned using the same methods explained in [9]. Briefly, they were scanned at 5X or 10X resolution by a TissueScope scanner (Huron Technologies International Inc., Waterloo, Canada), and sampled at 0.5 μm/pixel. A typical lumpectomy slide used in this study, with 15-20 cm^2, scanned at 10X is around 6GB (50,000 x 40,000 pixels).

ISBN: 978-87-996443-0-8

Patches of 512 x 512 pixels were collected from each slide by overlaying a grid of uniformly spaced but randomly placed boxes (Fig. 1). The horizontal and vertical distances between pairs of patches were 1000 and 500 pixels respectively. Adaptive thresholding was used as a preprocessing step to remove clearly irrelevant regions, such as large areas of fat and paraffin (Fig. 1). This reduced the number of patches the pathologist needed to review.

A graphical user interface was developed to capture the biological information within every patch. This interface allowed the expert pathologist to scroll through the images stored in the database and evaluate the presence of any information within the patches.

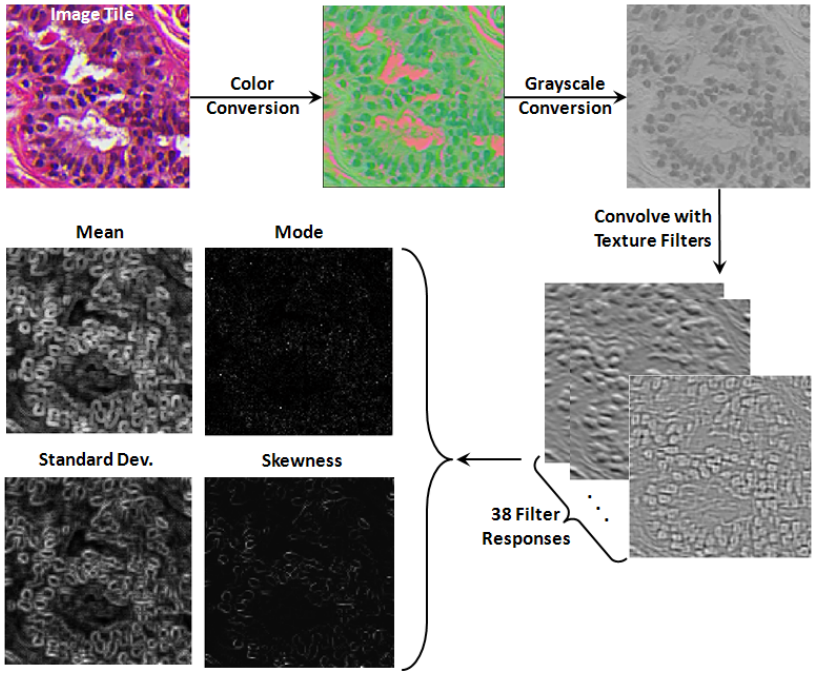


Fig. 2. Feature extraction pipeline: each image tile (128 x 128 pixels in this demonstration) is converted from RGB to La*b* colorspace and then converted to grayscale. The grayscale tile is then convolved with texture filters and further compressed into 4 statistical feature tiles (see the text).

3.2 Visual Pattern Mining and Extraction

The main goal of pattern mining is to find and extract useful visual information from image datasets that usually represents some interesting uncovered relationships among objects in the images. We concentrate on texture features in

ISBN: 978-87-996443-0-8

order to eliminate colour dependences in the analysis of pathology images. The diagnostically relevant regions usually contain higher variability in their texture due to aggregation of nuclei and lymphocytes. In order to find the textures in pathology images, the image patches were divided to smaller tiles, the tiles were converted to La*b* colorspace, and their grayscaled conversion were considered for further analysis. Images from La*b* space were converted to grayscale by eliminating the hue and saturation information while keeping the luminance. The contrast was increased between nuclei and background in grayscaled tiles and hence was more suitable for texture analysis, see Fig. 2. We employed the maximum response (MR) filter bank [6] to highlight textures in the grayscale tiles. The textures include regions where there is a sudden change in the intensities of the grayscale tiles. The MR filter bank consists of 38 filters: 2 anisotropic edge and bar filters in 6 orientations and 3 scales, and 2 rotationally symmetric Gaussian and Laplacian of Gaussian. The filters in the bank are convolved with all of the small image tiles (taken from large patches) making 38 filtered images per image tile, each highlighting textures with different scales in different directions.

Since the epithelial and lymphocytic components are relatively small and circular, their texture responses are maximum in all scales and directions of the filters in the bank. To capture the variability of the found textures, a set of pixel-wise statistical features were derived from all the filtered images: absolute mean, standard deviation, mode and skewness. This compresses the information relating to texture variability contained in the 38 filtered images into just 4 statistical feature tiles. In order to avoid some of the unwanted regions (noise) entering into the training pipeline and biasing our classifier training, a threshold was set on the tiles' texture response energy. The energy (absolute sum of mean responses) of the textures was high for the tiles where there was a high density of nuclei or lymphocytes. The threshold was optimized using preselected data and was kept constant for the rest of experiment.

Finally, in order to convert the feature images to numerical vectors, their Discrete Cosine Transform (DCT) coefficients were found. DCT image representation is a standard technique for data/image compression, which expresses an image using the sum of cosine functions at different frequencies. The top most DCT coefficients were considered in order to form a 364 dimensional feature vector per image tile.

3.3 Visual Content Representation Using Bag of Visual Words

In order to reduce the dimensionality of the feature vectors and abstract them into even more compact forms, we employ the Bag of visual Words (BoW) method. This technique allows for modeling complex image contents without explicitly considering the object models and their relationships [7]. This method has been previously used for machine-learning in medical image analysis and computer vision [7, 8].

BoW was used to learn discriminative models for automatic image analysis. The details of this algorithm are shown in Fig. 3. During the training stage,

ISBN: 978-87-996443-0-8

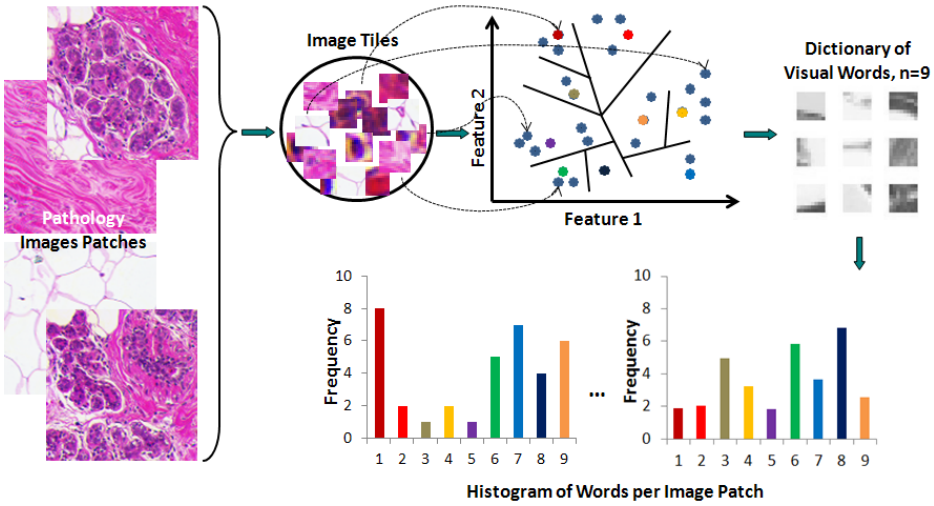


Fig. 3. Bag of visual Words content representation pipeline: image patches are split into smaller image tiles. Visual features are extracted from every image tile and mapped onto a multidimensional feature space (shown as 2D for simplicity). Centroids of the clustered feature space (coloured stars) are taken as the visual word dictionary elements. A histogram of words is generated using the feature dictionary for every image patch.

image patches are converted into tiles of a specific size depending on the extent of preferred isolation of visual patterns. Visual features corresponding to textures are retrieved from each tile. For any image tile, visual features are converted into a numerical feature vector as explained in the previous section. Therefore, every image patch is represented by a set of feature vectors depending on the number of tiles block dividing the image patch.

Feature vectors corresponding to all the patches are transformed into a multidimensional feature space and the space is clustered using a K-means algorithm to generate a feature dictionary. The centroids of the clusters represent a word in the feature dictionary. The feature dictionary summarizes the total number of feature types (words) to which every feature vector can be mapped. With the number of words and their spatial locations in the feature space, we can assign any new feature vector to a cluster, and find its representative word in the feature dictionary. The dimensionality of the feature vectors is reduced by generating histogram of words for each image patch in the dataset. In this way, we can distinguish among the image patches by comparing their histograms of words using any distance metrics. A classification model is trained to be able to distinguish the difference between relevant and irrelevant image patches.

ISBN: 978-87-996443-0-8

3.4 Classification of Diagnostically Relevant Regions

A dictionary of size $n = 25$ was generated during training and a linear support vector machine (SVM) was employed for the classification task. We employed the linear SVM algorithm in libsvm [10] and its parameter was optimized using one patient dataset for the penalty parameter of the error term [10]. The optimized parameter was tested on the remaining 9 patient dataset. Our method was trained and tested against a set of 1000 image patches taken from 10 patient dataset previously evaluated by an expert pathologist.

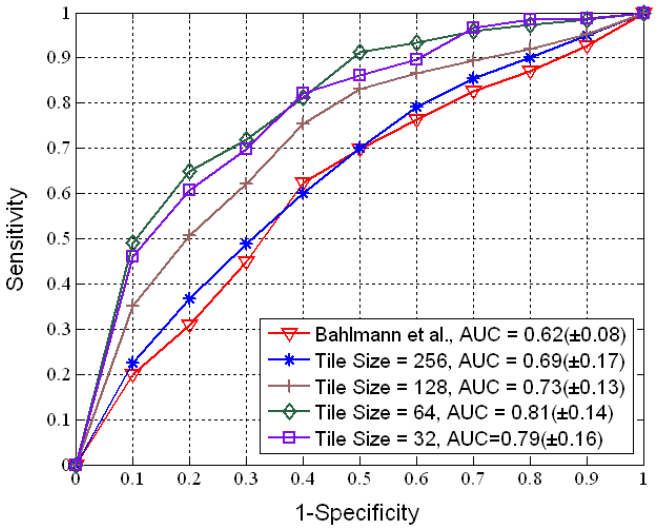


Fig. 4. ROC curves comparing different tile sizes, and the work presented in [5].

4 Results and Discussion

A 9-fold one-patient-out cross validation scheme was followed to validate our method. Fig. 4 shows the mean receiver operating characteristic (ROC) curve of our method at different tile sizes compared with the method explained in [5]. From Fig. 4, it is clear that our method with the mean area under the curve (AUC) of 0.81(±0.14) at 64 x 64 pixels tile size out-performed the work in [5] with a mean AUC of 0.62(±0.08). The tile size had the largest impact on the performance of the method. Using our method with the tile size of 64 x 64 pixels we can achieve a 91% sensitivity, while 62% of the total image patches will be kept for further analysis (by a computer algorithm for diagnostic assessment) and 38% will be discarded after classification.

ISBN: 978-87-996443-0-8

We found that, changing the dictionary size did not make a lot of difference (of the order ∼3%) to the classification results. This may be because there are only two classes no matter how many words in the dictionary are used to define them. In addition, the more we can isolate the patterns, the better the classification rates will be as shown in Fig. 4. Also in our study, image patches were selected in such a way that they may contain different extents of various tissue components. Our method has been shown to work well with these types of images that contain many classes of information. However, there might be some cases in which the epithelial component covers a very small portion of the patches, while a large portion is covered with irrelevant information (mainly stroma). In this case, using various patch sizes for training or a multiple instance learning method could further optimize the results.

