Point similarity measure based on mutual information

Peter Rogelj and Stanislav Kovačič

Faculty of Electrical Engineering, University of Ljubljana
Ljubljana, SI-1000, Slovenia
`peter.rogelj@fe.uni-lj.si,stanislav.kovacic@fe.uni-lj.si`

Abstract. Registration of multi-modality images requires similarity measures that can deal with complex and unknown image intensity dependencies. Such measures have to rely on statistics, and consequently, they require relatively large image regions to operate. This makes the detection of localized image discrepancies difficult. As a solution we propose point similarity measures, which can measure similarity of arbitrarily small image regions, including similarity of individual image points. In this paper we present a point similarity measure derived from the mutual information. In addition to its extreme locality it can also avoid the interpolation artifacts and improve the spatial regularization to better suit the spatial deformation model.

1 Introduction

A criterion function used in registration procedures comprises some measure of image similarity, which is used for measuring the quality of image match, and possibly a geometric regularization, which prevents unrealistic deformations. However, not only the criterion function, also its implementation may have an important influence on the registration results. The influence of implementation is the most obvious for multi-modality similarity measures, when they are used to detect localized image discrepancies.

The ability to detect local image discrepancies is crucial for the success of non-rigid registration. The most straightforward approach to detect local image discrepancies is to measure the similarities of small image regions. In the case of conventional multi-modality similarity measures, e.g. mutual information measures [15, 2, 13], this is not directly applicable. Due to their statistical nature they require relatively large image regions to operate and thus cannot directly detect local image properties. Although these measures are not local, they are still locally sensitive, which enables assessment of local discrepancies by measuring global image similarity, i.e. similarity of the whole images, subject to local deformations. This is a general approach used by many authors, [12, 8, 11] to mention a few. The main problem with this approach is its high computational cost, which originates in large number of recomputations of global image similarity. This practically limits the dimensionality of the transformation and thus the locality of image discrepancies that can be corrected. In this work we focus

on an alternative approach, which follows the basic idea of detecting local image discrepancies by measuring similarity of small image regions. To make this approach possible we introduced point similarity measures [10]. They are designed to measure similarity of arbitrarily small image regions, including similarity of individual image points, which also holds in the case of multi-modality data.

In this paper we present point similarity measure derived from the mutual information [15, 2], and compare it with the original mutual information measure in terms of interpolation artifacts and spatial regularization.

2 Point similarity measures

We define point similarity measures as measures that can measure similarity of individual image points. Obviously they can also be used to measure similarity of image regions of arbitrary size. Point similarity measures can be derived from global similarity measures. Here we derive a point similarity from the mutual information, which is the most frequently used multi-modality similarity measure.

2.1 Point similarity measure derived from the mutual information

Mutual information (MI) of two images (A and B) is defined by marginal and joint entropies:

$$MI = H(A) + H(B) - H(A, B), \tag{1}$$

and can be computed as follows:

$$MI = \sum_{\mathbf{i}} p(\mathbf{i}) \log \left(\frac{p(\mathbf{i})}{p(i_A)\, p(i_B)} \right). \tag{2}$$

Here, i_A and i_B are image intensities of images A and B, $\mathbf{i} = [i_A, i_B]$ denotes an intensity pair, $p(i_A)$ or $p(i_B)$ are marginal intensity distributions and $p(\mathbf{i})$ is a joint intensity distribution. Eq. (2) can be rewritten in the following form:

$$MI = \sum_{\mathbf{i}} \frac{N_{\mathbf{i}}}{N} \log \left(\frac{p(\mathbf{i})}{p(i_A)\, p(i_B)} \right) = \frac{1}{N} \sum_{v} \log \left(\frac{p(\mathbf{i}(v))}{p(i_A(v))p(i_B(v))} \right), \tag{3}$$

where $N_{\mathbf{i}}$ is the number of occurrences of intensity pair $\mathbf{i}$ and N is the total number of intensity pairs in the image, which equals the number of overlapping image voxels. Furthermore, $\mathbf{i}(v) = [i_A(v), i_B(v)]$ stands for image intensities located at voxel v. Note that the final summation is taken over the spatial image coordinates instead of the intensities. Thus, the global similarity MI can be treated as an average of point similarities $S_{MI}(v)$, defined for each voxel v.

