Similarity Metrics for Intensity-Based Registration Using Breast Density Maps

Eloy García[1(✉)], Arnau Oliver[1], Yago Diez[2], Oliver Diaz[1], Xavier Lladó[1], Robert Martí[1], and Joan Martí[1]

[1] Institute of Computer Vision and Robotics, University of Girona, Girona, Spain
egarcia@silver.udg.edu
[2] Tokuyama Laboratory GSIS, Tohoku University, Sendai, Japan

Abstract. Intensity-based registration algorithms have been widely used in medical image applications. This type of registration algorithms uses an object function to compute a transformation and optimizes a measure of similarity between the images being registered. The most common similarity metrics used in registration are sum of squared differences, mutual information and normalized cross-correlation. This paper aims to compare these similarity metrics, using common registration algorithms applied to breast density maps registration. To evaluate the results, we use the protocols for evaluation of similarity measures proposed by Škerl et al. They consist in defining a set of random directions in the parameter space of the registration algorithm and compute statistical measures, such as the accuracy, capture range, number of maxima and risk of non-convergence, along these directions. The obtained results show a better performance corresponding to normalized cross-correlation for the rigid registration algorithm, while the sum of squared difference obtains the best result for the B-Spline method.

1 Introduction

Breast density maps is a type of image in which the intensity pixels values correspond to the glandular tissue thickness presented in the mammogram. Figure 1 shows an example of these images, obtained using the commercial software Volpara^{TM}. Establishing a correlation between two density maps, alike to mammograms, can improve the detection and follow-up of breast diseases. This is not an easy task. Due to the different patient positioning during the acquisition, image registration is needed to compensate the errors yielded [4,7].

Several registration algorithms have been proposed to establish spatial correspondence between images [3,5]. The registration problem is expressed as a minimization of an energy composed of a regularization and a similarity term. Registration algorithms can be divided into two categories: feature- and intensity-based algorithms. In particular, intensity-based registration algorithms consist in maximizing the matching criterion, usually a similarity metric between images. This kind of algorithms can be decomposed into three principal components [1]:

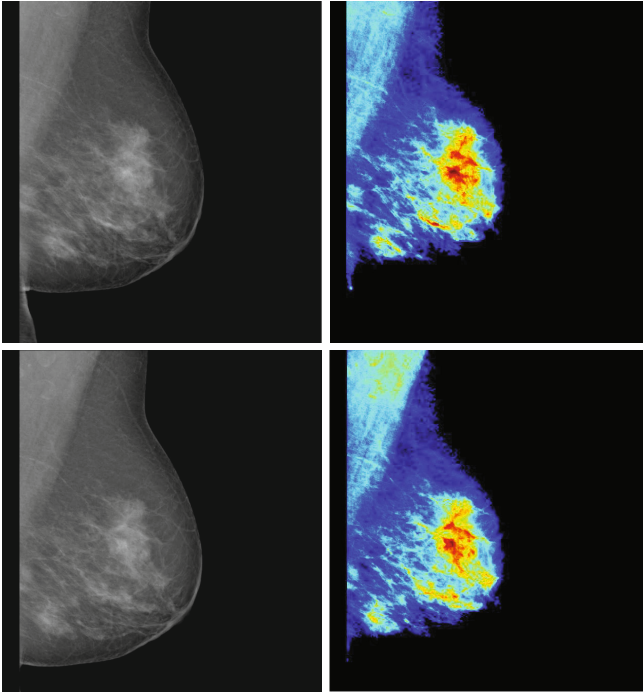– the similarity metric, which specifies the agreement between the images,

Fig. 1. Pair of repeated mammograms (left), acquired within a short time interval, and their respective density maps (right) computed by the commercial software Volpara^{TM}. Red color represents denser areas (Color figure online).

– the transformation model, which defines the transformation of the source image to match the target,
– the optimization strategy, which improves the parameters of the transformation model, based on the metric.

