

Improving a CAD System Using Bilateral Information

Meritxell Tortajada, Arnau Oliver, Yago Díez, Robert Martí, Joan C. Vilanova, and Jordi Freixenet

Abstract— Computer Aided Detection (CAD) mammographic systems are used in medicine to assist radiologists in the evaluation of mammographic images. The aim of this work is to compare the results of a developed single-image CAD system with a new one, dual-image CAD, that adds registration information of bilateral mammographic images in the training step of the former system. The evaluation of the different registration methods is performed using similarity measures. Receiver Operating Characteristic (ROC) analysis and Free Receiver Operating Characteristics (FROC) analysis are used to compare the results of both CAD systems. At a sensitivity of 80%, the false positives per image was 1.68 for the single-image CAD system and 0.90 for the dual-image CAD system. The results shows the benefits of integrating bilateral information into the CAD system.

I. INTRODUCTION

A CAD system is a set of automatic or semi-automatic tools developed to assist radiologists in the detection and/or classification of mammographic abnormalities [1]. Although the idea of developing computer systems to assist radiologists in the detection and classification of breast cancer is not recent, the actual development of full field digital mammographic imaging systems has been a catalyst in the increase of such computer systems [2].

Detection of abnormal structures or architectural distortions in mammograms can be performed by comparing images of the same patient using an image registration methodology. In this work we focus on the comparison of left and right mammograms of the same view and patient taken at the same time, also known as bilateral comparison. This practice is justified from a clinical point of view by Kopans [3], who makes two important observations when comparing different mammograms of the same woman: though one breast may be larger than the other, internal structures are quite symmetric over broad areas, and overlapping tissue structures that form summation shadows and normal tissue variations on the mammogram highlight asymmetries. Also, the radiologists Tabár and Dean [4] developed a method for looking for lesions based on comparing corresponding regions of the left and right mammographic images to detect differences that could be lesions. Fig. 1 shows an example of this registration. Fig. 1(b) is the mammogram which is registered with Fig. 1(a). The resulting mammogram is shown in Fig. 1(c), while Fig. 1(d) and Fig. 1(e) shows the difference of the images after and before of the registration.

This work was supported in part by Ministerio de Educación y Ciencia of Spain under Grant TIN2007-60553 and by the CIRIT and CUR of DIUIE of Generalitat de Catalunya under Grant 2008SALUT00029.

M. Tortajada, A. Oliver, Y. Díez, R. Martí, J.C. Vilanova, and J. Freixenet are with Institute of Informatics and Applications, University of Girona, 17071 Girona, Spain txell@atc.udg.edu

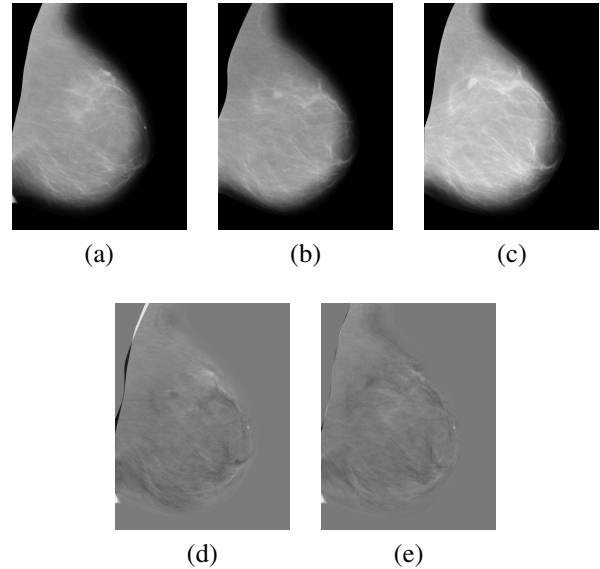


Fig. 1. Affine MR & BSP MR registration example of mammographic images: (a) fixed image, (b) moving image, (c) registered image, (d) difference image before registration "(a)-(b)" and (e) difference image after registration "(a)-(c)"

Note that Fig. 1(e) is more homogeneous than Fig. 1(d), indicating that the mammogram after registration is more similar to the target mammogram without registration.