5 Conclusion

To the best our knowledge there has been no previous work to automatically identify diagnostically relevant regions from digital WSIs using texture analysis in the images. Using texture analysis methods we can develop algorithms that are less sensitive to colour variations and therefore achieve good classification rates.

References

1. Bloom H. J. G., and Richardson, W. W.: Histological Grading and Prognosis in Breast Cancer: A Study of 1409 Cases of which 359 have been Followed for 15 Years. In: British Journal of Cancer, 359- 377, 1957
2. Lomenie N., and Daniel Racoceanu: Ontology-Enhanced Vision System for New Microscopy Imaging Challenges. In: Advances in Intelligent and Soft Computing, Vol. 120, 2012
3. Cosatto E. *et al.*: Grading Nuclear Pleomorphism on Histological Micrographs. In: The International Conference on Pattern Recognition, 2008
4. Petushi S. *et al.*: Automated Identification of Microstructures on Histology Slides. In: IEEE International Symposium on Biomedical Imaging, 424- 427, Vol. 1, 2004
5. Bahlmann C. *et al.*: Automated detection of diagnostically relevant regions in H&E stained digital pathology slides. In: SPIE Medical Imaging, Proc. SPIE Medical Imaging, Vol. 8315, 2012
6. Geusebroek, J.M. *et al.*: Fast Anisotropic Gauss Filtering. In: IEEE Transactions on Image Processing. 12(8):938-943, 2003
7. Cruz-Roa, A. *et al.*: pattern mining in histology image collections using bag of features. In: Artificial Intelligence In Medicine , Vol. 52, 2011
8. Varma, M., and Zisserman, A.: A statistical approach to material classification using image patch examplars. In: IEEE Transaction on Pattern Analysis and Machine Inteligence, Vol. 31, Is. 11, 2009
9. Clarke, G. M. *et al.*: Increasing specimen coverage using digital whole-mount breast pathology: Implementation, clinical feasibility and application in research. In: Computerized Medical Imaging and Graphics. Vol. 35, 2011
10. Chang, C.C. and Lin, C.J.: LIBSVM: A library for support vector machines. In: ACM Transactions on Intelligent Systems and Technology. Vol. 2, 2011

ISBN: 978-87-996443-0-8

A fully automatic lesion classification in breast ultrasound

Eugene Walach, Pavel Kisilev, Dan Chevion, Ella Barkan, Sivan Harary,
Sharbell Hashaul, Ami Ben-Horesh, and Asaf Tzadok,
IBM, Haifa Research Lab, Haifa, Israel

and Irith Hadas-Halpern and Irena Nikitin,
Radiology Department, Shaarei Zedek Hospital, Jerusalem, Israel

Abstract. In this paper we present a fully automatic lesion detection method trained to detect various lesions in breast ultrasound images, and to classify them as either benign or malignant. The automatic segmentation of lesions is based on the multi-modal appraoch combining Maximally Stable Extreme Regions (MSER) technique with posterior patches analysis. For each candidate, from the segmentation step, we calculate a set of novel robust descriptors. Based on these descriptors, we then train an Support Vector Machine (SVM) classifier to distinguish between the two classes of lesions. We tested our method on a set of over 400 diffrent cases (single image per case) collected from different acquisition devices. We proved the robustness of the proposed method by showing high precision-recall figures despite the wide variety of the data acqusition devices.

Keywords: Breast ultrasound, automatic lesion detection

1 Introduction

Worldwide, breast cancer comprises just under 30% of all diagnosed cancers in women, making breast cancer the second leading cause of death for women. Mammography is currently the most common modality for screening and detecting breast cancer. However, a large portion of the breast lesions found in mammograms is benign. To improve the specificity, doctors often examine the suspicious lesions using ultrasound (US) imaging. Moreover, ultrasound is known to increase cancer detection sensitivity in particular for women with dense breasts. Accordingly, a great deal of effort has been devoted to improving breast cancer diagnostic tools. A technological review of commonly used methods and new experimental techniques was conducted in [1]. The improvements to diagnostic tools can be divided into two categories: enhancements to the imaging equipment and the introduction of computerized image analytics. In this paper we deal with the second category, namely, the computer-aided diagnosis (CAD) system. A survey of CAD systems for breast US was conducted in [2]. Some encouraging results have been reported for automated analysis of 3D images [3]. However, since conventional B-scan systems are, by far, the most prevalent ones

ISBN: 978-87-996443-0-8

in the field, we have chosen to focus on CAD systems for B-scan images. These systems typically compute a variety of breast image features and use a variety of classification techniques to distinguish between malignant and benign tumors [4, 5]. These features include the shape of the tumor, its texture, and sometimes acoustic properties. Unfortunately, in many cases, the efficacy of US CAD systems tends to be limited due to high dependency on the specific image acquisition system.

Our work aims to mitigate such difficulties by focusing on a robust set of features (including internal calibration) and on the addition of recently introduced image attenuation features [6]. To test these concepts we have created a fully automatic image analysis system. The system receives, as an input, an ultrasonic breast image. The tumor is located. Then a set of 49 robust features (including both features mimicking human physicians and attenuation features that are invisible to the human eye) is extracted. Finally, a feature vector is fed into a machine learning classifier (SVM based), which provides a final determination of the malignant tumors. We tested the system on a benchmark of 400 independent cases having almost equal numbers of malignant and benign tumors. To verify system robustness we worked with images coming from four different hospitals using five different image acquisition systems. The system performance of the system on the above benchmark was 73% .

This paper is organized as follows. In Section 2, we describe our automatic image segmentation algorithm. Section 3 presents the proposed descriptors calculated for each one of the candidates found by the segmentation method. Section 4 describes our experimental results and conclusions.

2 Multi-modal approach to automatic segmentation

Watershed based segmentation techniques are quite popular in the computer vision (see for example [7]). Maximally Stable Extremal Regions (MSER) ([8]) technique can be viewed as an variant of the watershed approach. In MSER segmentation, we seek a range of thresholds that leaves the watershed basin effectively unchanged. In other words, MSER determines regions based solely on an extremal property of the intensity function in the region and on its outer boundary.

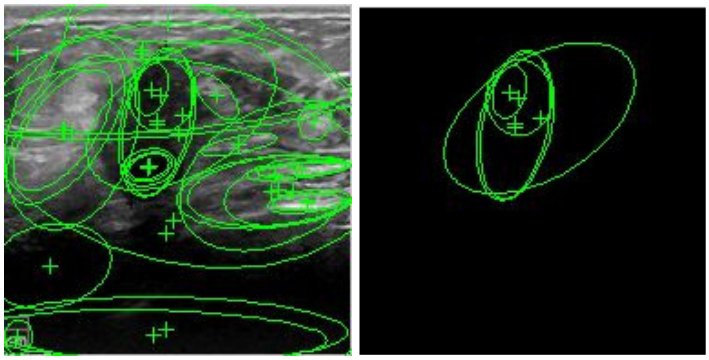
First stage of the process is that of image enhancement and pre-processing. It performs two functions: Time-Gain compensation and speckle removal.

Then MSER procedure yields blobs of pixels that represent candidates for further classification. Figure 1 on the left depicts MSER detected blobs. For display purposes, each blob candidate is represented by the bounding ellipse. Next, we need to filter out most of the candidates and leave only the most relevant ones. For each candidate, we compute its center of mass, its area, its mean and standard deviation of its intensities, the length of its eigenvectors for the pixel distribution, and its number of close neighbors. We then train the Gaussian mixture model and use it in order to determine the most probable

ISBN: 978-87-996443-0-8

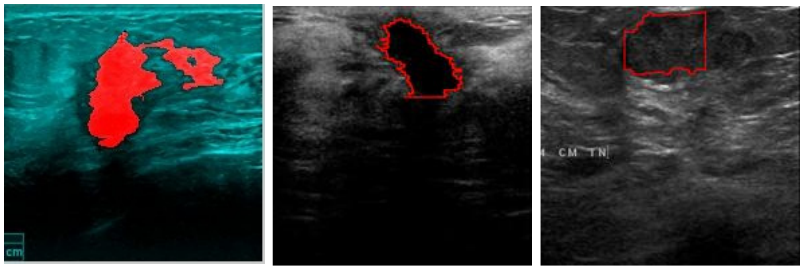
candidates. Figure 1 on the right shows the most likely candidates chosen by our method. The result of the final segmentation is shown in Figure 2 on the left.

Unfortunately, in many practical cases, there is no gray-level value that would define tumor border. In such cases classical MSER approach is bound to fail. Accordingly, in order to augment MSER results, we have applied a complementary approach. For that purpose, we analyze the entire image finding location of the breast tissue and estimating its median gray level. Then we identify patches of posterior attenuation (or amplification) that indirectly indicate tumor presence. The Figure 2 depicts two examples of successful application of this approach. Figure 2b (at the center) shows tumor identified due to its posteriorattenuation and Figure 2c (on the right) shows the case of posterior amplification.



(a) (b)

Fig. 1. Left: Several candidates detected by MSER algorithm; Right: the best candidate chosen.



(a) (b) (c)

Fig. 2. Segmentation results: a) MSER based segmentation, b) tumor detected based on the posterior attenuation, c) tumor detected based on posterior amplification.

ISBN: 978-87-996443-0-8

4 ...

3 Feature Extraction

In this section we describe the features that are extracted from the segments which were detected by the automatic segmentation method described in the previous section. These features are divided into two categories: the first is designed to mimic the decision-making process performed by a radiologist. The second is designed to add additional information that is not accessible to the radiologist. These features are based on estimating the acoustic attenuation. In the following subsections we describe both those categories in detail.

3.1 BI-RADS Features

When radiologists examine an US image of a breast tumor they grade it using BI-RADS (breast imaging - reporting and data system). In US the following characteristics of the mass are examined: shape, acoustic transmission (posterior enhancement/shadowing), margins, and echogenicity.

Shape The shape of the mass is the most important characteristic examined by radiologists. Malignant tumors tend to have more irregular and lobular shapes. To assess the shape of the mass we first used basic features that are the area of the mass and its height to width ratio. To keep this ratio in the range $[0, 1]$ we defined it as follows:

$$\begin{cases} H/2W, & H < W \\ 1 - W/2H, & W > H \end{cases} \tag{1}$$

where W,H are the mass width and height respectively. If the ratio is close to 0, then the mass is a long horizontal ellipse, which is usually the case with benign masses. Ratio close to 1 means a long vertical ellipse, which often implies malignancy.

Another shape feature is the curvature along the mass boundaries. A larger curvature indicates either that the mass is small or that it is lobular. Therefore we compare the curvature to that of a circle with the same perimeter as the mass.

The remaining six shape features are calculated by fitting an ellipse to the mass borders. The features that are calculated using this ellipse are: the ellipse orientation, the ratio between the minor axis and the major one, and the difference between the mass border and the ellipse. We calculate four different types of differences: RMS, L1, the maximal error, and the ratio between the perimeters. A larger difference implies irregular shape, which may indicate malignancy.

Acoustic transmission The posterior of the mass is an important characteristic when assessing the risk for malignancy. Strong enhancement and edge shadowing are common in benign masses (such as cysts), while posterior shadowing is common in malignant tumors. To assess the level of the posterior enhancement/shadowing, we examine the area below the mass. We refer to three

ISBN: 978-87-996443-0-8

different areas in the image, as presented in Figure 3. First, we determine rectangular area below the tumor. Its height is equal to that of the tumor and its width is equal to that of the tumor. Center 60Remaining 40% are split between the two edge shadowing areas (marked by the magenta lines). The green lines mark the areas that are used for comparison.

