

Image Registration by Maximization of Combined Mutual Information and Gradient Information

Josien P.W. Pluim, J.B. Antoine Maintz, and Max A. Viergever

Image Sciences Institute, University Medical Center Utrecht
Utrecht, The Netherlands
{josien,twan,max}@isi.uu.nl

Abstract. Despite generally good performance, mutual information has also been shown by several researchers to lack robustness for certain registration problems. A possible cause may be the absence of spatial information in the measure. The present paper proposes to include spatial information by combining mutual information with a term based on the image gradient of the images to be registered. The gradient term not only seeks to align locations of high gradient magnitude, but also aims for a similar orientation of the gradients at these locations.

Results of combining both standard mutual information as well a normalized measure are presented for rigid registration of three-dimensional clinical images (MR, CT and PET). The results indicate that the combined measures yield a better registration function than mutual information or normalized mutual information per se. The accuracy of the combined measures is compared against a screw marker based gold standard, revealing a similar accuracy for the combined measures to that of the standard measures. Experiments into the robustness of the measures with respect to starting position imply a clear improvement in robustness for the measures including spatial information.

1 Introduction

Mutual information is currently a popular registration measure, which has been shown to form the base of accurate and robust registration methods by several independent studies [1, 2, 3, 4, 5], and in particular by the Retrospective Registration Evaluation Project (RREP), an international study comparing the accuracy of sixteen registration methods against a screw marker gold standard [6]. However, failure of the measure in certain situations has also been reported [7, 8, 9, 10]. Such situations can arise when the images are of low resolution, when the overlapping part of the images is small or as a result of interpolation methods. A possible solution to failure of mutual information may be to include spatial information, something that is not contained in the measure. Rueckert et al. [11] recently proposed ‘higher-order mutual information’, which incorporates spatial information by forming four-dimensional intensity

histograms. We propose to include spatial information by combining mutual information with gradient information. Image gradients by themselves have been shown to be useful registration criteria [10, 12].

2 Method

2.1 Mutual Information

The mutual information of two images is the amount of information that one image contains about the other or vice versa. When transforming one image with respect to the other such that their mutual information is maximized, the images are assumed to be registered.

The mutual information I of two images A and B can be defined as

$$I(A, B) = H(A) + H(B) - H(A, B). \quad (1)$$

Here, $H(A)$ and $H(B)$ denote the separate entropy values of A and B respectively. $H(A, B)$ is the joint entropy, i.e. the entropy of the joint probability distribution of the image intensities. In this paper, we use the Shannon measure of entropy, $-\sum_{p \in P} p \log p$ for a probability distribution P . The joint probability distribution of two images is estimated by calculating a normalized joint histogram of the grey values. The marginal distributions are obtained by summing over the rows, resp. the columns, of the joint histogram.

Recently, it was shown that the mutual information measure is sensitive to the amount of overlap between the images and *normalized* mutual information measures were introduced to alleviate this problem. Examples of such measures are the normalized mutual information $Y(A, B)$ introduced by Studholme et al. [7]

$$Y(A, B) = \frac{H(A) + H(B)}{H(A, B)}$$

and the entropy correlation coefficient $ECC(A, B)$ used by Maes et al. [2]

$$ECC(A, B) = \frac{2 I(A, B)}{H(A) + H(B)}.$$

These two measures have a one-to-one correspondence and we will therefore only use $Y(A, B)$ in this paper.

2.2 Incorporating Gradient Information

Image locations with a strong gradient are assumed to denote a transition of tissues, which are locations of high information value. The gradient is computed on a certain spatial scale. We have extended mutual information measures (both standard and normalized) to include spatial information by multiplying the mutual information with a gradient term. The gradient term is based not only on the magnitude of the gradients, but also on the orientation of the gradients.

The gradient vector is computed for each sample point $\mathbf{x} = \{x_1, x_2, x_3\}$ in one image and its corresponding point in the other image, \mathbf{x}' , which is found by geometric transformation of \mathbf{x} . The three partial derivatives that together form the gradient vector are calculated by convolving the image with the appropriate first derivatives of a Gaussian kernel of scale σ . The angle $\alpha_{\mathbf{x}, \mathbf{x}'}(\sigma)$ between the gradient vectors is defined by

$$\alpha_{\mathbf{x}, \mathbf{x}'}(\sigma) = \arccos \frac{\nabla \mathbf{x}(\sigma) \cdot \nabla \mathbf{x}'(\sigma)}{|\nabla \mathbf{x}(\sigma)| |\nabla \mathbf{x}'(\sigma)|}, \quad (2)$$

with $\nabla \mathbf{x}(\sigma)$ denoting the gradient vector at point \mathbf{x} of scale σ and $|\cdot|$ denoting magnitude.