The aim of this work is twofold. Firstly, we want to test different registration techniques which have not been usually applied to mammography. Secondly, with the best registration technique, we will convert a previously developed single-image CAD into a new dual-image CAD that takes the bilateral information into account. Further, we will compare the results of both CADs in order to show the benefits and drawbacks of taking this information into account. To date, no commercially available CAD systems incorporate information from different images of the same patient using image registration.

This paper is structured as follows: In section II, we describe the used methodology: the single-image CAD, the image registration algorithms, and the dual-image CAD resulting from the integration of registration information into the single-image CAD. Subsequently, results are presented, providing details on the data and experiments (Sec. III). Finally, discussions and conclusions are provided (Sec. IV).

II. METHODOLOGY

A. Single-image CAD system

Fig. 2 depicts our previously developed single-image CAD [5]. The CAD follows a model-based approach where

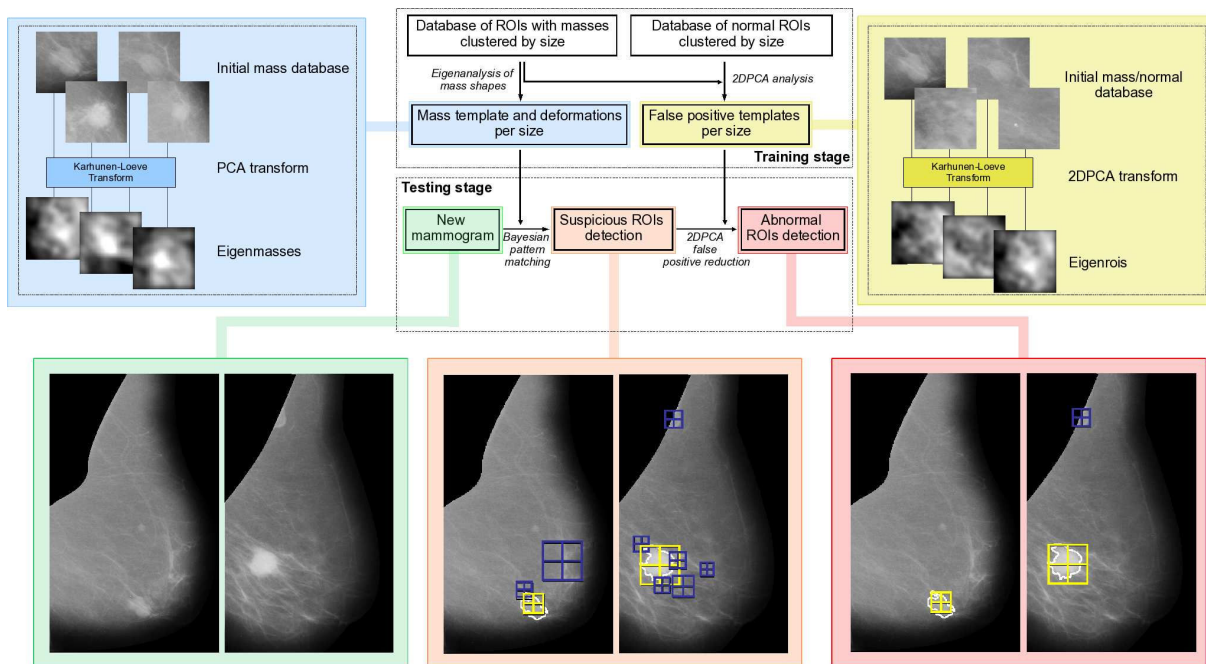


Fig. 2. Layout of the single-image CAD system (top center flow diagram), depicting graphically the main steps with details in the specific boxes. The bottom row shows the results of the CAD system: the original images, the template matching and the false positive reduction step. Note, however, that in the second mammogram a false positive remains.

an initial training stage to learn the morphology of the masses on the current database is needed. After this step, the system is able to detect masses in new mammograms.

The training stage is divided in two steps (Fig. 2). First, using a database of Regions of Interest (ROI), which each contain a mass in the center, the system automatically learns their size and shape. The approach is based on using the Karhunen-Loeve transform to take the variation of the mass shapes into account. The output of this initial step is a template per each size. Templates are defined by their mean shape and possible deformations. The second step of the training stage is directed to false positive reduction, which in this case means to distinguish between normal tissue and real masses. For this purpose the training images should now include instances of real masses and also instances of ROIs being normal tissue. Hence, a second database is constructed by adding normal ROIs to the mass database. Regarding the number of mass ROIs, three times as many normal ROIs were included for each size-cluster. This ratio results in a good compromise between the performance of the false positive reduction algorithm and the huge variability of false positive ROIs. Therefore, the output of this second learning step is a classifier (k-Nearest Neighbor) trained to predict if a ROI contains a mass or normal tissue.