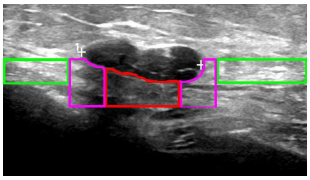
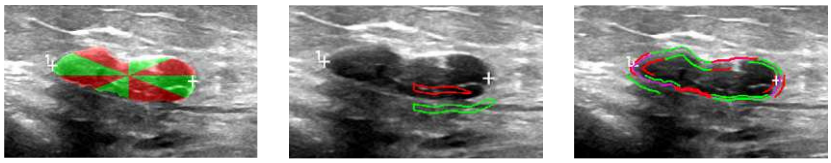


Fig. 3. Acoustic transmission: the red line marks the posterior area, the magenta marks the areas of possible edge shadowing, and the green marks the side area used as reference.

Margins Sharp margins can indicate a benign tumor and visa versa. To assess the sharpness of the boundaries, we divided the mass into 8 sectors of 45 degrees each, as shown in Figure 4(a). For each sector, we have computed two sharpness features. Using erosion and dilation of the segment mask, we found areas inside and outside the mass and computed the corresponing mean intensity values. Figure 4(b) depicts an example of such areas for one of the sectors. Sharpness was measured by the distance (in pixels) between the 10% and 90% lines, and the slope of the margin which wasdefined as intensity difference divided by the distance in pixels.



(a) (b) (c)

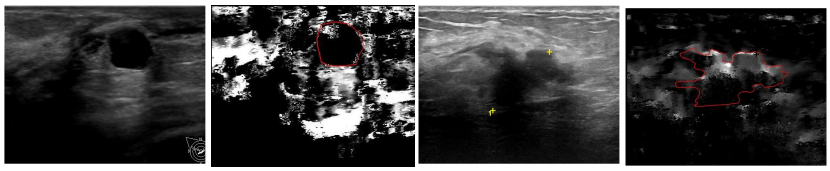
Fig. 4. Margins sharpness. (a) The sectors. (b) The areas used to determine the inside and outside intensities of the sector. (c) The 10% and 90% lines for each sector.

Echogenicity Another important characteristic of masses that is examined by doctors is their echogenicity. However, since ultrasonic images are inherently qualitative (rather than quantitative), one can not evaluate echogenicity based on the echo amplitude values only. Accordingly, echogeneicity is determined by comparing tumor echo levels to that of the reference fat tissues.

3.2 Attenuation Features

Studies such as [9, 10] show that acoustic attenuation measurements can distinguish between malignant and benign tissues, and can therefore be used as an

ISBN: 978-87-996443-0-8



(a) (b) (c) (d)

Fig. 5. Examples of attenuation images: a) original US image of the benign tumor, b) attenuation map of the image a, c) original US image of the malignant tumor, d) attenuation map of the image c. Image d features high attenuation levels with high degree of in-homogeneity..

effective basis for a CAD system. Tissue attenuation can be calculated using transmission US in a tomographic manner [11]. However, clinical US systems produce B-scans, which are based on backscattering rather than transmission. Consequently, authors of [12] developed an alternative approach transforming B-scans into the corresponding attenuation images. In [6] it was demonstrated that this approach can facilitate detection of malignant breast tumors. Figure 5 illustrates this approach. On the left raw benign image is depicted followed by its attenuation map. Fig 5c and d present raw and attenuation images of a malignant tumor. In general, malignant tumors feature high attenuation values and heterogeneous structure. In this paper, we adopted the same approach for creation of attenuation maps. However, we utilized a different set of attenuation-based features, in combination with more conventional image-based features, to differentiate between benign and malignant tumors. As explained in [6], since our system works with images processed by the Time-Gain compensation curve, our method produces relative attenuation values (rather than absolute ones). Hence, a typical tumor will include both positive and negative attenuation values. To express intra-tumor distribution of the attenuation values we have computed two vectors: vertical vector V_{ver} and circular vector V_{circ}. Each of these two vectors had 9 components (features). Vector V_{ver} has been created by computing median values of positive attenuation pixels in each line of the tumor image. In turn vector V_{circ} has been computed via the sequential removal of tumor layers and by computing the median values of all the positive pixels inside each layer. Both vectors have been normalized and used in the classification process. Generally speaking vector V_{ver} represents vertical attenuation distribution perpendicular to the direction of the tissue layers, which tend to have a horizontal nature. High attenuation values in the vicinity of the posterior tumor wall tend to be associated with malignant tumors. In turn vector V_{circ} depicts a radial tumor structure so that identification of the characteristic dome-type structure facilitates detection of benign calcifications. In addition, we have used two more global tumor features. The first one represents the radius of the maximal blob having homogeneous attenuation (the presence of such a blob tends to indicate malignancy). The second global feature represents the % of pixels having neg-

ISBN: 978-87-996443-0-8

ative attenuation (a high % indicates that a given tumor has low attenuation raising the probability of the cyst).

In the above description, we have emphasized the intuitive meaning of each feature. However, in reality all the features (49 overall: 9 shape, 3 Transmission, 16 margins, 1 echogeneicity and 20 attenuation) are used as an input to the automatic classifier. Hence, exact location of the multi-dimensional decision planes cannot be easily visualized.

4 Experimental results

We tested the performance of our system on a set of 436 different cases (one ultrasound image for each case) from five different acquisition devices with significantly different characteristics (e.g. resolution, dynamic range, compression mechanism etc.). The ground truth for each image was determined by the pathology results. First, we tested the discriminative ability of our features. For that purpose we have calculated our features for manual segmentation created by radiologists. Then We tested the performance of the system using automatic segmentation results.

We randomly divided our benchmark into training and testing subsets, then performed a complete training and testing cycle. We have repeated this experiment 100 times and averaged the performance of our system over 100 such independent runs. This experiment shows the level of adaptivity and the average performance of our method.

The binary classification into one of the classes, that is, benign or malignant, using 'manual' masks yields 87% classification accuracy; this is the upper bound on the overall system performance. The results of the similar test using automatic tumor detection are summarized in Figure 6. It shows ROC curves, that is, false negative versus false positive errors, of 100 random subsampling trials, and its averaged ROC curve for benign/malignant classification. The equal error rate of the binary classification into benign or malignant classes, using 'automatic' masks yields 73% classification accuracy. Although, it is about 14% lower than the 'manual' based classification, it is still a reasonable classification accuracy for a fully automatic system.

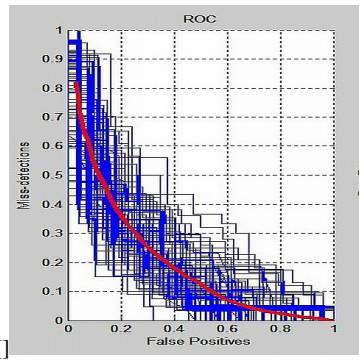
We conclude that it is possible to build a robust CAD system for ultrasonic breast images capable of reliable detection of malignant lesions and able to work with different image acquisition systems. However, additional work is necessary to verify these results on a larger data set.

References

1. Nover, A.B., Jagtap, S., Anjum, W., Yegingil, H., Shih, W., Shih, W., Brooks, A.D.: Modern Breast Cancer Detection: A Technological Review. International Journal of Biomedical Imaging (2009)
2. Chenga, H.D., Shana, J., Jua, W., Guo, Y., Zhang, L.: Automated Breast Cancer Detection and Classification Using Ultrasound Images: A Survey. Pattern Recognition 43(1) (2010) 299–317

ISBN: 978-87-996443-0-8

8 ...



[]

Fig. 6. False negative vs. false positive errors for benign/malignant binary classification experiment. Blue curves represents separate random experiments. Red curve represents an average of 100 random experiments.

3. Chen, D., Hsiao, Y.: Computer-aided Diagnosis in Breast Ultrasound. Elsevier Journal of Medical Ultrasound 16(1) (2008) 46–56
4. Cui J, Sahiner B, C.H.N.A.P.C.H.L.Z.C.S.J.: A new automated method for the segmentation and characterization of breast masses on ultrasound images. Med Phys 36(5) (2009) 1553–1565
5. Joo S, Yang YS, M.W.K.H.: Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features. Trans Med Imaging 23(10) (2004) 1292–300
6. Sivan Harary, E.W.: Identification of malignant breast tumors based on acoustic attenuation mapping of conventional ultrasound images. Proceedings of MICCAI-MCV (2012) 168–177
7. Shih, F.Y., Cheng, S.: Automatic seeded region growing for color image segmentation. Image and Vision Computing 23(10) (2005) 877–886
8. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing 22(10) (2004) 761–767
9. Goss, S.A., Johnston, R.L., Dunn, F.: Compilation of empirical ultrasonic properties of mammalian tissues II. Journal of The Acoustical Society of America 68(1) (1980) 93–108
10. Chang, C.H., Huang, S.W., Li, P.C.: Attenuation Measurements for Ultrasonic Breast Imaging: Comparisons of Three Approaches. IEEE Ultrasonics Symposium (2008) 1306–1309
11. Li, C., Duric, N., Huang, L.: Breast Imaging Using Transmission Ultrasound: Reconstructing Tissue Parameters of Sound Speed and Attenuation. BMEI (2008)
12. Walach, E., Shmulewitz, A., Itzchak, Y., Heyman, Z.: Local Tissue Attenuation Images Based on Pulsed-Echo Ultrasound Scans. IEEE Trans. Biomed. Eng. 36(2) (1989) 211–221

ISBN: 978-87-996443-0-8

Automatic Detection of Architectural Distortion in Mammograms using Sparse Overcomplete Dictionaries of a Curvelet Descriptor

Fabián Narváez, Cesar Poveda*, and Eduardo Romero

Computer Imaging and Medical Application Laboratory - Cim@Lab,
Departament of Medicine,
Universidad Nacional de Colombia, Bogotá, Colombia
* Instituto Nacional de Cancerología, Bogotá, Colombia
http://www.cimlaboratory.com

Abstract. Mammography aims at the detection of early breast disease stages, a condition for which the radiological illness signs are usually blurred. In particular, the architectural distortion, a common cause of false-negative on mammograms, arise as the change of normal oriented texture of the breast, which is usually hidden by the intricate breast anatomy. This paper presents a novel detector of Regions of Interest (RoIs) with a certain degree of architectural distortion, i.e., an associated normal control (NC) or architectural distortion (AD) membership value. The approach carries out a multiscale image decomposition using the curvelet transform and representing the marginal curvelet subband as the parameters of a Generalized Gaussian Density (GGD) that approximates the subband coefficient distribution. A set of training images serves to construct NC and AD dictionaries in that multiscale space from randomly selected RoIs. A mammogram is decomposed into overlapping patches that are projected to the multiscale space, reconstructed with the two dictionaries and assigned to any of the two classes with the smallest ℓ_2 norm of the reconstructing vector α. The membership function is computed as the Kullback−Leibler distance between any multiscale represented RoI patch and the closest basis of the dictionary, set by the largest projection coefficient in the α vector. We demonstrate the effectiveness of the proposed descriptor by classifying AD in a set of 19 RoIs (4269 patches) of the MIAS database and we obtained a sensitivity and specificity rates of 65 % and 82 %, respectively while the number of false positives per image was 0.9.

Keywords: Breast Cancer, BI-RADS, Architectural Distortion, Curvelet transform, Kull-back Leibler Divergence

1 Introduction

Breast cancer is fully tractable if diagnosis is achieved early, reason for breast cancer screening programs have been introduced in most health systems. Mammography, as part of these screening programs, is considered the most cost-effective method for visualizing abnormalities in the very early stages [21, 5].