$$MI = \frac{1}{N} \sum_{v} S_{MI}(v), \tag{4}$$

$$S_{MI}(v) = \log\left(\frac{p(\mathbf{i}(v))}{p(i_A(v))p(i_B(v))}\right). \tag{5}$$

In general, the point similarity can be estimated not only for image voxels, but for each image point pair $[A(\mathbf{x}_1), B(\mathbf{x}_2)]$ from the corresponding intensity pair $\mathbf{i}(\mathbf{x}_1, \mathbf{x}_2) = [i_A(\mathbf{x}_1), i_B(\mathbf{x}_2)]$, such that $S_{MI}(\mathbf{x}_1, \mathbf{x}_2) = S_{MI}(\mathbf{i}(\mathbf{x}_1, \mathbf{x}_2))$, where

$$S_{MI}(\mathbf{i}) = \log\left(\frac{p(\mathbf{i})}{p(i_A)p(i_B)}\right). \tag{6}$$

The function $S_{MI}(\mathbf{i})$ is called a point similarity function and is an estimate of the intensity dependence between the images when they are correctly registered. The measurement of point similarity therefore consists of two steps. In the first step the point similarity function $S_{MI}(\mathbf{i})$ is estimated from the whole images A and B. In the second step point similarity $S_{MI}(\mathbf{x}_1, \mathbf{x}_2)$ is obtained from the point similarity function $S_{MI}(\mathbf{i})$ by simply pointing to a certain value by the corresponding image intensity pair $\mathbf{i}(\mathbf{x}_1, \mathbf{x}_2)$, as illustrated in Fig. 1. Note that the point similarity function needs to be estimated only once for all the measurements of point similarity at some image configuration, which makes the measurement of similarity computationally efficient.
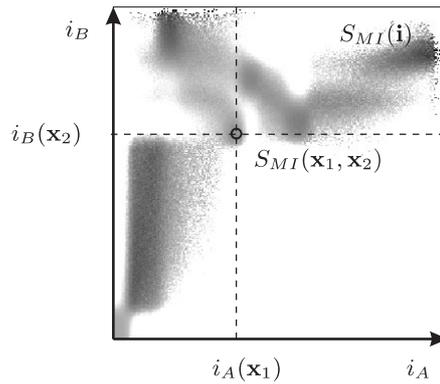
Fig. 1. Measurement of point similarity. The point similarity function $S_{MI}(\mathbf{i})$ is estimated from the whole images and defines the similarity with respect to the image intensities (darker color represents higher similarity). For example, a point similarity $S_{MI}(\mathbf{x}_1, \mathbf{x}_2)$ can be obtained from the point similarity function $S_{MI}(\mathbf{i})$ by pointing to a certain value by corresponding image intensity pair $\mathbf{i} = [i_A(\mathbf{x}_1), i_B(\mathbf{x}_2)]$.

2.2 Similarity of an image region

According to the Eq. (4) the mutual information of the whole images can be computed from point similarities by averaging. Let us rewrite it in the following

form:

$$MI = \frac{1}{N} \sum_v S_{MI}(v) = \sum_{\mathbf{i}} \frac{N_{\mathbf{i}}}{N} S_{MI}(\mathbf{i}) = \sum_{\mathbf{i}} p(\mathbf{i}) S_{MI}(\mathbf{i}). \qquad (7)$$