The definition of the similarity criterion relies on the nature of image gray-level dependencies [6] and, moreover, the most popular optimization algorithms, such as the gradient-descent algorithm, use the surface defined by the metric in the parameter space to reach the optimum. Therefore, the similarity metric is one of the factors that affect the quality of registrations [8]. Intensity-based metrics can define non-convex topologies. Defining a suitable metric reduces the likelihood of obtaining a suboptimal solution by means of locating a small number of maxima, reducing the risk of non-convergence or maximizing the capture range of the global maxima during the optimization. Figure 2 shows a cost-function surface, from a rigid registration method, using the three similarity metric analyzed in our study: sum of squared differences (SSD), mutual information (MI) and normalized cross-correlation (NCC). Notice that the optimal solution does not change widely its position in the search space. However, sub-optimal solutions (i.e. local maxima) are found using some metrics but not others.
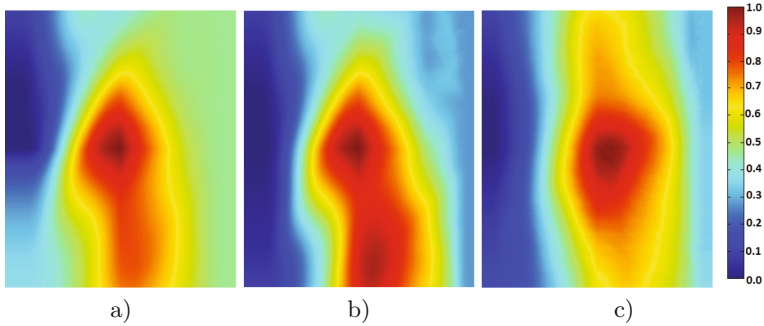
Fig. 2. (a) Sum of squared difference, (b) Mutual information and (c) Normalized cross-correlation applied to measure similarity between two density maps within the search space of a rigid registration process. The higher the values, the higher the agreement between images. The global optimal value is localized in the center of the images, while a local optimum value (red area below the global value) is visible in (b) (Color figure online).

The aim of this paper is to evaluate these three similarity metrics to lead the temporal registration of Volpara^{TM} Density Maps. We perform the registration using two widely used algorithms: affine and B-Spline registration methods. To evaluate the suitability of the metric, we test the protocol for evaluation of similarity measures proposed by Škerl et al. for rigid [8] and non-rigid [9] registration.

The rest of this document is organized as follows: Sect. 2 presents a brief review of the employed methodology, such as the registration algorithms, similarity metrics and evaluation, Sect. 3 describes the results obtained. The paper ends with the discussions and conclusions.

2   Methodology

2.1   Registration Methods

Affine Registration Algorithm. Affine registration carries out the rotation, translation, and scaling over the whole image. It is frequently used as an initialization step for a posterior non-rigid registration, although in some contexts it could also be used as a stand-alone registration tool given that it is less affected by registration artefacts [2].

B-Spline Free-Form (BSFF) Deformation Algorithm. This method uses a mesh of control points that are deformed using B-Spline interpolation looking for the maximization of a similarity measure. The degree of deformation of the mesh can be modeled with the resolution of the mesh. This produces deformation that, although local in nature, maintains coherence between neighboring points [2].

During this work, we use 5×5 nodes on the grid, considering just those within the images and not those corresponding to the boundary conditions.

2.2 Similarity Metrics

Sum of Squared Difference Metric. The SSD metric computes the squared differences between gray intensity values corresponding at each pixel. SSD is defined as:

$$SSD = \frac{1}{N \cdot M} \sum_{(x,y)} (I_1(x,y) - I_2(x,y))^2 \tag{1}$$

where $I_1(x,y)$ and $I_2(x,y)$ represent the intensity gray value -i.e. the breast density thickness- at pixel (x,y) for each image, I_1 and I_2, and $N \cdot M$ is the number of pixels of the images. When both images are identical, the sum of square difference is equal to zero.