ISBN: 978-87-996443-0-8

Mammography allows detection of some abnormalities such as calcifications, bilateral asymmetry, masses and architectural distortion [5], being architectural distortion the third most common mammographic finding. This breast imaging term is associated to parenchymal asymmetry and usually amounts to any change in the usual parenchymal pattern. This includes spiculations radiating from a point and focal retraction or distortion at the edge of the parenchyma [4]. Likewise, focal retraction is considered as easier to perceive than spiculated distortion and could be categorized as malignant or benign. However, mammographic interpretation is a really difficult examination, with high intra and inter-observer variabilities. Studies have shown that between 10% and 25% of breast cancers are not detected [4]. Scarring from previous surgery, radial scar and crossing breast structures are usually misdiagnosed as architectural distortions. Architectural distortions have indefinite characteristic features, e.g. number of spicules, width and length of spicules and density. An agreement to reduce such variability resulted in the *Breast and Imaging Report and Database System* (BI-RADS), designed by the the American College of Radiology , as a standard description to report breast lesions and categorize different pathologies as well as their severity level [17].

Automatic aid diagnosis systems for mammography categorization is becoming largely used in actual clinical scenarios to complement the clinician first diagnosis impression. Computer Assisted Diagnosis Systems (CAD) have decreased variability diagnosis, becoming a well accepted clinical practice to assist radiologists interpreting mammograms when they search micro-calcification clusters [13]. However, architectural distortion is still a very open problem. Several automated methods for AD detection have been reported, all of them relying on the fact that AD is a group of aligned structures with different orientation. RoIs have been characterized [11, 3, 20, 18, 10] by segmenting the breast, sharpening the spicules, filtering in the Radon domain, extracting the mean direction, filtering with a Bank of Gabor filters, extracting potential AD with the phase portrait model or the fractal dimension and calculating the Haralick's texture measures. All these approaches use a Support Vector Machine to set an optimal classification threshold. In spite of the multiple techniques, the proposed solutions are still unsatisfactory. There is either a low sensitivity or a high number of false positives per image (FPI). Ichikawa et al. [11] reported a sensitivity for detection of only 56 % while the number of false positives per image obtained by Sampat et al. [20] was 14. The results of Rangayyan et al. [18] showed a sensitivity of 84 % but with a high false-positive rate of 7.4 per image. The commercial CAD systems [9] correctly marked a 90 % of mass and calcification cases, whereas the sensitivity of AD detection was reported to be only of 40 %. The R2 Image Checker system [7] successfully identified 49 % cases containing AD. The CADx Second Look system [1] successfully detected 33 % cases of AD with 1.27 false positives per image [19]. Additionally, only 48 - 60 % of ADs that are biopsied are found to be cancer [19].

In this paper, we proposed a new approach to detect and discriminate AD lesions in a mammogram. Given a particular mammogram, the method obtains

ISBN: 978-87-996443-0-8

a normalized mammogram membership function, which is thresholded to determine to which class a RoI belongs. This function is found by firstly splitting the mammogram into a set of overlapped patches, which are projected to the curvelet space. The subband coefficients are well approximated by the GGD parameters, achieving thereby a considerable dimensionality reduction, i.e., a subband from a 64×64 patch, usually composed of 60 coefficients, is described by two parameters. A particular patch is then set to any of the two classes by the smallest ℓ_2 sparse reconstructing vector. Once the class has been assigned, the patch is characterized by the Kullback−Leibler distance of the patch to the basis dictionary with the largest sparse representation coefficient. Finally, the membership function is obtained by normalizing the distances associated to any patch of the mammogram.

The rest of this article is organized as follows: after this introduction, next section presents the methodology, then results are shown and last section discusses future work and conclusions.

2 Methodology

2.1 The method

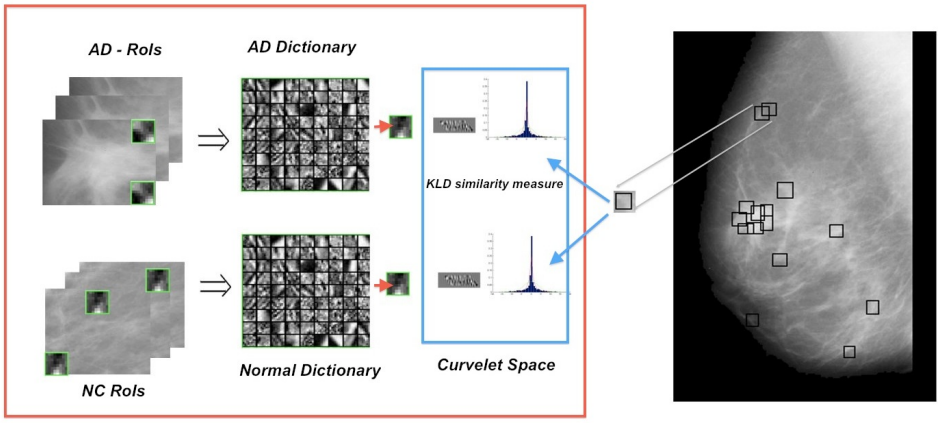


Fig. 1. Method overview: red block shows the process of construction of dictionaries (AD-NC), and automatic detection of architectural distortion based on a curvelet descriptor of different patches

The whole method searches to characterize a particular RoI by projecting the mammogram data to the curvelet space, thereby improving data sparsity. Two overcomplete dictionaries also projected to the curvelet space, representing the AD and NC classes in the curvelet space, serve as a reference conceptual

ISBN: 978-87-996443-0-8

frame (see figure 1), i.e., the minimal distance to each of the two dictionaries will define the classification problem. The curvelet sparsity is then additionally used to reduce dimensionality by approximating the subband coefficients with two parameters of a Generalized Gaussian Distribution. The result is that the whole curvelet decomposition, with for instance four scales and eight directions, is represented as a vector with 32 values. Provided that each curvelet sub-band is approximated by a probability distribution function, The Kullback−Leibler divergence is used as the metrics to introduce a distance notion between a RoI and each of the spaces spanned by the dictionaries. Once the RoI is characterized by the set of couples (μ, σ) representing the subbands of the curvelet decomposition, the Kullback−Leibler distance to the closest basis of the two dictionaries defines a membership function that will define the particular region classification.

2.2 RoI Pre-processing

Every image was herein stretched to the maximum and minimum gray level values ($[0, 255]$), but adaptively equalizing the histogram so that structural details were preserved. Resultant images were smoothed out by a median filter to remove the remaining noise [16].

2.3 The curvelet transform: statistical characterization

The curvelet transform is a multiscale decomposition [6] with a set of redundant bases which optimally represent 2D curves. In addition to the usual information about scale and location, already available from a wavelet, each of these curvelet bases is able to capture information about orientation while fulfilling a parabolic anisotropic scaling law $width \approx length^2$, that assures curves at different scale levels conserve their geometrical relationships [6]. A curvelet in the frequency domain, is constructed as the product of two windows: the angular and the radial dyadic frequential coronas. The angular window corresponds to a directional analysis, i.e., a Radon transform, and the radial dyadic window is a bandpass filter with cut frequencies following the parabolic anisotropic scaling law [6].

In general, the subband curvelet coefficient distribution is characterized by a sharper peak at zero with smooth tails. This leptokurtic pattern is associated to the sparse property of this transformation, i.e., few coefficients have high probability of being used to represent. It has been shown that the parameters of a generalized Gaussian density can fully describe the marginal distribution of the subband curvelet coefficients. The GGD reads as

$$p(x; \theta, \gamma) = \frac{\gamma}{2\theta\Gamma(1/\gamma)} e^{-(|x|/\theta)^{\gamma}}$$

where $\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1} dt$, $z > 0$ is the Gamma function, θ is the variance and γ is related to the decreasing rate of the GGD. The parameters θ and γ are estimated from the subbband data through Maximum Likelihood, as detailed in [8]. The parameters (θ,γ) may be used as descriptor of the probability density function of the energy levels inside each curvelet subband.

ISBN: 978-87-996443-0-8

2.4 Similarity measure

The similarity between subband curvelets is measured using the Kullback-Leibler divergence (KLD) of the corresponding GGDs:

$$D(p(.;\theta_1,\gamma_1)||p(.;\theta_2;\gamma_2)) = \log\left(\frac{\gamma_1\theta_2\Gamma(1/\gamma_2)}{\gamma_2\theta_1\Gamma(1/\gamma_1)}\right) + \left(\frac{\theta_1}{\theta_2}\right)^{\gamma_2}\frac{\Gamma((\gamma_2+1)/\gamma_1)}{\Gamma((1/\gamma_1)} - \frac{1}{\gamma_1}$$

where (θ_1,γ_1) and (θ_2,γ_2) are the GGD parameters estimated for each subband. This metric does not require additional normalization and shows good performance in other multiscale domains [8].

2.5 Dictionary Construction

The next step is to build dictionaries $\mathbf{D}_{AD}$ and $\mathbf{D}_{NC}$ for the architectural distortion and normal controls, respectively. For doing so, it was selected a set of N RoIs with different architectural distortion lesions or normal controls. Then, a set of K 64×64 patches were randomly selected from each RoI and transformed to the curvelet space. Each subband of any patch is approximated by the couple of the GGD parameters so that the patch is characterized by a vector of couples. Finally, the characterized patches are stored as columns of the Dictionaries $\mathbf{D}_{AD}$ and $\mathbf{D}_{NC}$.

2.6 Class Characterization

Sparse representations attempt to identify the constituent parts of an object and use them for reconstructing any scene with a certain accuracy. These parts, denoted as basis functions or atoms, are usually arranged in overcomplete dictionaries with a larger number of elements than the effective dimensionality of the input space, thereby representing a wider range of image phenomena [15, 14, 12]. Formally, consider a $n\times m$ matrix $\mathbf{D}$, where each column is a possible image in $\mathbb{R}^n$ (atomic images), a dictionary of atoms. The projection of an image x onto the space spanned by $\mathbf{D}$ yields a weighting vector α ($x=\mathbf{D}\alpha$). Furthermore, if α is sparse (with $k_0 \ll m$ nonzeros), this produces a linear combination of k_0 atoms with varying weights. To find the adequate α, the optimization problem to solve is $\mathcal{G}_1(\mathbf{D},x,\lambda)$, which has the form

$$\mathcal{G}_1(\mathbf{D},x,\lambda):\ \min_{\alpha}\lambda\|\alpha\|_1^1+\frac{1}{2}\|x-\mathbf{D}\alpha\|_2^2$$

The solution consists in finding the sparsest vector α that weights x as a linear combination of atoms from $\mathbf{D}$, using the norm ℓ_1 as a measure of sparsity.

After a random selection of patches from two different types of RoIs, namely architectural distortion and normal controls, two different dictionaries $\mathbf{D}_{AD}$ and $\mathbf{D}_{NC}$ were obtained. To identify whether a particular patch is architectural distortion or not, the patch is firstly characterized by the set of parameters of each of the curvelet subbands. The patch feature vector is then reconstructed using

ISBN: 978-87-996443-0-8

both dictionaries and assigned to the class with the smallest ℓ_2 reconstructing vector. Afterward, the patch is characterized with the distance between its curvelet representation and the most important atom for the reconstruction in the curvelet space. This distance is then normalized with respect to the distances of every patch in the image and such patch classified after a simple threshold. For the sake of classification, this threshold was herein set to 0.62 and found by automatically adjusting the minimum number of False Positives in the ROC curve.