The same principle could be used for computing similarity of smaller image regions, with the only difference that in this case the summation (averaging) runs only over the voxels in the region. Thus, the similarity S_R of the region R can be obtained by using probability distribution $p_R(\mathbf{i})$ of this region instead of probability distribution $p(\mathbf{i})$ of the whole images:

$$S_R = \frac{1}{N_R} \sum_{v \in R} S_{MI}(v) = \sum_{\mathbf{i}} p_R(\mathbf{i}) S_{MI}(\mathbf{i}), \qquad (8)$$

where N_R is the number of voxels in the region R. Note that in all the cases the point similarity function $S_{MI}(\mathbf{i})$ is estimated from the whole images, in contrast to mutual information, which uses only the region that is being measured. This is a small but important difference between point based similarity measures and global measures, which enables the first ones to better assess the intensity dependence between the images and thus to improve the quality of local similarity measurement, see Fig. 2.
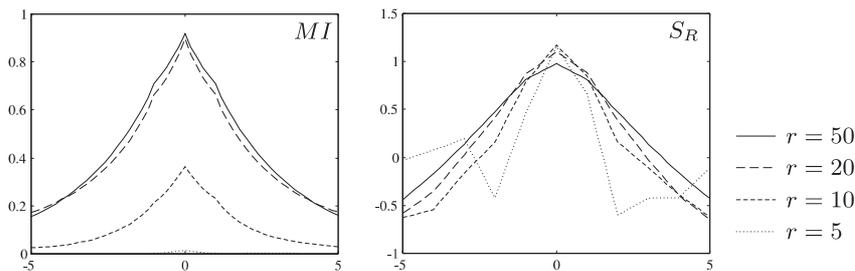
Fig. 2. Mutual information MI (left) and point based similarity measure S_R (right) with respect to image translation, for different sizes of image region r^3; $r = \{50, 20, 10, 5\}$. Point based similarity measure was always based on the same point similarity function $S_{MI}(\mathbf{i})$, obtained from the whole images at the correct image alignment. Note that the local sensitivity of the MI decreases with decreasing r, while the sensitivity of S_R remains practically the same.

3 Similarity and image transformation

Image registration methods search for such a transformation that maximizes the image similarity. Let us analyze how the similarity changes when transforming image B with transformation $\mathbf{T}$. The transformation moves each point $B(\mathbf{x})$ from its original position $\mathbf{x}$ for some displacement $\mathbf{T}(\mathbf{x})$ to a new position $\mathbf{x} + \mathbf{T}(\mathbf{x})$,

where it gets matched with a point $A(\mathbf{x} + \mathbf{T}(\mathbf{x}))$. One of the important issues is how the transformation changes the similarity of some small image region (or point) in case of subvoxel displacement $\mathbf{T}(\mathbf{x})$. A common phenomenon is the appearance of the interpolation artifacts, i.e. disproportionate change of similarity with respect to the transformation, which rules out the subvoxel accuracy. Some approaches that can reduce interpolation artifacts have been proposed [9, 6] but can be applied only when the region size used for measuring the similarity is large. In case of high-dimensional registration approaches, where gradient descent optimization method is usually used, the interpolation artifacts are even more problematic, because they can cause large image misregistration.

When observing a single point in image B, its point similarity $S_{MI}(A(\mathbf{x} + \mathbf{T}(\mathbf{x})), B(\mathbf{x}))$ may change due to two reasons. The first one is the change of point pair, and the second is a possible change of the intensity distributions, which changes the point similarity function $S(\mathbf{i})$, see Eq. (6).