Once the system has been trained, it is ready to detect masses in unseen mammograms. This testing stage is, again, divided in two steps (Fig. 2). The first stage is focused on the detection of all suspicious regions, while the second (false positive reduction) stage aims to classify the detected regions as normal tissue or masses. The detection step consists in matching the templates to regions in the mammograms. According to the used probabilistic framework, the objective

function to minimize is (see the original work for details [5]):

$$\lambda = \sum_{k=1}^N (\xi_k - 1) + \Omega(\psi^{s,\xi,d}, Y) \quad (1)$$

where ξ_i are the set of parameters that models the template deformations and $\Omega(\psi^{s,\xi,d}, Y)$ is the potential function that measures the agreement between the template deformed with parameters ξ_i and the image itself. Therefore, this function λ consists of two terms: a first one that measures the deviation of the deformed template from the prototype (hence penalizing larger deviations from the template), and a second one which describes the fitness of the deformed template to the boundaries of the image (as more closer the template to the image, less is the potential value). The output of this step is, hence, a set of detected suspicious regions. In order to ensure that these regions are real masses, the false positive reduction step is applied. Each suspicious region is cropped from the image and used as the input to the classifier. Therefore, only the regions classified as being real masses are marked in the mammogram.

Note that in the original approach of Freixenet et al. [5] we introduce breast density information in order to improve the results of the algorithm. However, in this work we do not use this information since we want to test the improvement of using bilateral information into the CAD algorithm. In further works, we will merge both steps.

B. Registration methods

Different intensity based image registration algorithms for bilateral mammographic images are evaluated. Transformation parameters are recovered by maximizing a similarity

measure using an optimization approach. Sum of squared differences (SSD) and Mutual information (MI) have been used as metrics [6]. Based on their application to medical imaging, the following image registration methods are evaluated in this work:

Global Methods: *Rigid* (translation and rotation, 4 parameters in 2D) and *Affine* transformations (allowing additional shearing independent in each dimension and scaling, for a total of 6 parameters) have been evaluated from the perspective of global methods, in which all pixels suffer the same transformation.

B-splines: *B-splines* is a deformable registration method where pixels are transformed locally, having a different transformation depending on their local similarity and position. B-spline (BSP) free form deformations [7] has been selected, due to its wide popularity in medical applications although not widely tested in x-ray mammographic images. The BSP algorithm is based on deforming an image by modifying a mesh of control points following a maximization of a similarity measure. These control points define a mesh of smooth and continuous BSP functions with limited support (modifying a control point only affects neighboring points).

Multi-resolution and Algorithm Combination: Although registration methods are often used independently it is commonly accepted that results can be improved in terms of accuracy and robustness by using a multi-resolution (MR) approach or by combining different approaches. MR is based on registering the images in a lower resolution, propagating parameter estimation into a higher resolution and registering again. This often avoids local minima in the parameter search space and reduces computational time. Algorithm combination exploits the benefits of the different methods, for instance using a global and a local method. In this case, the global method recovers for main pose and scale differences and the local method accounts for localized non-linear deformations.

We have evaluated *Affine MR*, *BSP MR*, *Affine* in combination with *BSP* (*Affine & BSP*) and *Affine MR* in combination with *BSP MR* (*Affine MR & BSP MR*).

C. Dual-image CAD system

In order to benefit from including registration information, we propose to integrate this information in the detection approach. This information can be integrated either as *a priori* or *a posteriori* information:

- In the *a priori* case, the registration is performed before the detection starts. Hence, the comparison of both left and right mammograms can be inserted in the probabilistic detection framework by modifying the potential function.
- In contrast, in the *a posteriori* case, the detection step is executed independently of the registering algorithm. At the end of the algorithm, however, the bilateral information can be used as a way to reduce the number of false positives. If a detection is found in the same position in both mammograms, this would probably be due to the same internal breast structure.