Mutual Information Metric. MI is a measure of the mutual dependence between both images, related with Shannon entropy, and defined by the following equation:

$$MI = \sum_{i,j} p_{i,j} log \frac{p_{i,j}}{p_i \cdot p_j} \tag{2}$$

where p_i, p_j are the probability distributions of the individual images and $p_{i,j}$ is the joint probability distribution. During this work, the probabilities are obtained by using a histogram divided into 64 bins. Thus, MI allows to consider non-linear differences in intensity gray values. Regarding the evaluation of similarity, higher mutual information values imply higher agreement between images.

Normalized Cross-Correlation Metric. NCC is a standard statistical measure used to calculate whether two datasets are linearly related. It represents the 2D version of the Pearson's correlation coefficient and it is defined as follows:

$$NCC = \frac{\sum_{(x,y)} (I_1(x,y) - \bar{I}_1)(I_2(x,y) - \bar{I}_2)}{\sqrt{\sum_{(x,y)} (I_1(x,y) - \bar{I}_1)^2} \sqrt{\sum_{(x,y)} (I_2(x,y) - \bar{I}_2)^2}} \tag{3}$$

where $\bar{I}_1$ and $\bar{I}_2$ represent the mean of the intensity pixel values in the images I_1 and I_2 respectively.

2.3 Metric Evaluation

The evaluation protocols, for rigid and non-rigid registration algorithms, proposed by Škerl et al. are based on systematic simulations of the transformation state along random directions in the parameters space of the registration method. The following sections briefly expose these methods. For further references see [8,9].
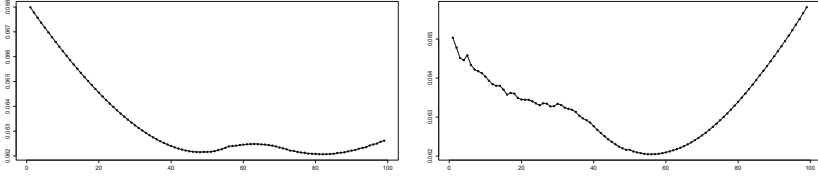
Fig. 3. Normalized SSD along two lines in the parameter space, composed of 100 points. The solution obtained during the registration corresponds to the middle (point number 50). Notice that the maximum similarity is obtained when SSD is the minimum.

Affine Registration Evaluation. Once the affine registration has reached an optimal solution, we can define a hypersphere in the parameter space. Within the hypersphere, we can define a random direction crossing from one side to the other, and traversing the center of the sphere -i.e. the optimal solution- like the radius of a wheel. Finally, we compute the similarity metric using the parameters defined along this direction. Figure 3 shows two of these lines, once the results have been normalized.

During this work, we use 50 random directions, using 100 points per line. The parameter space was defined considering the degrees of freedom of the whole image. Hence, the source image was allowed to slide in the interval $[-L_X, L_X]$ in the X direction, being L_X the length of the target image in the X direction, and $[-L_Y, L_Y]$ in the Y direction, being L_Y the length of the target image in the Y direction. The rotation parameter is defined between $-\pi$ and $+\pi$ radians. Finally, the scale factor vary from 0.5 and 1.5 due to the mammographic acquisition does not allow big changes in the breast shape.

BSFF Deformation Evaluation. In this case, localizing a hypersphere in the parameter space is a more challenging problem. Therefore, to define the direction to obtain the sample point, first, all nodes composing the mesh are initialized using a random displacement from that obtained during the registration. Then each node, consecutively, is taken as a control point. Around the control point, a circle with radius R is defined. Similarly to the previous method, we use this circle to vary the control point position, sampling along one defined direction.

For each node, 20 lines composed of 100 points were defined. In this case, the parameter space was defined in a node-wise way. Nodes are initialized considering the neighborhood, to avoid the overlap or inversion of position between two nodes. Thus, the search space of the node N_i is within the interval $[X_E, X_W]$ in the X direction, being X_W the x position of the node at left and X_E of the node at right, and $[Y_N, Y_S]$ in the Y direction, being Y_N the y position of the node above and Y_S of the node at below.