Let us assume that the point similarity function $S_{MI}(\mathbf{i})$ does not change and that point similarities $S(A(\mathbf{x} + \mathbf{T}(\mathbf{x})), B(\mathbf{x}))$ change only because points in image B are compared with different points in image A. However, due to the discrete nature of the images and due to the image transformation, grid points in image B do not match exactly with grid points in image A, and measuring of point similarities requires interpolation. In case of mutual information there are two interpolation methods commonly used: interpolation of intensities and partial volume interpolation. Interpolation of intensities can also be employed in case of point similarities. However, the interpolation of intensities assumes a linear intensity dependence, which may not necessarily comply with the intensity dependence estimated from the images, and can cause interpolation artifacts, as shown in Fig. 3. To avoid the interpolation artifacts, we propose to interpolate similarities instead of intensities. This is related to partial volume interpolation, which may be used for estimation of global intensity distributions. Instead of interpolating the unknown intensity $i_A(\mathbf{x} + \mathbf{T}(\mathbf{x}))$ from intensities of neighboring voxels, we directly interpolate the point similarity $S(A(\mathbf{x} + \mathbf{T}(\mathbf{x})), B(\mathbf{x}))$ from similarities of point $B(\mathbf{x})$ to neighboring grid points in image A. The weights remain the same as in the case of interpolation of intensity. This approach results in a linear relationship between the point similarity and point displacement in a range of one image voxel, thereby avoiding the interpolation artifacts, see Fig. 3. The difference between results obtained by using different interpolation methods is illustrated in Fig. 4.

The point similarities could also change due to the change of the point similarity function. In general, the transformation $\mathbf{T}$ changes the marginal and joint intensity distributions $p(i_A)$, $p(i_B)$ and $p(\mathbf{i})$. Consequently, if the point similarity function $S_{MI}(\mathbf{i})$ is recomputed using the updated distributions as defined in Eq. (6), then it changes as well. The relation between the transformation and the change of intensity distributions is complex, nonlinear and depends on the information of the whole images. Furthermore, the relation between the intensity distributions and the corresponding point similarity function is not linear either (see Eq. (6)). Nevertheless, $S_{MI}(\mathbf{i})$ is always an approximation of the same
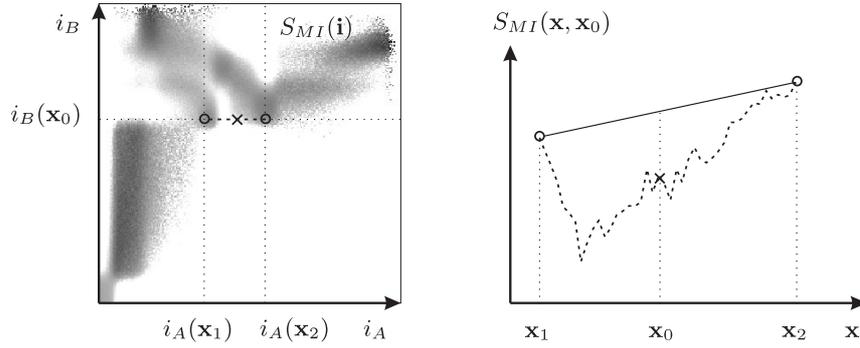
Fig. 3. Illustration of measuring point similarity with interpolation. Similarity between some voxel point $B(\mathbf{x}_0)$ and corresponding point $A(\mathbf{x}_0)$ requires interpolation. In case of interpolation of intensity, an intensity $i_A(\mathbf{x}_0)$ is interpolated from intensities of neighboring points $i_A(\mathbf{x}_1)$ and $i_A(\mathbf{x}_2)$, and the point similarity is $S(\mathbf{x}_0, \mathbf{x}_0) = S(i_A(\mathbf{x}_0), i_B(\mathbf{x}_0))$ (cross mark). Because the interpolated intensity does not comply with the complex intensity dependence estimated from the images, the similarity does not have a correct meaning, which introduces interpolation artifacts (dashed line). The problem can be solved by using interpolation of similarity, which interpolates the similarity $S(\mathbf{x}_0, \mathbf{x}_0)$ from point similarities $S(\mathbf{x}_1, \mathbf{x}_0)$ and $S(\mathbf{x}_2, \mathbf{x}_0)$, such that interpolation of intensity is not required and interpolation artifacts do not appear (solid line).
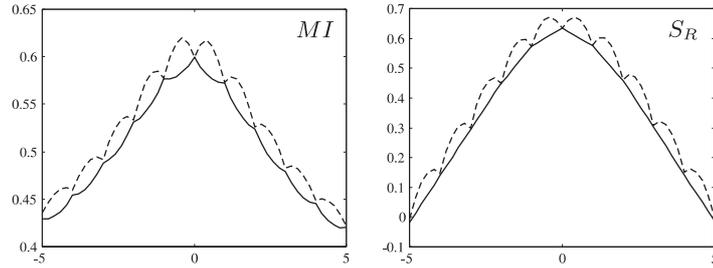
Fig. 4. An example of mutual information MI (left) and point based similarity S_R (right) with respect to image translation, for the two different interpolation methods. The dashed lines denote interpolation of intensity, while the solid lines denote partial volume interpolation for MI (left) and interpolation of similarity for S_R (right).