For multimodal images, the different imaging techniques can lead to a tissue having different intensities in either image. As a result, the gradients of the images can point in different directions. However, since the images fundamentally depict the same anatomical structures, gradients in two multimodal images – at least in principle – will have the same orientation and either identical or opposing directions. Consequently, we use the following weighting function w , which favours both very small angles and angles that are approximately equal to π (see figure 1(a)):

$$w(\alpha) = \frac{\cos(2\alpha) + 1}{2}. \quad (3)$$

Furthermore, the different imaging processes of different modalities imply that multimodal images do not necessarily depict the same tissue transitions. Hence, strong gradients that emerge with a certain imaging technique may be absent or less prominent with another technique. Since we are only interested in including strong gradients that appear in *both* images, the angle function is multiplied by the minimum of the gradient magnitudes. Summation of the resulting product for all samples gives us the gradient term with which we multiply the mutual information measure. Two examples of the gradient measure (before summation) for different combinations of multimodal images can be found in figure 1(b). Tissue transitions that are depicted in both modalities are emphasized.

The proposed registration measure becomes

$$I_{new}(A, B) = G(A, B) I(A, B) \quad (4)$$

with

$$G(A, B) = \sum_{(\mathbf{x}, \mathbf{x}') \in (A \cap B)} w(\alpha_{\mathbf{x}, \mathbf{x}'}(\sigma)) \min(|\nabla \mathbf{x}(\sigma)|, |\nabla \mathbf{x}'(\sigma)|). \quad (5)$$

Similarly, the combination of normalized mutual information and gradient information is defined: $Y_{new}(A, B) = G(A, B) Y(A, B)$.

3 Results

We illustrate the behaviour of the known and the newly proposed measures by showing registration functions for two multimodal matching problems: MR to

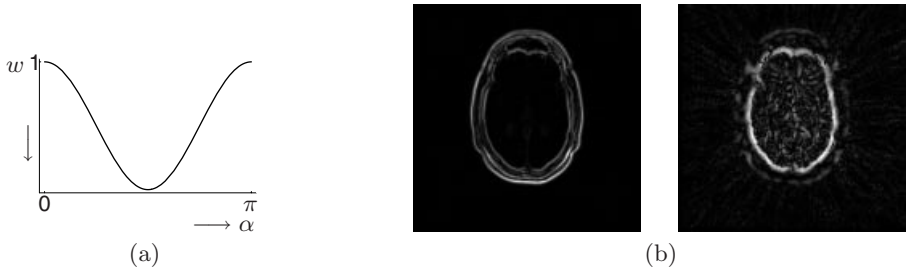


Fig. 1. (a) Weighting function for gradient angles, (b) Examples of the gradient function G per pixel for MR-T1/CT (left) and MR-T1/PET (right).

CT and PET to MR. We make use of the datasets from the aforementioned RREP study, which consist of pairs of either CT or PET and MR (PD, T1 and T2) images, seven pairs each. For most MR images, a *rectified* version is also available, corrected for scaling and intensity inhomogeneity. Following the examples of registration functions, we evaluate the accuracy of the new measures by comparing the registration results against a screw marker based gold standard. Finally, we give an indication of the behaviour of the registration functions in six dimensions, by traversing the search space from different starting positions and comparing the number of local maxima and the errors encountered for the different measures.

In the computation of the gradients, our choice of scale was motivated by past research on edge based measures for image registration [12], which demonstrated the best performance of edge based measures at smaller scales. Searching a trade-off between small scale and image resolution, we have opted for a σ of 1.5 mm for all images.

3.1 Registration Functions

MR-T1 and CT

Although MR images depict different anatomical details than CT images, there generally are corresponding structures – and hence corresponding gradients – in both images. Figure 2 shows some examples of MR-CT registration functions, with the zero position corresponding to the gold standard solution. We first show an example of a well-defined mutual information function (rotation around an in-plane axis, top row) and find that the function is not significantly altered by the inclusion of gradient information. As mutual information is sensitive to the number of samples used, the registration function is generally less smooth for images of lower resolution, for example, images that have been downsampled for use in a multiresolution registration method. Indeed, when subsampling the images by a factor of three (middle row), both standard and normalized mutual information functions deteriorate, while the functions for the combined measures are virtually unchanged.