Statistics. First, during the evaluation, parametric space features and similarity values are normalized to avoid preferred variables. For instance, the translation

parameters in millimeters are much larger than the rotation parameters in radians. This fact can induce to an error during the conclusions. To evaluate the similarity metrics, we use the following features:

– Accuracy (ACC), defined as the Euclidean distance between the initial optimal position X_0 and the global maxima $X_{n,max}$ in the line n. If the initial position is the actual global maxima, $ACC = 0$.

$$ACC = ||X_{n,max} - X_0|| \tag{4}$$

– Distinctiveness of optimum (DO) is the average change of the similarity metric SM within distance r, close to the global maxima. DO shows the sharpening of the similarity metric around the global maxima.

$$DO(r) = \frac{2 * SM(X_{n,max}) - SM(X_{n,max-r/\sigma}) - SM(X_{n,max+r/\sigma})}{2r} \tag{5}$$

where σ represents the distance between two consecutive points along a probing line.
– Capture range (CR) is the distance between the global maxima and the closest minima along the line.

$$CR = min(||X_{n,max} - X_{n,min}||) \tag{6}$$

– Number of maxima (NOM) is the number of maxima -i.e. suboptimal solution- in each line. We have changed this feature (originally number of minima) considering that, in our case, the maxima represents a suboptimal solution.
– Risk of Non-convergence (RON) is the average of positive gradients $d_{n,m}$ within distance r from the global maxima in each line.

$$RON(r) = \frac{\sum_{n,m} d_{n,m}}{2r} \tag{7}$$

Therefore, the better a similarity measure, the smaller the accuracy, number of maxima and risk of non-convergence and the larger the values of capture range and distinctiveness of optimum.

3   Results

The dataset was composed of 21 pairs of mammographic images (42 FFDMs in total) from 21 women, including 14 pairs of CC and 7 pairs of MLO projections. Each image pair corresponds to mammograms acquired within a very short time interval, few minutes. We used the commercial software Volpara^{TM} (Volpara Solutions[1], Wellington, New Zealand) to extract the density map of the mammogram. Both images were registered using affine registration and BSFF deformations. All registration methods were implemented using the InsightToolkit (ITK v.4.8.0) libraries. The optimization followed a gradient

---

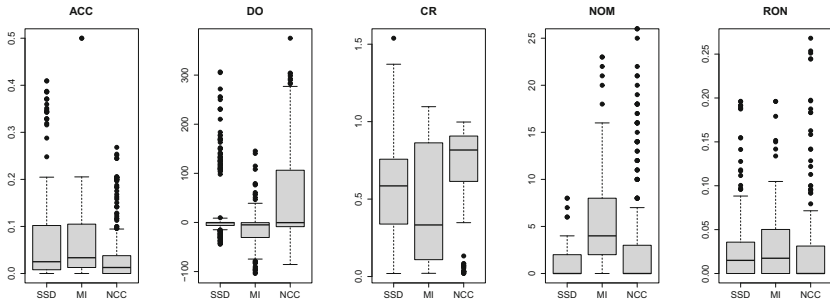[1] http://volparasolutions.com/.

Fig. 4. Statistical measures for affine registration. The better a similarity measure, the smaller the ACC, NOM and RON and the larger the values of DO and CR.
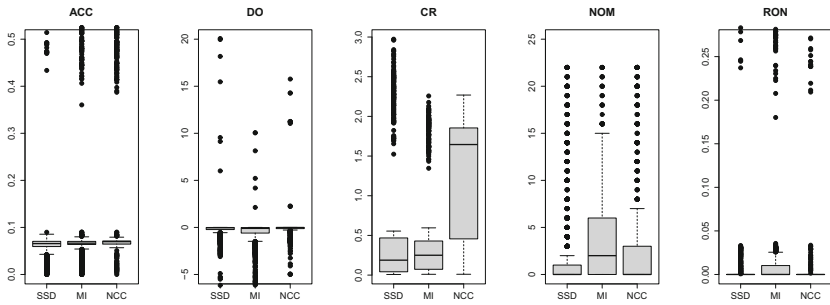

Fig. 5. Statistical measures for B-Spline registration. The better a similarity measure, the smaller the ACC, NOM and RON and the larger the values of DO and CR.

descent approach, while linear interpolation was performed for pixel interpolation. Data analysis and statistical tests were carried out using the statistical software R (v.3.0.3).