intensity dependence of the images when they are correctly registered. The only difference between the obtained point similarity functions is in the quality of the estimation, which depends on the level of global image mismatch. When the point similarity function is estimated at better match, the similarity better distinguishes between correct matches and mismatches, while the positions of maxima that correspond to different tissue types in the point similarity function do not change, as shown in the experiment performed using simulated Brainweb

images [4] in Fig. 5. Consequently, the registration based on point similarity measures always tends towards the transformation that would be obtained when using point similarity function estimated from the registered images, see Fig. 6. Therefore, the complex nonlinear relation between the transformation and the similarity, which also reflects in interpolation artifacts, can be avoided by keeping the point similarity function fixed. However, when the point similarity function is obtained at large misalignment, the sensitivity of similarity measure is low. Therefore, to avoid the interpolation artifacts and still achieve good sensitivity, we propose to recompute the point similarity function only once per registration step or registration iteration.

To summarize, to avoid the interpolation artifacts we keep the point similarity function fixed and use interpolation of similarity instead of interpolation of intensity.
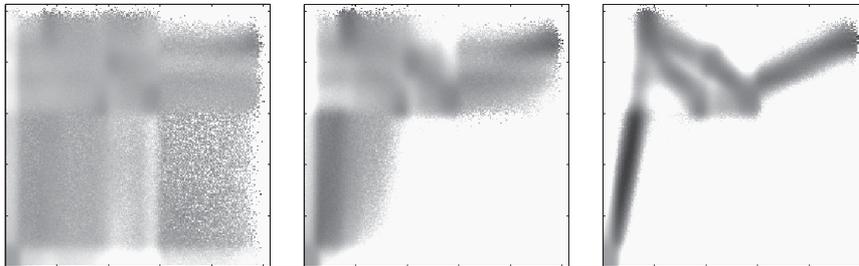
Fig. 5. Point similarity functions for simulated MRI-T1 and MRI-PD images of the head, at different levels of image mismatch: 10 mm displacement (left), 2 mm displacement (middle), and registered images (right). Darker color represents higher similarity. Note that the positions of maxima that correspond to different tissue types in point similarity function do not change.

4 Locality and spatial deformation models

Point similarity measures push the limits of the locality into extreme. Consequently, similarity of one point does not presume any spatial relation with neighboring image points. However, as stated by some authors ([3, 7]), matching of individual image points is ill-posed if they are matched independently. Registration with point similarity measures therefore requires regularization, which can be performed by a suitable spatial deformation model. Spatial deformation models are commonly used in high-dimensional registration, (for review of high-dimensional registration approaches see [14, 5]). The model can follow physical properties of deformable materials (e.g. elasticity or viscosity) or simplified/fictious properties (e.g. Gaussian regularitzation). Nevertheless, the majority of such models can be performed by convolution filtering, as proposed
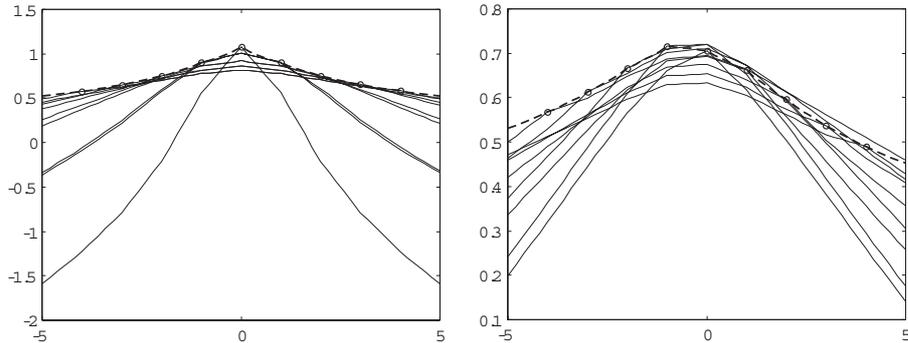
Fig. 6. Mutual information (dashed line) and point based similarities obtained using different estimations of point similarity function $S(\mathbf{i})$ (solid lines), with respect to image displacement. Similarity is measured between MRI-PD and MRI-T1 data, using simulated images (left) and real images of the head (right). Point similarity functions $S(\mathbf{i})$ were estimated at different image displacements. At this displacement the point based similarity equals the mutual information (marked with circles). All point based similarities reach the maximum at displacement 0, i.e. where images are correctly registered.