In the bottom row of figure 2 the behaviour of the measures for translation in the slice direction is shown. The local minima in the mutual information and normalized mutual information functions are a result of interpolation. The interpolation method used (linear interpolation) influences the entropy measures by blurring noise and other small structures. Fluctuations in the registration function occur as a result of grid-aligning transformations, for which interpolation is not applied. Here the images have equal slice thicknesses and the local minima correspond to transformations that align the image slices (see [14] for a more detailed explanation).

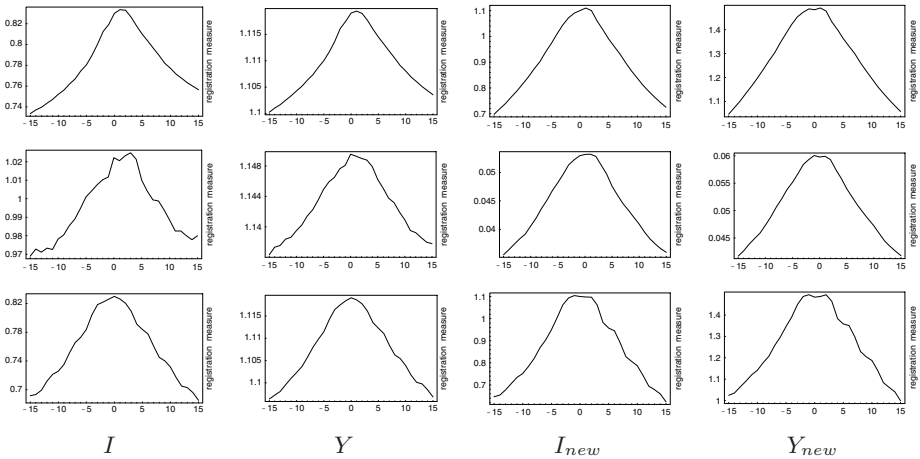


Fig. 2. Registration functions for MR-T1 and CT matching. From top to bottom: (i) rotation around an in-plane axis, (ii) *idem*, images subsampled by a factor of three in each dimension and (iii) translation in slice direction.

MR-T1 and PET

Registration of MR and PET images is more likely to end in misregistration, both because of the fewer similarities between the image contents and because of the lower intrinsic resolution of PET images. The RREP PET images have a relatively low sampling resolution in the slice direction. As a result, the registration functions are particularly ill-defined for the out-of-plane rotations, as can be seen in the top two rows of figure 3. Since it is (at least partly) an overlap problem, the normalized mutual information measure performs rather better. However, for both standard and normalized mutual information, it is obvious that optimization of such functions will not be robust. By including gradient information the registration functions for out-of-plane rotation are vastly improved (rightmost two columns).

In the third row (in-plane rotation), the position of the global optimum for the combined measures is closer to the gold standard solution compared to the global

optimum of standard mutual information. The registration functions for in-plane translations (rows four and five) are well-defined for all measures. Interpolation-induced local minima are found in the registration functions for translation in the slice direction (bottom row). The inclusion of gradient information reduces the artefacts, as can be seen in the rightmost two functions.