3 Experimental results

3.1 Experimental setup

A total of 38 mammograms, 19 containing AD lesions, extracted from the *Mini-iMias Database* [2] were used for the evaluation. Images were curvelet transformed, with 4 scales and 32 orientations, resulting in a descriptor with 66 subbands for each patch. Dictionaries were constructed by randomly sampling 250 overlapped patches from a set of ten images containing the AD lesion, each patch having 64×64 pixels. Evaluation was then performed in the remaining set of 9 mammograms.

3.2 AD detection and False Positive minimization

Once the membership function is computed for the whole mammogram, a simple threshold defines the patch classification. This threshold is automatically determined by setting a value which minimizes the number of False Positives in the ROC curve with the restriction that the area under the curve should be maximal.

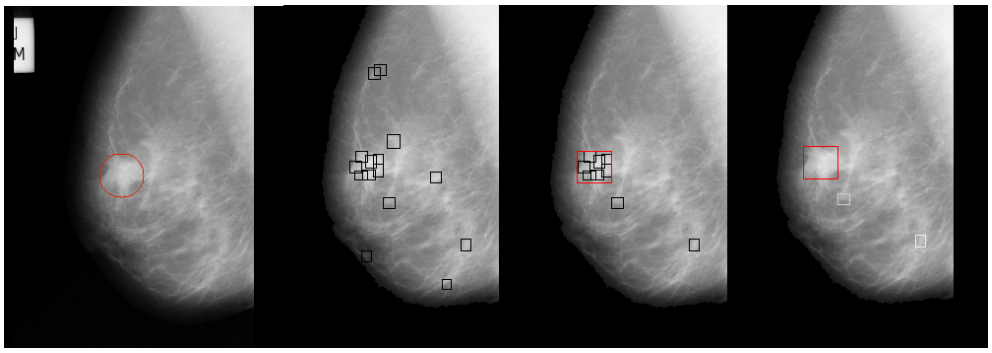


Fig. 2. Left panel shows the selected AD while the sequence of panels illustrate different numbers of False Positive after setting the binary membership function to several thresholds

ISBN: 978-87-996443-0-8

3.3 Sensitivity and False Positive Rate

Sensitivity, specificity and a minimal number of False Positives was determined in an independent set of 9 images

Table 1.

	Sens.	Spec.	NFP
AD	65	82	0.9
NC	86	92	0.93

4 Conclusion and future works

We have introduced a new descriptor for AD lesions, based on curvelets and a statistical model of the curvelet coefficients for RoIs extracted from mammography images. By applying the curvelet transform and adjusting the levels of energy for each subband to a generalized Gaussian model, we obtained a robust representation which captures the edge distribution at different orientation and scales. Experimental results indicate that the new feature improves detection and classification performance. Future works include improving the feature with invariance to rotation and scale and extensive experimentation in large mammograms databases.

5 Acknowledgment

This work was partially funded by the Ecuadorian government through " La Secretaria Nacional de Ciencia y Tecnología y Educación Superior (SENESCYT)", Grant: 20110958 "CONVOCATORIA ABIERTA 2011"

References

1. (2012), http://www.icadmed.com
2. et al, J.S.: The mammographic image analysis society digital mammogram database exerpta medica. vol. 1069, pp. 375–378 (1994)
3. Ayres, F.J., Rangayyan, R.M.: Characterization of architectural distortion in mammograms. IEEE Enginering in Medicine and Biology Magazine pp. 59–67 (2005)
4. Bird, R., Wallace, T., Yankaskas, B.: Analysis of cancers missed at screening mammography. Radiology 178, 234–247 (1992)
5. Buseman, S., Mouchawar, J., Calonge, N., Byers., T.: Mammography screening matters for young women with breast carcinoma. Cancer 97, 352–358 (2003)
6. Candes, E., Demanet, L., Donoho, D., Ying, L.: Fast discrete curvelet transforms. Multiscale Modeling and Simulation 5(3), 861–899 (2006)
7. Checker:, T.R.T.I.: Image checker (2012), http://www.r2tech.com

ISBN: 978-87-996443-0-8

8. Do, M., Vetterli, M.: Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance. IEEE Transactions on Image Processing 11(2), 146–158 (2002)
9. Garvican, L., Field, S.: A pilot evaluation of the r2 image checker system and users' response in the detection of interval breast cancers on previous screening films. Clinic Radiology 56(10), 833–837 (2001)
10. Guo, Q., Shao, J., Ruiz, V.: Investigation of support vector machine for the detection of architectural distortion in mammographic images. Journal of Physics: Conference Series(15) (2005)
11. Ichikawa, T., Matsubara, T., Hara, T., Fujita, H., Endo, T., Iwase, T.: Automated detection method for architectural distorion areas on mammograms based on morphological processing and surface analysis. In: Processing, I. (ed.) Medical Imaging. vol. 5370 (2004)
12. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE-Transactions on signal processing 41(12), 3397–3415 (1993)
13. Nishikawa, R.M.: Current status and future directions of computer-aided diagnosis in mammography. Computerized Medical Imaging and Graphics 31, 224–235 (2007)
14. Olshausen, B.: Principles of image representation in visual cortex. No. 1603-1615, MIT (2003)
15. Olshausen, B.A., Field, D.J.: Sparse coding whit an overcomplete basis set: A strategy employed by v1? Vision Res. Pergamon 37(23), 3311–3325 (1997)
16. Petrick, N., Chan, H.P., Sahiner, B., Wei, D.: An adaptive density weighted contrast enhancement filter for mammographic breast mass detection. IEEE Trans. Med. Imaging 15 1, 59–67 (1996)
17. of Radiology (ACR), A.C.: Illustrated Breast Imaging Reporting and Data System (BI-RADS). ACR (1998)
18. Rangayyan, R.M., Prajna, S., Ayres, F.J., Desautels, J.E.L.: Detection of architectural distortion in prior screening mammograms using gabor filters, phase portraits, fractal dimension and texture analysis. Int J CARS. Springer (2008)
19. Rangayyana, R.M., Ayres, F.J., Desautelsa, J.E.L.: A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. Journal of the Franklin Institute 344, 312–348 (2007)
20. Sampat, M.P., Whitman, G.J., Markey, M.K., Bovik, A.C.: Evidence based detection of spiculated masses and architectural distortions. In: Medical Imaging. Image Processing, vol. 5747, pp. 26–37. SPIE (2005)
21. Verma, K., Zakos, J.: A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. IEEE Transactions on Information Technology in Biomedicine 16, 219–223 (2002)

ISBN: 978-87-996443-0-8

Learning to detect lesion boundaries in breast ultrasound images

Pavel Kisilev*, Ella Barkan*, Greg Shakhnarovich** and Asaf Tzadok*

*IBM Research - Haifa, Israel, **TTI-Chicago

Abstract. This paper presents a novel method for automatic lesion detection in breast ultrasound images; the method performs multi-stage learning of lesion-specific boundaries represented by a bag of robust features. The proposed method can be seen as an edge pruning procedure that leaves only object-specific edges and filters out the rest. It can be combined with segmentation algorithms that rely on edge information. We show an example of such combination with one of the state-of-art segmentation algorithms; our method yields improved segmentation results. The proposed method is tested on a set of 400 breast ultrasound images, with the goal to automatically detect lesion boundaries. However, we believe that our method can be used by radiologists as an assistance tool during examination routine, in which case it may help to better localize lesions and document the findings. The performance of our method is compared to a state-of-art object boundary classification algorithm; we show that our method outperforms it in different tests.

Keywords: Automatic lesion detection, breast ultrasound image, boundary probabilities, bag of features, cascade classification

1 Introduction

Object detection and segmentation is an active area of research in computer vision, and has various applications in the medical imaging domain. Many object segmentation methods rely on object boundary detection as the first step of an algorithm. One of the most common methods for this first step of edge detection is the Canny edge detector [1], which usually provides consistent results. On the other hand, one of the the current state-of-art in the edge detection is '*gPb*-owt-ucm', the probability of boundaries contour detection and segmentation method of [2] which is trained to detect edges in natural images, and then to combine the edges into segments.

Object boundaries in natural images are usually well defined, and, therefore, objects can be clearly distinguished from their background. In contrast, in medical images, objects such as lesions or even organs do not always have fully closed contours with clear boundaries. The reasons for that include limitations of acquisition methods and devices (for example, having low resolution and/or low dynamic range), occlusions of other tissues and organs, and others. In these cases, standard segmentation methods may produce segments that

ISBN: 978-87-996443-0-8

have no physical meaning. In particular, such methods merge neighboring areas if they are 'similar' enough, or if there are no sufficiently strong boundaries detected between the areas. In these cases a better strategy would be to detect most probable object boundaries, by learning their typical properties from available examples, and then merge them into closed segments by some ad-hoc approach. We therefore present a new such approach to lesion-specific boundary detection; it is based on multi-stage cascade learning. We apply our approach to the problem of automatic detection of lesions in breast ultrasound images.

The paper is organized as follows. In the next section we give a detailed description of the proposed method. Then, we present various experiments where we compare the performance of our method with state-of-art methods, and conclude with the summary.

2 The proposed method

In the next subsection we give a high-level overview of the proposed approach, and then provide more details in the following subsections.

2.1 Overview of the method

The proposed method of learning-based localization of objects can be summarized as follows. First, an edge detection procedure is applied to an image. Edge detection can be performed by any known approach, for example, Canny edge detector, or more sophisticated methods that consider not only intensity changes but also other image characteristic changes, such as color and texture as in [2]. The edges obtained in the first step are used as an initial boundary guess. Next, boundary descriptors are calculated for individual pixels on the edges (an extension to the case of group of pixels is described below). Given a set of *training* images with object boundary ground truth available (e.g., manually created in advance), we collect positive and negative examples of edge pixel descriptors. Based on these examples, we train a cascade classifier and use the learned models to classify boundaries with a score that reflects the likelihood of a particular boundary or its section to belong to an object in question. An example of the output of our method in the case of lesion detection from breast ultrasound (US) is shown in Figure 1, right. Finally, we may want to aggregate the resulting positively classified boundaries' sections into closed segments or areas. As in the case of initial edge detection, the last step can utilize one of the known contour aggregation methods such as ObjCut [3], or Ultrametric contour map of [4], which we use in our experiments.

The contribution of our method is twofold: 1) we propose a new description of lesion boundaries represented by a feature set which is robust to various changes in object appearance as a result of variability in characteristics of image acquisition devices, and 2) we present a new cascade-type classifier which is trained to detect lesion-specific boundary sections and to merge these boundary sections at each consecutive classification stage.

ISBN: 978-87-996443-0-8

2.2 Feature channels and initial edge detection

Our goal is to train the system to detect object specific boundaries. We start with description of the boundary feature channels. We generate additional channels from a given image as follows. Suppose the image contains only intensity channel, such as in ultrasound or X-ray. If an image has three color channels, such as in colonoscopy, we can use them in a similar way. Using the intensity channel, the 1^{st} feature channel, we calculate the following additional channels that describe various textural characteristics and local directional coherence features. The first additional, that is the 2^{nd} feature channel, is the texton-based texture map as used in [2]. The next three channels are based on the local entropy of the pixel neighborhood [5], calculated at three scales (window sizes of 7x7, 15x15, and 35x35). These channels characterize well textures at various resolutions and thus encode multiscale texture properties. The 6^{th} feature channel is the phase congruency from [6]. This channel provides fine edge information that, in our case, represent a complementary description of textures in an image, and encode local directional characteristics.