by Bro-Nielsen [1]. For example, an incremental elastic registration can be performed iteratively as follows:

$$\mathbf{T}(\mathbf{x})^{(n+1)} = \mathbf{T}(\mathbf{x})^{(n)} + G_E * \frac{\partial S(\mathbf{x})}{\partial \mathbf{T}(\mathbf{x})}, \tag{9}$$

where G_E denotes a filter with impulse response of the elastic media, S stands for some similarity measure, and n is the iteration number.

Some regularization is also provided by similarity measures when they operate on larger image regions. The similarity of an image region can be obtained by averaging the point similarities, see Eq. (8). However, the averaging over a region surrounding a point $\mathbf{x}$ can also be performed by convolution filtering with some spatial filter G_R,

$$S_R(\mathbf{x}) = G_R * S_{MI}(\mathbf{x}). \tag{10}$$

Larger the region is, wider is the impulse response of the filter G_R and more global information is extracted from the point similarities. For example, when the region spreads over the whole images and the obtained similarity equals the global mutual information, only a global image properties, appropriate for global registration (e.g. rigid) are extracted. The averaging therefore represents a kind of regularization, which extracts more global knowledge from multiple more localized image features. The regularization caused by the similarity measures is substantial, such that when the region size used for measuring the similarity is large, the additional regularization with spatial deformation model may not be necessarily required, as in [7].

When the similarity of an image region is used in combination with a spatial deformation model, the Eq. (9) can be rewritten:

$$\mathbf{T}(\mathbf{x})^{(n+1)} = \mathbf{T}(\mathbf{x})^{(n)} + G_E * \frac{\partial G_R * S_P(\mathbf{x})}{\partial \mathbf{T}(\mathbf{x})} = \mathbf{T}(\mathbf{x})^{(n)} + G_E * G_R * \frac{\partial S_P(\mathbf{x})}{\partial \mathbf{T}(\mathbf{x})}. \quad (11)$$

Here, S_P is some point similarity measure, e.g. S_{MI}. Similar results could also be obtained for other high-dimensional registration methods. The regularization is therefore duplicated, which means that the final effect does not directly follow the spatial deformation model, see Fig. 7. Point similarity measures solve this problem. They are not regularized by G_R, such that regularization remains only in the domain of spatial deformation model, which gives a full control over the transformation properties.

$$G_E \qquad\qquad G_R \qquad\qquad G_E * G_R$$
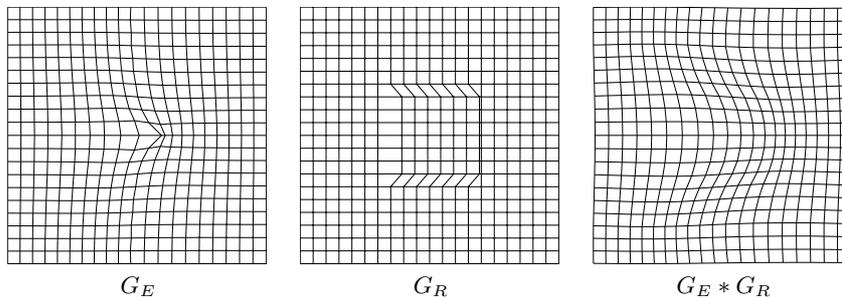
Fig. 7. An example of convolution filters used for regularizing non-rigid registration: elastic filter G_E (left), filter G_R that corresponds to region averaging (middle), and their convolution $G_E * G_R$ (right).