Note however, that in the second case, the sensitivity of the CAD can not be improved. Hence, in this work, we focused on the first case, i.e. introducing bilateral information in an *a priori* way.

In order to obtain the bilateral information, we register both left and right mammograms using one of the above mentioned approaches. More specifically, if we want to detect masses in the left mammogram with the CAD algorithm, this left mammogram is the reference image, and we register the right mammogram to be similar to the left. In contrast, if we analyse the right mammogram, the registered (transformed) image is the left one. Therefore, suspicious regions are found by subtracting the original mammogram with the registered one. We actually discard negative values since brighter structures in the bilateral mammogram would produce artifacts in the testing mammogram (note that otherwise, a mass present in the bilateral mammogram would affect the analyse one).

There are different options to introduce this subtracted image to our single-image CAD. In this work, we multiply the original mammogram to be analysed with this subtracted image. In some sense, we are dealing with this bilateral information as an enhancement procedure. Hence, as the result of the multiplication, regions of the mammogram with similar structures are darkened, while regions with different structures are highlighted. Therefore, using this bilateral information, we introduce *a priori* information to the potential function Ω used in Eq. 1. This step fits in Fig. 2 in the *testing stage* before *Bayesian pattern matching*.

III. RESULTS

A. Data

We used two different databases: the Mammographic Image Analysis Society (MIAS) database [8] and the Digital Database for Screening Mammography (DDSM) [9]. For the MIAS database we have obtained detailed annotations of the masses, while for the DDSM database this is not available. This determines that we can only use the MIAS database for detailed evaluation, but DDSM for training (see [5] for further details). The experimental results presented in this paper were performed using a subset of 208 mammograms extracted from MIAS, including 104 pairs of left and right Medio-Lateral Oblique (MLO) mammograms. In total there were 52 mammograms containing at least one mass.

B. Registration methods

For evaluating the results of registration methods we computed similarity metrics before and after registration to obtain an indication of how similar the bilateral images were. A higher similarity after image registration was expected and the method with the highest similarity was considered the best candidate to be integrated with the single-image CAD system. For all the images in the database we calculated SSD and MI metrics. Table I presents the numeric results for the complete database for both metrics. As can be seen, results are useful to compare registration methods using the same metric but is not possible to compare results between metrics due to the fact that the results are related

	SSD		MI	
	Mean	SD	Mean	SD
Before registration	2369.33	1656.87	0.73	0.24
Rigid	795.84	502.77	0.83	0.27
Affine	503.03	310.15	1.05	0.21
BSP	277.51	131.69	1.34	0.23
Affine MR	788.25	476.13	1.05	0.21
BSP MR	218.56	100.93	1.34	0.23
Affine & BSP	276.69	143.14	1.37	0.21
Affine MR & BSP MR	221.30	111.29	1.38	0.22

TABLE I

EVALUATION OF THE ANALYSED REGISTRATION METHODS.

to the metric optimized by the algorithm. With both metrics the standard deviation (SD) is large with respect to mean values. This can be explained by the heterogeneity of the images and, nevertheless, SD decreases after registration. The methods that obtained the best results were *BSP MR* for SSD metric and *Affine MR & BSP MR* for MI metric. Once the best methods for each metric were chosen, another criteria, visual assessment, was performed to determine the metric to be used. Difference images after registration were reviewed by different observers to evaluate dissimilarities from a global point of view where MI obtained better results than SSD [10]. Therefore, *Affine MR & BSP MR* method with MI metric was the one used to test the proposed dual-image CAD system.

C. CADs comparison with and without bilateral information

To perform the quantitative evaluation we used ROC and FROC analysis. In ROC analysis, a graphical curve represents the true positive rate as a function of the false positives rate. Moreover, the percentage value under the curve (A_z) is an indication for the overall performance of the observer, and is typically used to analyse the performance of the algorithms. On the other hand, in FROC analysis the Lesion Localisation Fraction (LLF) is plotted against the Non-Lesion Localisation Fraction [11]. Note that in this analysis the definition of a detected region is needed. In this paper we assume that a region is detected if the overlap between that region and the suspicious region is at least 50%.