The first step in our method is an edge detection; it needs to find most of the object boundaries, including the weak ones, even in expense of many noisy detections (which will be filtered out by our learning-based method). We experimented with two different approaches to perform the initial edge detection. In the first approach, we used the state-of-art probability of boundaries algorithm from [2]. The second approach is quite simple: we applied Canny edge detector to each one of the six channels described above, and took a pixel-wise maximum of the six resulting edge maps from those six channels. Notice that the first method is trained to detect edges in natural images. We found empirically that the second approach was better suited to medical images, as it found consistently more weak edges than the first approach[1].

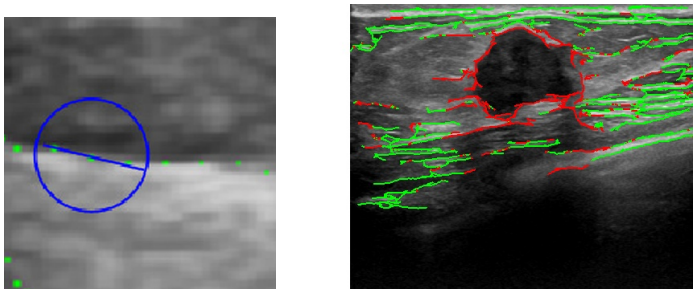


Fig. 1. Left: an example of sampled points on boundaries from the 1^{st} stage of classifier, and of a disk ROI around one such point. Right: final edge classification for the whole image obtained at the 8^{th} cascade classification stage. Red edges classified as object-related, green edges classified as non-object related.

[1] In general, it is possible to train the method in [2] on medical images; we leave it for the future research

ISBN: 978-87-996443-0-8

2.3 Boundary descriptors

We now turn to the definition of boundary descriptors. The boundary map from the edge detection step contains separate *edge sections* (or, edge segments). For each such section, we take every m^{th} pixel with coordinates (x,y) on the corresponding boundary section. We then place a circular disc of radius R at (x,y), and split it into two halves according to the major boundary direction (see illustration in Figure 1, left). We found that sampling each 20^{th} pixel on boundaries and a disk of $R=20$ yields good results. Given the two halves of the disk for a given pixel (x,y), for each one of the six information channels we calculate two histograms of channel values obtained from both halves of the disk. We write it compactly as a list of histogram bins with non-zero probability, i.e.

$$\{(z_i, p_i^{k,ch})\}_{i=1}^{n}$$

where z_i is a bin-center, $p_i^{k,ch}$ is the corresponding probability mass for that bin for the k^{th} probability distribution ($k = 1, 2$), corresponding to the k^{th} half, $ch = 1...6$ is the channel number, and n is the number of such bins. Note that we have fixed both the number of bins and the bin-centers across the two distributions. This standard technique is quite prevalent in object recognition, for example in the Bag of Visual Words technique (see e.g. [7]).

Further, for each one of the six channels, we calculate the D-Field pairwise descriptor which was recently proposed in [8]. In particular, for bin i of the first distribution D-Field δ_i is defined by

$$\delta_i = \sum_j f_{ij}(z_j - z_i)$$

where f_{ij} is the *flow* matrix between two distributions obtained as a by-product of computation of the Earth Mover's Distance (EMD) [9]. Intuitively, the D-Field descriptor captures – for each bin – where its probability mass moves.

In general terms, the D-Field descriptor captures the relationship or transformation between two probability distributions of two halves of the disk. One of its advantages argued in [8], is its robustness to various changes in image appearance that may occur as a result of different acquisition devices having very different characteristics such as dynamic range, resolution, and others.

Finally, the bag of boundary features for a pixel on the boundary with coordinates (x,y) and a corresponding disk around it comprises of: 1) two histograms (for each half of the disk) per each of the six channels, 2) D-Field descriptors per each channel, 3) the EMD distance between the two distributions (for each half of the disk) per each channel, and 4) the coordinates (x,y). Overall, it yields $(6 \times 3 \times n) + 8$ where n is the number of bins in the histograms.

2.4 Cascade classifier of boundary features

In this section, we describe the structure of the proposed cascade classifier[2] As we mentioned above, we first identify separate edge sections where each pixel

[2] We describe main ideas, omitting some technical details because of the lack of space.

ISBN: 978-87-996443-0-8

(or, alternatively, each m-th pixel) is a center of the local neighborhood that has a disk shape. After we calculated a descriptor for each one of the pixel areas, we train the 1^{st} classification stage.

Training. Given a set of training images with marked inner and outer object boundaries (the ground truth), we label pixels on the boundary as positive examples if they are situated in between those two contours, and as negative ones otherwise. A training set of 300 images yields quite a large amount of training samples - around 1.5 million - with the number of positives about 20 times smaller than negatives. Since the number of features is much smaller than the number of samples, it is more convenient and efficient to use a linear SVM classifier. We divide the full set of examples into two subsets; one is used in the odd classification stages, and the other one is used in the even stages. We start with training the 1^{st} classification stage on the 'odd' subset as follows: we use weighted cost that penalizes the errors in positive examples about 10 times higher than those in negative ones. This asymmetric cost assures that nearly all positive samples are correctly identified, in expense of large number of errors in the negative class. This is, however, exactly what we are up for, since we propagate the results from this 1^{st} stage to the next, 2^{nd} classification stage. In particular, we now use the model obtained from the 1^{st} classification stage to classify the examples in the 'even' subset. We then use all the positively classified samples to train the 2^{nd} stage model. This time, however, instead of taking separate edge pixel areas, we merge neighboring pairs of pixel areas whose centers are 20 pixels apart on the same edge section. In fact, we calculate the same descriptors described above, but using a neighborhood combined of two disks around the two closed-by pixels (centers of disks), and a 'corridor' between them. We train the 2^{nd} stage model on the 'even' subset, and use the obtained model to classify the examples from the 'odd' subset. Positively classified examples from the 2^{nd} stage are used then to train the 3^{rd} classification stage; at this stage, we use pairs of pairs of pixels. This hierarchical process continues till there are no boundary sections remaining that can be merged and classified. At each training stage we optimize the SVM parameters using 10-fold cross validation to prevent possible overfitting.

Testing. Having trained the hierarchy of models, the testing (that is, object boundary detection) is straight forward. As before, the 1^{st} step is the edge detection applied to an inquiry image. We then start from classifying individual edge pixels using the 1^{st} stage SVM model. Next, positively classified neighboring pixel areas are merged, and the pairs are classified using the 2^{nd} stage model, and so force.

3 Experiments

The system performance was tested on 400 ultrasound images collected from 4 different acquisition devices, with about 100 images from each device. Ground truth was created manually by a radiologist who marked inner and outer lesion contours. The two contours required to count for the uncertainty in cases of lesions without clear boundaries, and for the variability in radiologists' markings.

ISBN: 978-87-996443-0-8

During the training stage we label edge sections found between inner and outer contours as positive examples, and all others as negatives.

We performed two types of tests evaluating the accuracy of the boundary detection. In the first, 'random subsampling' test, we divided the whole set randomly into training and testing subsets with roughly 300 and 100 images respectively. We then performed 50 such complete training and testing cycles for each such set, and averaged the performance of our system over 50 independent runs. This experiment provides figures of performance of the method when we train and test the system with images from the same device or a set of devices. This test verifies the robustness and adaptivity of the method. In the second, 'blind' test, we separated training and testing sets of images according to the acquisition devices. We then used images from three different devices for training and images from the other one for testing in a leave-one-out manner, and averaged the results over the 4 runs. It simulates a situation wherein the algorithm needs to perform in an environment with various devices whose images are not available during the training stage. This test verifies the ability of our method to generalize well to previously unseen images.

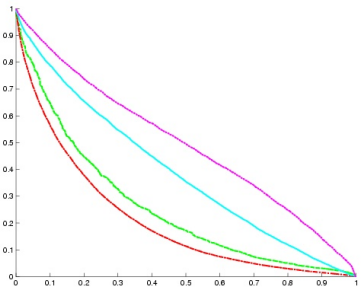


Fig. 2. False negative versus false positive errors: averaged ROC curves of lesion/non-lesion edge classification. Red and Green: our method in 'random subsampling' and in 'blind' tests respectively; Blue and Magenta: method of [10] in 'random subsampling' and 'blind' tests respectively.

Although there is extensive literature on the breast ultrasound (e.g., a survey in [11]), it is difficult to compare different methods, since algorithms and datasets are usually proprietary, and not available for use. We therefore compare the performance of our method with general segmentation and boundary detection methods. We test our method against object boundary classification method from [10]. This method was designed for natural images, but this is the only other approach that we are aware of, that learns object-specific *boundaries* of general objects (versus algorithms trained to detect specific objects like cars, faces etc., and use object specific features). In [10], the authors use a patch centered on the edge, rotated so that its x-axis is aligned with the edge tangent; color channel values from the patch are used as the local edge descriptor. Since our images have one intensity channel, we added additional channel of texture, to have a more 'fair' comparison. We trained a linear SVM classifier as done in [10], for each set of training images (described above), and applied the trained model for detection of

ISBN: 978-87-996443-0-8

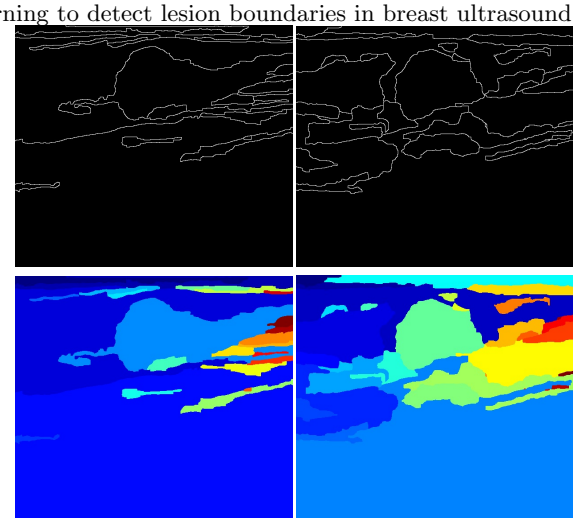


Fig. 3. Contours (upper) and segments (bottom): based on boundaries from [2] (left), and on the proposed method (right).

lesion boundaries in the testing set. We judge the performance of the competing methods by comparing their ROC curves (false negative versus false positive errors) obtained from classification of edges into lesion/non-lesion classes. The summary of these experiments is given in Figure 2; the ROC curves are averaged over either 50 or 4 runs, according to the test performed (as explained above). Obviously, simple descriptors from [10] cannot provide good detection of lesion boundaries since similar patches may appear anywhere in an image. In the case of random subsampling experiment, the classifier based on [10] succeeded to learn and separate 2 classes, though with high error; in the case of 'blind' test, it fails completely. In contrast, our method learns well from examples: in the random subsampling test, it yields 27% mean equal error rate; in the 'blind' test, our method shows its robustness and yield 33% mean equal error rate.

In another experiment, we compared our method to the state-of-art '*gPb*-owt-ucm' edge detection and segmentation method of [2]. We replaced the boundary detection method, known as the probability of boundaries 'gPb' algorithm, in '*gPb*-owt-ucm' method, with our boundary detection method, and used the contour based segmentation of 'owt-ucm' algorithm to combine detected boundaries into closed contours forming segments. Figure 3 shows an example of such comparison; the *gPb*-owt-ucm fails to localize correctly the lesion, while our method outlines nicely the important region, using the estimated probabilities of lesion boundaries.

4 Conclusions

In this paper we presented a novel method for automatic lesion detection in breast ultrasound images. The method is based on multi-stage learning of object-specific boundaries represented by a bag of robust features. The proposed method

ISBN: 978-87-996443-0-8

can be combined with segmentation algorithms that rely on object edge information. We show an example of such combination of our method with one of the state-of-art contour based segmentation algorithms. Our method improves upon the results of the original algorithm.