5 Conclusion

In this paper we presented a point similarity measure derived from the mutual information. This measure can directly detect localized image discrepancies, by measuring similarity of small image regions or individual image points. It is computationally efficient, as the point similarity function is computed only once for a given image configuration. The other advantage is the ability to avoid the interpolation artifacts. That is because point similarity measures may be estimated using the same estimation of image intensity dependencies, regardless of the actual image transformation. This could also be advantageous in the case of rigid registration. Finally, point similarity measures do not constrain the transformation, which is especially important when performing a high-dimensional registration. In this case the regularization remains only in the domain of a spatial deformation model, which does not interfere with the similarity measures.

References

1. M. Bro-Nielsen. *Medical Image Registration and Surgery Simulation*. PhD thesis, Department of Mathematical Modelling, Technical University of Denmark, 1996.

2. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information processing in medical imaging 1995*, pages 263–274. Kluwer Academic, 1995.

3. T. Gaens, F. Maes, D. Vandermeulen, and Suetens P. Non-rigid multimodal image registration using mutual information. In W.M. Wells, A. Colchester, and S. Delp, editors, *Proceedings of the 1st International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, number 1496 in Lecture Notes in Computer Science, pages 1099–1106, MIT, Cambridge, MA, USA, October 1998. Springer-Verlag.

4. R.K.-S. Kwan, A.C. Evans, and G. B. Pike. An extensible MRI simulator for post-processing evaluation. In *Visualization in Biomedical Computing (VBC'96)*, volume 1131 of *Lecture Notes in Computer Science*, pages 135–140. Springer-Verlag, May 1996.

5. H. Lester and S. R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–149, 1999.

6. B. Likar and F. Pernuš. A hierarchical approach to elastic registration based on mutual information. *Image and Vision Computing*, 19:33–44, 2001.

7. J. B. A. Maintz, H. W. Meijering, and M. A Viergever. General multimodal elastic registration based on mutual information. In K.M. Hanson, editor, *Medical Imaging 1998: Image Processing*, volume 3338 of *Proc. SPIE*, pages 144–154. SPIE Press, Bellingham, WA, 1998.

8. D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank. Non-rigid multimodality image registration. In M. Sonka and K.M. Hanson, editors, *Medical Imaging 2001: Image Processing*, volume 4322 of *Proc. SPIE*. SPIE Press, Bellingham, WA, 2001.

9. J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Interpolation artefacts in mutual information based image registration. *Computer Vision and Image Understanding*, 77(2):211–232, 2000.

10. P. Rogelj and S. Kovačič. Similarity measures for non-rigid registration. In M. Sonka and K.M. Hanson, editors, *Medical Imaging 2001: Image Processing*, volume 4322 of *Proc. SPIE*, pages 569–578. SPIE Press, Bellingham, WA, 2001.

11. G. K. Rohde, A Aldroubi, and B. M. Dawant. Adaptive free-form deformation for inter-patient medical image registration. In *The Proceedings of the SPIE Symposium on Medical Imaging 2001*, 2001.

12. D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, August 1999.

13. C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86, 1999.

14. P. Thompson and A. W. Toga. Warping strategies for intersubject registration. In I. Bankman, editor, *Handbook of Medical Image Processing*. Academic Press, 1999.

15. P. Viola and W. Wells III. Alignment by maximization of mutual information. In *Proceedings of the 5th International Conference on Computer Vision*, pages 16–23, 1995.