Without taking the bilateral information into account and using the above mentioned dataset, our CAD system obtained $A_z = 0.716$. In contrast, when adding this information this value increased to $A_z = 0.852$. On the other hand, Fig. 3 shows the results of the FROC analysis without and with using the bilateral information. Note that at lower sensitivities this information was not useful. However, at higher sensitivities the dual-image CAD improved the single-image one. For instance, at a LLF of 80% the single-image CAD obtained 1.68 false positives per image, while the dual-image CAD reduced to 0.90, and at a LLF of 88% the false positives per image were 1.85 and 0.99, respectively. The obtained results show the benefits of including the bilateral information, increasing the sensitivity of the CAD at less false positives per image.

IV. CONCLUSIONS AND FUTURE WORKS

This paper has presented a comparison for a mass detection CAD system when using or not registration information

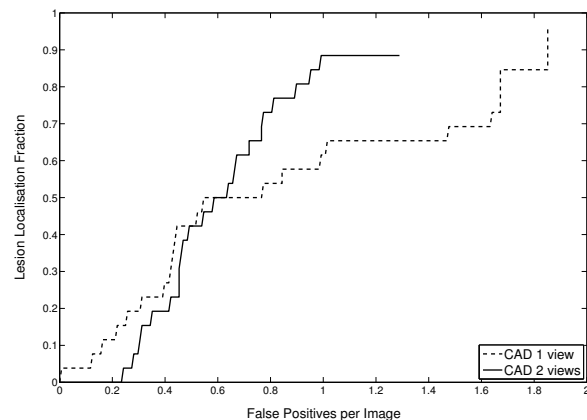


Fig. 3. FROC analysis for the CAD system without (CAD 1 view) and with (CAD 2 views) considering bilateral information.

of bilateral mammograms. According to the obtained results, the introduction of the registration information as *a priori* information is considered as an improvement of our single-image CAD system.

Future work will focus on two different ways. We will investigate firstly, about the influence of introducing breast density information in the training step of our CAD and secondly, about the use of the registration information as a false positive reduction method. Furthermore, we would like to extend this work using a full-field digital database, where we expect to improve the results since it is known that digital technology allows a better discrimination of the internal breast tissues [12].

REFERENCES

- [1] T.W. Freer and M.J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12860 patients in a community breast center", *Radiology*, vol. 220, pp. 781-786, 2001.
- [2] C.M. Kuzmiak, G.A. Millnamow, et al., "Comparison of full-field digital mammography to screen-film mammography with respect to diagnostic accuracy of lesion characterization in breast tissue biopsy specimens", *Acad. Radiol.*, vol. 9, pp. 1378-1382, 2002.
- [3] D. Kopans, *Breast Imaging*, Lippincott-Raven, Philadelphia; 1998.
- [4] L. Tabar, C.J.G. Fagerbert, et al., "Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare", *Lancet*, vol. 1, pp. 829-832, 1985.
- [5] J. Freixenet, A. Oliver, et al., "Eigendetection of masses considering false positive reduction and breast density information", *Med. Phys.*, vol. 35, no. 5, pp. 1840-1853, 2008.
- [6] B. Zitova, "Image registration methods: a survey", *Image Vision Comput.*, vol. 21, no. 11, pp. 977-1000, 2003.
- [7] D. Rueckert, L.I. Sonoda, et al., "Nonrigid registration using free-form deformations: application to breast MR images", *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712-721, 1999.
- [8] J. Suckling, J. Parker, et al., "The Mammographic Image Analysis Society digital mammogram database", *IWDM*, pp. 211-221, 1994.
- [9] M. Heath, K. Bowyer, et al., "The Digital Database for Screening Mammography", *IWDM*, pp. 212-218, 2000.
- [10] Y. Díez, A. Oliver, et al., "Comparison of Registration Methods Using Mammographic Images", *ICIP*, to appear, 2010.
- [11] H.J. Yoon, B. Zheng, et al., "Evaluating computer-aided detection algorithms", *Med. Phys.*, vol. 34, no. 6, pp. 2024-2038, 2007.
- [12] U. Fischer, F. Baum, et al., "Comparative study in patients with microcalcifications: full-field digital mammography vs screen-film mammography", *Epidem. Rev.*, vol. 11, no. 12, pp. 2679-2683, 2002.