Figure 4 shows the results obtained for affine registration. The median and quartile deviation of the boxplot represent the accuracy and robustness for each metric. For instance, in this case, NCC shows a better performance in all of cases. The median of the NCC distributions for ACC, NOM and RON are smaller than those corresponding to the other metrics and, similarly, they are bigger for CR and DO. However, the same distribution shows a higher quartile deviation for NOM and a large number of outliers for NOM and RON.

On the other hand, Fig. 5 shows the results corresponding to the B-Spline registration. In this case, the results obtained with SSD shows a better performance in ACC, NOM and RON; NCC obtains the best results of CR.

4   Discussion and Conclusion

In this work, we have evaluated the three most common image similarity metrics used for intensity-based registration of density maps. To achieve this goal,

we have used the protocols proposed by Škerl et al. [8,9]. In our experience, these protocols show a different performance with respect to the registration algorithm. During the evaluation of the affine registration, the protocol shows a good performance. Modifying the registration parameters allows a sampling of all the features in the hyperspace and an accurate interpretation of the parameter space around the localized optimal solution. Furthermore, ACC, CR and NOM show NCC as a robust and stable metric to lead the affine registration. The risk of non-convergence is smaller while the capture range and the accuracy of this metric is high, defining NCC as the preferred metric.

During the B-Spline registration, the number of nodes has a big impact in the likelihood of localizing a suboptimal solution. The more nodes, the smaller the local deformation and, therefore, the higher the number of maxima at each line. Furthermore, the nodes situated in the black background of the mammogram have a small impact over the metric but high in the statistics exposed. For instance, a large number of points in the black background could produce a high ACC because the movement of these points have a small impact over the final result of the similarity metric. However, these points are considered similarly to those with a high impact over the metric in the statistics. We consider that the proposed protocol is not suitable in this case. Establishing the impact of each node is mandatory to avoid wrong results or misinterpretations.

Finally, the metrics used show different behavior with respect to the registration algorithms. While NCC shows a clearly better performance during the affine registration, SSD may be the best option for B-Spline registration, due to the better performance obtained for ACC, NOM and RON. However, there is not a significant difference with respect to those obtained using NCC, which obtains a clearly better value for CR.

References

1. Crum, W.R., Hartkens, T., Hill, D.: Non-rigid image registration: theory and practice. Br. J. Radiol. **77**, 140–153 (2004)
2. Diez, Y., Oliver, A., Lladó, X., Freixenet, J., Martí, J., Vilanova, J.C., Martí, R.: Revisiting intensity-based image registration applied to mammography. IEEE Trans. Inf. Technol. Biomed. **15**(5), 716–725 (2011)
3. Guo, Y., Sivaramakrishna, R., Lu, C.C., Suri, J.S., Laxminarayan, S.: Breast image registration techniques: a survey. Med. Biol. Eng. Comput. **44**(1–2), 15–26 (2006)
4. Hill, D.L., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical image registration. Phys. Med. Biol. **46**(3), R1 (2001)
5. Maintz, J.A., Viergever, M.A.: A survey of medical image registration. Med. Image Anal. **2**(1), 1–36 (1998)
6. Roche, A., Malandain, G., Ayache, N.: Unifying maximum likelihood approaches in medical image registration. Ph.D. thesis, Inria (1999)

7. Rueckert, D., Schnabel, J.A.: Medical image registration. In: Deserno, T.M. (ed.) Biomedical Image Processing. Biological and Medical Physics, Biomedical Engineering, pp. 131–154. Springer, Heidelberg (2010)
8. Škerl, D., Likar, B., Pernus, F.: A protocol for evaluation of similarity measures for rigid registration. IEEE Trans. Med. Imaging **25**(6), 779–791 (2006)
9. Škerl, D., Likar, B., Pernus, F.: A protocol for evaluation of similarity measures for non-rigid registration. Med. Image Anal. **12**, 42–54 (2008)