The proposed method was tested on breast ultrasound images, with the goal to automatically detect suspicious lesion boundaries. However, we believe that the proposed method can be used by radiologists as an assistance tool during the examination routine. In this case, our method may help to better localize lesions and document the findings. Furthermore, since the method is general, we are planning to apply it to the problem of automatic lesion detection in mammogram images, and, possibly, in other modalities.

References

1. Canny, J.: A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on (6) (1986) 679–698
2. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on **33**(5) (2011) 898–916
3. Kumar, M.P., Ton, P., Zisserman, A.: Obj cut. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 1., IEEE (2005) 18–25
4. Arbeláez, P., Maire, M., Fowlkes, C., Malik, J.: From contours to regions: An empirical evaluation. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 2294–2301
5. Awate, S., Tasdizen, T., Whitaker, R.: Unsupervised texture segmentation with nonparametric neighborhood statistics. Computer Vision–ECCV 2006 (2006) 494–507
6. Kovesi, P.: Phase congruency: A low-level image invariant. Psychological research **64**(2) (2000) 136–148
7. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, Ieee (2007) 1–8
8. Kisilev, P., Freedman, D., Wallach, E., Tzadok, A., Naveh, Y.: Dflow and dfield: New features for capturing object and image relationships. In: Pattern Recognition (ICPR), 2012 21st International Conference on, IEEE (2012) 3590–3593
9. Rubner, Y., Tomasi, C., Guibas, L.: The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision **40**(2) (2000) 99–121
10. Prasad, M., Zisserman, A., Fitzgibbon, A., Kumar, M., Torr, P.: Learning class-specific edges for object detection and segmentation. Computer Vision, Graphics and Image Processing (2006) 94–105
11. Nover, A.B., Jagtap, S., Anjum, W., Yegingil, H., Shih, W., Shih, W., Brooks, A.D.: Modern breast cancer detection: A technological review. International Journal of Biomedical Imaging (2009)

ISBN: 978-87-996443-0-8

Image Quality in automated breast ultrasound images: a preliminary study for the development of automated image quality assessment

Julia Schwaab[1], Yago Diez[2], Joan Martí[2], Robert Martí[2], Jan van Zelst[3], Bram Platel[4], Tao Tan[3], Johannes Gregori[1], Stefan Wirtz[4], Johanna Kramme[4], and Matthias Günther[1,4]

[1] mediri GmbH, Heidelberg, Germany
[2] University of Girona, Spain
[3] Radboud University Nijmegen Medical Centre, the Netherlands
[4] Fraunhofer MEVIS, Bremen, Germany

Abstract. Effectiveness of breast cancer screening programs is enhanced by adding 3D automated breast ultrasound (ABUS) or breast MRI to standard X-ray mammography for women at high risk or with dense breasts. However, these modalities are relatively sensitive to image artifacts and evaluation of these supplemental images requires additional manpower. Online automatic quality assessment (AQUA) in these images could be useful. In this work, the quality criteria that could be considered by an AQUA system for ABUS images were defined. First algorithms that are able to detect the described artifacts and quality aspects are presented, i.e. methods for nipple and rib detection are proposed. Furthermore, a data base of 402 3D ABUS images of 42 women was collected and classified according to the above mentioned quality aspects. In this way, a data base for development of software tools could be established.

Keywords: Automated Breast Ultrasound, Automated Image Quality Assessment, Image Artifacts.

1 Introduction

Breast cancer is the most common cancer that affects women [1]: currently, approximately one in eight women develops breast cancer during her lifetime. Experiences of the last years have shown that screening programs reduce mortality significantly [2, 3]. However, still about one third of breast cancers are detected in between screenings [4] and approximately 25-30% of screen-detected cancers can be detected retrospectively on earlier mammograms [5]. This motivates clearly that the standard X-ray mammography breast cancer screening should be further improved, e.g. by adding ABUS (automated breast ultrasound) or breast MRI examinations for women at high risk or with dense breasts [6, 7, 8].

However, ABUS and breast MRI are susceptible to image artifacts and thus, are in demand for proper quality assurance. Visual inspection of the image data by trained

adfa, p. 1, 2011.

ISBN: 978-87-996443-0-8

personnel is still the predominant method to ensure high quality images. But this common practice binds manpower of highly qualified personnel and hardly yields quantitative or objective judgments, as it strongly depends on the individual human visual system.

When ABUS images are acquired, the technician controls by hand the typical ultrasound parameters (scanning depth, contrast, time gain). Furthermore, the technician is responsible for the positioning of the ABUS transducer on the woman's breast and for the number of volumetric data sets that are acquired per breast and even though the personnel is well trained, there may occur image artifacts due to suboptimal ultrasound settings. Furthermore, wrong placement of the transducer or the patient as well as patient movement can cause artifacts.

In the last years, several promising approaches towards an automated and objective image quality assessment for MRI images have been made, e.g. image noise extraction, patient motion and ghost artifact detection [9, 10]. However, such efforts are mainly restricted to study specific brain MRI applications. To our knowledge, there is no flexible and general applicable automated quality assessment tool on the market. Especially in breast MR and US, no software application can be found that assesses image quality of screening data at the time of acquisition.

The aim of this work is to prepare the development of an automated quality assessment (AQUA) system that inspects ABUS (and breast MR) images directly after or even during acquisition such that potential artifacts can be detected immediately and, when indicated, the examination can be repeated immediately without having to recall the patient.

2 Material and Methods

As a start, the focus was put on ABUS image quality assessment. All 402 volumetric ABUS images of 42 women included in this study were acquired at Radboud University Nijmegen Medical Center, the Netherlands, using a Siemens ACUSON S2000 or U-Systems SomoVu ABUS device. Based on this dataset, specific quality aspects were defined in close cooperation between a physician, a radiologist, three physicists, three biomedical engineers and three experts in image analysis. These criteria were not only induced by the clinical needs of professional radiologists who are supposed to interpret the images, but also by the requirements of several CAD software tools as chest wall detection [11] or lesion classification algorithms. All images were classified by a budding radiologist according to the previously defined criteria. Furthermore, a dedicated test ABUS examination was performed and the imaging procedure was analyzed for potential causes of image artifacts.

So far, two algorithms for rib detection and for nipple detection were set up in a first version. In order to detect Ribs, a slightly modified version of the algorithm presented in [11, 13, 14] is used. In a first step, computations based on the eigenvalues of the smoothed Hessian (second order derivative matrix) are used to detect sheet-like structures. Both slice-wise 2D and full 3D eigenvalue computation was considered and the later was finally used. This type of procedure is known to produce a large

ISBN: 978-87-996443-0-8

number of false positives, so a false positive reduction step is mandatory. In this case, in addition to the steps already used in the aforementioned reference (breast mask and distance to the nipple) a "rib probability map" was considered. To be precise, this probability map was constructed by considering the points detected as ribs in each slice. Projecting all such slices along the coronal axis and adding them up in order to get a single image that sums up the number of times each pixel had been found to be a rib candidate and dividing for the total number of slices in order to get the probability map. This is based on the fact that, once a rib is detected, its shadow is also detected in all subsequent slices along the mentioned axis. Consequently, candidate pixels in low probability areas where considered to belong to isolated regions mistakenly considered as rib candidates in few slices.

In order to automatically determine the position of the nipple, a classical 2D nipple segmentation algorithm was used. First of all, a slice-wise nipple detection algorithm was run: 1) Lower intensity pixels were detected as candidates. 2) A canny edge detector was run in the whole slice to try to detect nipple boundaries. 3) Each candidate point was paired with its corresponding boundary and an active shape algorithm was run to try to fit each boundary to a circle. Those points that succeeded were labeled as possible nipple positions. False positives were reduced by relating the 2D information obtained at each slice: the position of the nipple may not differ more than a prearranged threshold.

Quality aspect	Rating options (and results in %)		
Number of acquired views	too few	enough	
Scanning Depth	too low (36)	good (53)	too high (11)
Resolution	too low (3)	acceptable (42)	good (55)
Noise	too prominent (14)	acceptable (75)	good (11)
Contrast	too low (1)	acceptable (41)	good (58)
Brightness	too low (5)	good (93)	to high (2)
Edge shadows	too prominent (20)	acceptable (64)	good (16)
Number of air artifacts	>1 (7)	1 (9)	0 (84)
Discontinuities	too prominent (1)	acceptable (2)	good (97)
Sinusoidal pattern	too prominent (10)	acceptable (34)	good (56)
Visibility of ribs	not visible (8)	acceptable (17)	good (75)
Nipple position	too near to edge (36)	acceptable (16)	good (48)
Nipple / duct shadow	too prominent (64)	acceptable (21)	good (15)

Table 1. Quality aspects of ABUS images and results of the manual classification process.

3 Results and Discussion

The image quality aspects that were found to be relevant in ABUS imaging and their relative frequency in the considered data set are listed in table 1 and described in detail in the following section. As the AQUA software is intended not only to be an image analysis tool in the sense of image processing but also to test the validity of the DICOM tags, the following quality criteria cover both these aspects.

ISBN: 978-87-996443-0-8

An AQUA tool should check whether the **number of acquired views** per breast corresponds to the breast (cup) size which is entered during the examination. The bigger the breast, the more different views have to be collected in order to represent the whole breast.

Additionally, the **scanning depth** has to be surveyed. As the longitudinal resolution of the ultrasound image is inversely proportional to the scanning depth, there is always a trade-off between these quantities. The radiologist can only be sure to see the whole breast (in anterior-posterior direction) and not to miss some centimeters of tissue when the chest-wall, i.e. the pectoral muscle, is clearly visible on the image. This could be tested by an automatic chest wall detection or pectoral muscle segmentation on the volumetric image data sets. The longitudinal **resolution** which is directly coupled to the scanning depth should not undergo a certain limit. The classification result in table 1 shows that the scanning depth was too low in 36% of the images.

It is generally accepted that **noise** on medical images resulting in low image quality can largely limit the diagnostic usefulness. One idea to determine the noise level is to measure contrast (defined, for example, as the "standard deviation of the pixel intensities"), run a noise filter and then compute contrast again. This way, a measure of noise based on the difference in contrast could be established. 14% of the considered images were classified as noisy.

The image **contrast** and the overall **brightness** have to lie within a certain range in order to allow best possible image analysis, both for clinical diagnosis, as well as for automated analysis lesion detection and classification. It might be useful to calculate a contrast value only in the foreground region of an image after a breast segmentation has discarded the background. As for the brightness, an AQUA tool could focus on the absolute intensities of the image (maximum intensities, intensity range) or try to analyze the distribution of intensities. In the latter case, one possibility would be to use shape measures such as skewness or kurtosis. According to the classification result (table 1), contrast and brightness are acceptable or good in most of the cases.

Depending on the position and angle of the transducer frame, there might be substantial black areas at the edges of the images because the breast was compressed in such a way that there was no more tissue to be imaged. These shadow regions in the borders of the image do not convey useful information and might indicate that not the whole of the breast is visible (see figure 3). These black areas will be referred as "**edge shadows**". In order to detect them, an edge shadow will be characterized as a set of rows (columns) adjacent to an image extreme holding that all their pixels are black (except for a small threshold to account for noise). The algorithm to measure these shadows, consequently, begins scanning at an image edge and proceeds row by row (column by column) until a non-black pixel is detected. The width (height) of the edge shadow is then determined as the number of rows (columns) times the dimensions of the pixel. The edge shadows were found to be too prominent in 20 % of the images.