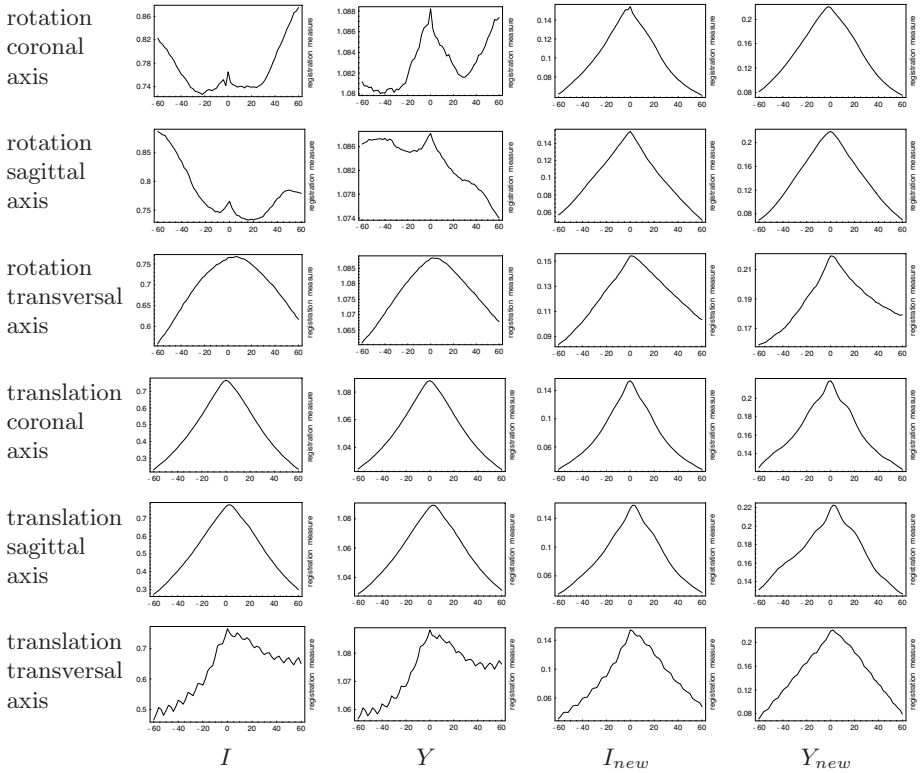


Fig. 3. Registration functions for PET to MR-T1 matching.

3.2 Accuracy

The accuracy of the proposed measures has been tested by comparison of the registration results against a screw marker based gold standard. Each pair of either CT or PET and MR (PD, T1 and T2, both rectified and nonrectified) from the RREP study was registered using the four measures discussed in this paper. The number of datasets for registration of CT to MR nonrectified and rectified was 21 and 20, respectively. For PET and MR, the numbers were 21 and 14. Screw marker based solutions were available for these registration cases. We define the accuracy measure as the averaged Cartesian distance between the

positions of the eight corner points of the image volumes with our solution and with the gold standard solution. The measure is in millimetres. To minimize the dependence of the results on the optimization method (Powell’s direction set method [13]), the starting position of all registration experiments was the gold standard transformation. Table 1 summarizes the registration results for all four measures. The average accuracy, standard deviation and number of mismatches are given. To avoid distortion of the results by outliers, image pairs that resulted in misregistration with any of the measures were excluded for *all* measures for computation of the results. This was the case for two MR and CT image pairs and two MR and PET image pairs.

From the results in table 1 can be concluded that the combined measures do not compromise the accuracy of mutual information based methods. Moreover, optimization of the combined measures did not result in any misregistrations.

Table 1. Registration results (in mm)

		nonrectified MR (n=21/21)			rectified MR (n=20/14)		
		average	std dev	failures	average	std dev	failures
MR/	I	2.4451	1.1737	1	1.3009	0.5254	1
CT	Y	2.1876	0.9485	0	1.1316	0.5410	0
	I_{new}	2.0410	0.8599	0	1.4318	0.4313	0
	Y_{new}	2.0737	0.7109	0	1.8048	0.7616	0
PET/	I	3.9304	1.8102	2	3.8436	1.8426	0
MR	Y	3.6407	2.0821	1	3.2228	1.4035	0
	I_{new}	3.6917	1.4735	0	2.6733	1.6286	0
	Y_{new}	3.5443	1.4355	0	3.2845	2.0767	0

3.3 Robustness

To illustrate the behaviour of the various registration functions in six dimensions, we have investigated the search space of a PET and an MR image. A gradient ascent optimization was started from 64 different positions: the corner points of a 6D hypercube with sides of 10 millimetres in length. The number of different maxima found was counted, where two end positions were considered identical maxima when the difference between each of the six transformation parameters was not more than the step size of the gradient ascent method (here: 1 mm or degree). The registration error of a maximum was identical to the measure defined in the previous section. Figure 4(a) shows the number of maxima found for each of the four registration measures, while (b) shows the mean (horizontal line) and standard deviation (vertical line) of the registration errors of all maxima. The number of maxima is considerably lower for normalized mutual information compared to the standard measure, while it decreases even further when applying the combined measures. Similar behaviour is found for the mean and standard deviations of the registration errors.