ISBN: 978-87-996443-0-8

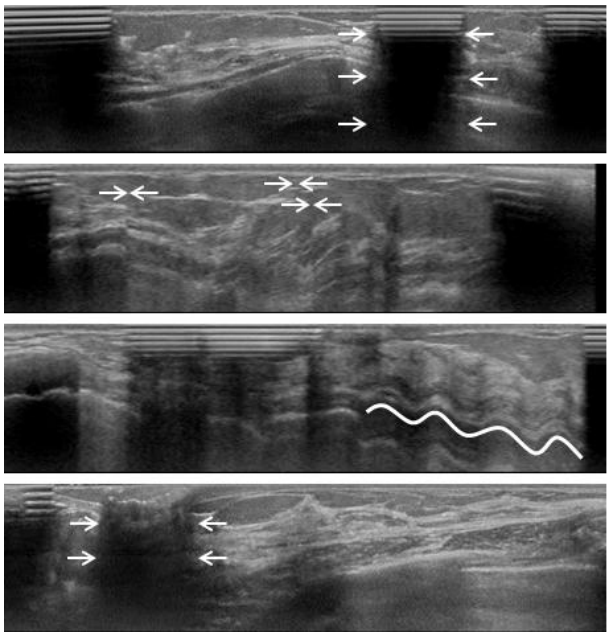


Fig. 1. Typical ABUS image artifacts. *1st line*: transversal view with air artifact due to lack of contact fluid. *2nd line*: sagittal view with discontinuities due to transducer bumping. *3rd line:* sagittal view with sinusoidal pattern due to respiration. *4th line*: transversal view with shadow due to nipple and ducts.

Another common artifact in ultrasound imaging are shadows (**air artifacts**) caused by a lack of contact gel. If there is no good contact between the transducer and the skin via a contact fluid, the sound waves are reflected by the air between transducer membrane and skin. In this way, a characteristic pattern of some bright and dark stripes on the image near the surface is produced which quickly ends up in a shadow column in longitudinal direction (see figure 1). These air artifacts can be detected by blob detection, i.e. region labeling, algorithms that consider the shape and size of the shadow in all three dimensions as well as the characteristic resonance artifact near the surface. 16 % of the classified images contain at least one air artifact.

As the volumetric image data set is reconstructed from several 2D ultrasound image slices which are collected one after another by the transducer scanning over the breast, it is crucial that the transducer moves smoothly at constant velocity through its frame. If, however, the pressure of the transducer on the skin is too high or if there are hard tissue structures in the breast, the transducer motion may be hampered. In consequence there will be **discontinuities** between the lines in the reconstructed images (see figure 1). Some kind of gradient image filter could detect the transducer bumping. However, the image rating showed that this effect does not appear very frequently (1 %).

Eventually, the success of the imaging process also depends on the cooperation of the patient. If the woman is breathing strongly, talking or coughing during the exami-

ISBN: 978-87-996443-0-8

nation, the volumetric image will show some kind of **sinusoidal pattern** in the reconstructed image plane (see figure 1). This may also be detected by customized gradient image filters. Groundwork in this area has already been performed by Boehler and Peitgen [12]. This pattern was detected in 10 % of the considered images.

In order to allow further automatic image analysis, i.e. chest wall detection or lesion segmentation, the images should fulfill some additional quality criteria even though these might be not this important for the radiologist, as for example the **visibility of ribs**. A chest wall detection algorithm developed by Tan et al. [11] needs at least three ribs to be imaged with still enough shadow behind them so that they can be segmented reliably. In order to decide whether or not a sufficient number of ribs is clearly visible, a first version of a rib detection algorithm was implemented (see section 2). In figure 2, some single steps and the result of the proposed algorithm are presented. In this case, the upper three ribs are segmented correctly while the bottom rib is detected in two separate parts. The correct counting in this case is still a pending challenge. According to the rating result, the ribs were not visible in 8 % of the images.

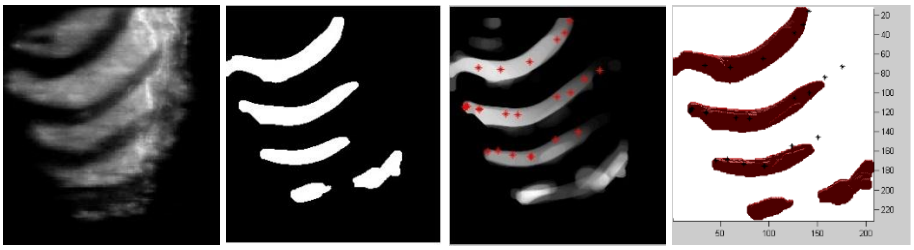


Fig. 2. The different stages in automated rib detection from left to right: 1) typical ultrasound slice in coronal view where ribs are visible, 2) binary map of this slice, 3) sum projection of all slice-wise binary maps result in the probability map, and 4) the segmented ribs.

Many times, a major concern in ultrasound breast images are the **nipple** and the **ducts**. On the one hand, the nipple is an important landmark which helps, e.g., combining different views. On the other hand, the nipple can cause severe shadows in the ultrasound images due to entrapped air and lack of contact between transducer and skin (see figure 1). The ducts are also filled with air and, thus, can produce elongated shadow regions along their axes. These shadows may cover important structures in the breast image and hinder a solid diagnosis. This was the case in 64 % of the classified images.

Furthermore, the position of the nipple relative to the rest of the breast in the image is a quality aspect that should be considered. The nipple being too close to the edge of the (laterally compressed) breast might constrain the view of the radiologist on important areas and induce uncertainty about the true edge of the breast. This effect was observed in 36 % of the images. Figure 3 presents an exemplary output of the proposed nipple detection algorithm (see section 2), which allows computing the distance to the edge.

ISBN: 978-87-996443-0-8

In order to translate the above described subjective human image quality criteria into an objective criterion, the single software tools will have to be trained by a machine learning algorithm on the classified image data set. As some artifacts only appear very rarely, it might be necessary to expand the image data base and to get the images classified by more than one radiologist.

As soon as the single AQUA tools are in a stable version, they might be integrated to clinical practice as add-on. Depending on the false positive rate, the examination time might be increased needlessly for some woman. However, saving the effort of recalling women some days later for a repeated examination should outweigh this effect. Of course, the possible increase in workload and costs will have to be investigated thoroughly.

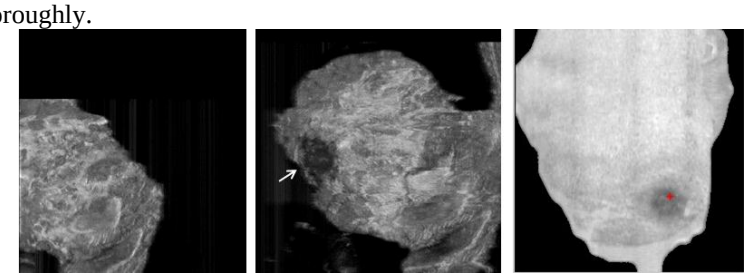


Fig. 3. Coronal views of the breast. *Left*: example for dominant "edge shadows" (the scan was aborted in this case). *Middle*: position of the nipple is very close to the edge of the breast. *Right*: example output image of the automatic nipple detection algorithm. The proposed nipple position is marked by a red star and coincides well with the true position.

In the end, it should be practical to rely on AQUA findings alone for repeating a scan. However, it remains to be seen during the course of this study whether the compromise of additional visual conformation might be necessary.

4 Conclusion

Several ABUS image artifacts have been defined in an iterative process and found to be relevant at least for either the interpretation of the radiologist or an automatic post-processing of the images. For many of the described artifacts, there exist already conceptual designs or software prototypes that will be able to detect them. The total of 402 test images was manually classified by a radiologist according to the previously defined quality criteria. In the future, the same definition process of quality criteria and manual image data rating will be performed for breast MR images. In this way, two classified image data bases will be established supporting the development of AQUA software tools. In the end, the different AQUA software tools will be combined and managed by a single pipeline that can process the image data directly after acquisition and produce a quality report or pronounce warnings, when indicated. This quality assessment software could enhance the overall image quality in breast imaging, hence simplify the interpretation by the radiologist and, finally, safe time and money.

ISBN: 978-87-996443-0-8

Acknowledgements

Parts of this work have been funded by the European Commission's FP7 Cooperation program project ASSURE (project number 306088).

References

1. J. Ferlay, E. Steliarova-Foucher, J. Lortet-Tieulent, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012. *Eur J Cancer*, 49(6):1374-1403, Apr 2013.
2. S. J. Otto, J. Fracheboud, A. L. M. Verbeek, et al. Mammography Screening and Risk of Breast Cancer Death: A Population-based Case-Control Study. *Cancer Epidemiol Biomarkers Prev*, 21(1):66-73, Jan 2012.
3. C. Nickson, K. E. Mason, D. R. English, et al. Mammographic Screening and Bresat Cancer Mortality: A Case-Control Study and Meta-analysis. *Cancer Epidemiol Biomarkers Prev*, 21(9):1479-1488, Sep 2012.
4. R. L. Bennett, S. J. Sellars, and S. M. Moss. Interval cancers in the NHS breast cancer screening programme in England, Wales and Northern Ireland. *Br J Cancer*, 104:571-577, Feb 2011.
5. J. D. M. Otten, N. Karssemeijer, J. H. C. L. Hendriks, et al. Effect of Recall Rate on Earlier Screen Detection of Breast Cancers based on the Dutch Performance Indicators. *J Natl Cancer Inst*, 97(10):784-754, May 2005.
6. M. T. Mandelson, N. Oestreicher, P. L. Porter, et al. Breast Density as a Predictor of Mammographic Detection: Comparison of Interval- and Screen-Detected Cancers. J Natl Cancer Inst, 92(13):1081-1087, May 2000.
7. L. Yaghjyan, G. A. Colditz, L. C. Collins, et al. Mammographic Breast Density and subsequent Risk of Breast Cancer in postmenopausal Women according to Tumor Characteristics. *J Natl Cancer Inst*, 103(15):1179-1189, Jul 2011.
8. K. M. Kelly, J. Dean, S.-J. Lee, et al. Breast cancer detection: radiologist's performance using mammography with and without automated whole-breast ultrasound. *Eur Radiol*, 20(11):2557-2564, Nov 2010.
9. W. S. Kerwin, J. Cai, and C. Yuan. Noise and motion correction in dynamic contrast-enhanced MRI for analysis of atherosclerotic lesions. *Magn Res*, 47(6):1211-1217, Jun 2002.
10. B. Mortamet, M. A. Bernstein, C. R. Jack Jr., et al. Automatic quality assessment in structural brain magnetic resonance imaging. *Magn Res*, 62(2):365-372, Aug 2009.
11. T. Tan, B. Platel, R. Mann., et al. Chest Wall Segmentation in Automated 3D Breast Ultrasound Scans, *Medical Image Analysis* (2012), doi: http://dx.doi.org/10.1016/j.media.2012.11.005
12. Boehler, T. and Peitgen, H.-O. Reducing Motion Artifacts in 3-D Breast Ultrasound Using Non-linear Registration. *MICCAI 2008, LNCS*, 5242:998-1005, 2008.
13. H. Huisman and N. Karssemeijer, Chestwall segmentation in 3D breast ultrasound using a deformable volume model. *IPMI 2007, LNCS*, 4584:245-256, 2007.
14. R.-F. Chang, Y.-W. Shen, J. Chen, et al. Rib detection for whole breast ultrasound image. *Proc. SPIE 6915, Medical Imaging 2008: Computer-Aided Diagnosis*, 691525, Mar 2008.

ISBN: 978-87-996443-0-8