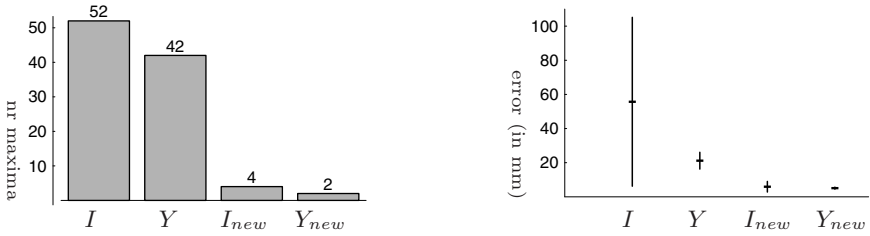


Fig. 4. Left: number of maxima; right: mean (horizontal line) and standard deviation (vertical line).

4 Discussion

We have proposed the adaptation of mutual information measures, by incorporating spatial information. The measure combines either standard or normalized mutual information with gradient information. The essence of the gradient information is that, at registration, locations with a large gradient magnitude should be aligned, but also that the orientation of the gradients at those locations should be similar.

The results presented in this study indicate that the combined measures yield registration functions outperforming both the standard mutual information function with respect to smoothness and attraction basin as well as a normalized mutual information measure. The functions of the combined measures are better defined, containing fewer erroneous maxima and leading to the global maximum from larger initial misregistrations. The measures perform better for low resolution images and can decrease interpolation induced local minima. Well-defined mutual information registration functions are not significantly altered by the inclusion of gradient information.

The accuracy of the combined measures was shown to be similar to that of standard and normalized mutual information. Image pairs that yielded mismatches with (normalized) mutual information, were accurately registered with either of the combined measures. The robustness of standard and normalized mutual information was studied by Studholme et al. [7], showing poor robustness for mutual information and good performance for normalized mutual information. However, we *have* encountered a mismatch and ill-defined registration functions for normalized mutual information (only for the nonrectified MR images, which were not included in Studholme's study). Experiments into the robustness of the measures with respect to starting position implied that including spatial information considerably improves robustness.

Several issues of the method can be improved upon or should be investigated further, including the dependence of the method on the scaling parameter in the gradient computation and the matter of differences in intrinsic resolution. PET images have a significantly lower intrinsic resolution than MR images and

it is possible the method can be improved upon by taking this difference into account.

5 Acknowledgments

The images were provided as part of the project, “Evaluation of Retrospective Image Registration”, National Institutes of Health, Project Number 1 R01 NS33926-01, Principal Investigator Prof. J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN. We thank the Laboratory for Medical Imaging Research in Leuven (especially Dr Frederik Maes) for kindly supplying us with their software for mutual information based registration.

This research was funded by the Netherlands Organization for Scientific Research (NWO).

References

- [1] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.
- [2] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.
- [3] C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, and R. L. Wahl. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis*, 1(3):195–206, 1997.
- [4] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical Physics*, 24(1):25–35, 1997.
- [5] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.
- [6] J. West *et al.* Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21(4):554–566, 1997.
- [7] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.
- [8] P. Thévenaz and M. Unser. Spline pyramids for inter-modal image registration using mutual information. In A. Aldroubi, A. F. Laine, and M. A. Unser, editors, *Wavelet Applications in Signal and Image Processing*, volume 3169 of *Proc. SPIE*, pages 236–247. SPIE Press, Bellingham, WA, 1997.
- [9] C. E. Rodríguez-Carranza and M. H. Loew. A weighted and deterministic entropy measure for image registration using mutual information. In K. M. Hanson, editor, *Medical Imaging: Image Processing*, volume 3338 of *Proc. SPIE*, pages 155–166. SPIE Press, Bellingham, WA, 1998.

- [10] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes. A comparison of similarity measures for use in 2D-3D medical image registration. *IEEE Transactions on Medical Imaging*, 17(4):586–595, 1999.
- [11] D. Rueckert, M. J. Clarkson, D. L. G. Hill, and D. J. Hawkes. Non-rigid registration using higher-order mutual information. In *Medical Imaging: Image Processing*, Proc. SPIE. SPIE Press, Bellingham, WA, 2000. In press.
- [12] J. B. A. Maintz, P. A. van den Elsen, and M. A. Viergever. Comparison of edge-based and ridge-based registration of CT and MR brain images. *Medical Image Analysis*, 1(2):151–161, 1996.
- [13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical recipes in C*. Cambridge University Press, Cambridge, UK, 1992.
- [14] 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.