Rapid Communications

Image Registration by Maximization of Combined Mutual Information and Gradient Information

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

***Abstract*—Mutual information has developed into an accurate measure for rigid and affine monomodality and multimodality image registration. The robustness of the measure is questionable, however. A possible reason for this is 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 as a normalized measure are presented for rigid registration of three-dimensional clinical images [magnetic resonance (MR), computed tomography (CT), and positron emission tomography (PET)]. The results indicate that the combined measures yield a better registration function does mutual information or normalized mutual information *per se*. The registration functions are less sensitive to low sampling resolution, do not contain incorrect global maxima that are sometimes found in the mutual information function, and interpolation-induced local minima can be reduced. These characteristics yield the promise of more robust registration measures. The accuracy of the combined measures is similar to that of mutual information-based methods.

***Index Terms*—Image gradients, image registration, image structure, multimodal images, mutual information.**

I. Introduction

Of the multitude of image registration measures that have been proposed over the years (see [1] for an extensive survey), mutual information is currently one of the most intensively researched measures. This attention is a logical consequence of both the favorable characteristics of the measure and the good registration results reported. Mutual information is an automatic, intensity-based measure, which does not require the definition of landmarks or features such as surfaces and that can be applied in retrospect. Furthermore, it is one of the few intensity-based measures that is well suited to registration of multimodal images. Unlike measures based on correlation of gray values or differences of gray values, mutual information does not assume a linear relationship among the gray values in the images.

Several independent studies have shown the suitability of mutual information as a registration measure for multimodal medical images [2]–[6]. Perhaps the best illustration of the performance of mutual information can be found in the Retrospective Registration Evaluation Project (RREP), an international study comparing the accuracy of 16 registration methods against a screw marker gold standard [7]. Registration of computed tomography (CT) and positron emission tomography (PET) brain images to magnetic resonance (MR) images was studied, with the experiments for each method being performed by the research group proposing that particular method. From the study, it can be concluded that the mutual information-based methods in the study were among the most accurate methods, achieving an accuracy approaching that of the screw marker-based gold standard.

Manuscript received January 18, 2000; revised June 17, 2000. This research was supported by the Netherlands Organization for Scientific Research. The Associate Editor responsible for the review of this paper and recommending its publication was M. Vannier. *Asterisk indicates corresponding author*.

*J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever are with the Image Sciences Institute, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands (e-mail: josien@isi.uu.nl).

Publisher Item Identifier S 0278-0062(00)07993-3.

However, mutual information is not a panacean method. Despite the general promising results, mutual information-based matching can result in misregistration [8]–[11]. The mutual information registration function can be ill-defined, containing local maxima. This can occur, for example, when the images are of low resolution, when the images contain little information, when there is only a small region of overlap or as a result of interpolation methods. Because it is evident that mutual information *can* register images well in general, but still fails occasionally, research into improving the method is ongoing. Improvements of different aspects of the method have been suggested, such as multiresolution methods [8], a different entropy measure [9], invariance with respect to overlap [11], and "higher-order" mutual information, using co-occurrence matrices of neighboring voxels' intensities [12].

In this paper, we propose an adaptation of the mutual information measure to include spatial information that is contained in each of the images separately. Mutual information does not contain this spatial information (except in the interpolation of gray values). Setting aside interpolation, a random reshuffling of the image voxels (identical for both images) yields the same mutual information value as for the original images. We combine mutual information with a gradient measure to provide spatial information. Image gradients by themselves have been shown to be useful registration criteria [10], [13].

II. Method

A. Mutual Information

The mutual information of two images is a combination of the entropy values of the images, both separately and jointly. One interpretation of entropy is as a measure of dispersion of a probability distribution. A distribution with only a few large probabilities has a low entropy value; the maximum entropy value is reached for a uniform distribution. The entropy of an image can be computed by estimating the 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 gray values. The marginal distributions are obtained by summing over the rows, respectively, the columns, of the joint histogram.

The definition of the mutual information I of two images A and B combines the marginal and joint entropies of the images in the following manner:

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

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. Correct registration of the images is assumed to be equivalent to maximization of the mutual information of the images. This implies a balance between minimization of the joint entropy and maximization of the marginal entropies. The joint entropy is minimal when the joint distribution is minimally dispersed, i.e., when it is crisp. This corresponds to registration, because any misalignment of the images will both introduce new combinations of gray values and decrease the probabilities of the "correct" combinations. The overall result is a more dispersed joint probability distribution.

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 overcome this problem. Examples of such measures are the normalized mutual information introduced by Studholme *et al.* [11]

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

and the entropy correlation coefficient used by Maes *et al.* [3]

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

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

B. 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 that is present in each of the images. This extension is accomplished 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.

Simply applying mutual information to gradient images would seem a logical solution to incorporating spatial information, were it not that the registration function would probably have a narrow attraction range and that a lot of information from the gray value images is discarded. We therefore propose a combination of mutual information and gradient information.

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

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

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, because 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 favors both very small angles and angles that are approximately equal to π (see Fig. 1):

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

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. Because 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. Multiplication was preferred over addition, because addition of the terms would require both terms to be normalized. Some examples of the gradient measure (before summation) for different combinations of multimodal images can be found in Fig. 2. Tissue transitions that are depicted in both modalities are emphasized.

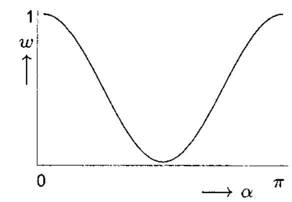

Fig. 1. Weighting function for gradient angles.

Fig. 2. Examples of the individual contributions of voxels to the gradient function G for, respectively, MR-T1 and MR-T2, MR-T1 and CT, and MR-T1 and PET. The slices shown are extracted from three-dimensional images. The scalar G, with which mutual information is multiplied, is found by summing over all voxels.

The proposed registration measure becomes

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

with

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

Similarly, the combination of normalized mutual information and gradient information is defined as

$$Y_{new}(A, B) = G(A, B)Y(A, B).$$

C. Optimization

Optimization of the registration function is done using Powell's method [14]. This method repeatedly iterates the dimensions of the search space, performing one-dimensional optimizations for each dimension, until convergence is reached.

Because we rigidly register image volumes, the search space is six-dimensional, i.e., three rotations and three translations.

III. Results

To show the performance of our proposed measure, we consider a variety of registration problems: MR to MR, MR to CT, and MR to PET. The images involved are those used in the aforementioned RREP comparison study. This set consists of pairs of CT and MR (PD, T1, and T2) images of seven patients and pairs of PET and MR (PD, T1, and T2) of seven patients. Also included are *rectified* MR sets, corrected for scaling and intensity inhomogeneity.

In Section III-A, we will highlight some typical registration functions for different registration problems, comparing the behavior for mutual information, normalized mutual information, and the combination measures. In Section III-B, we evaluate the accuracy of the com-

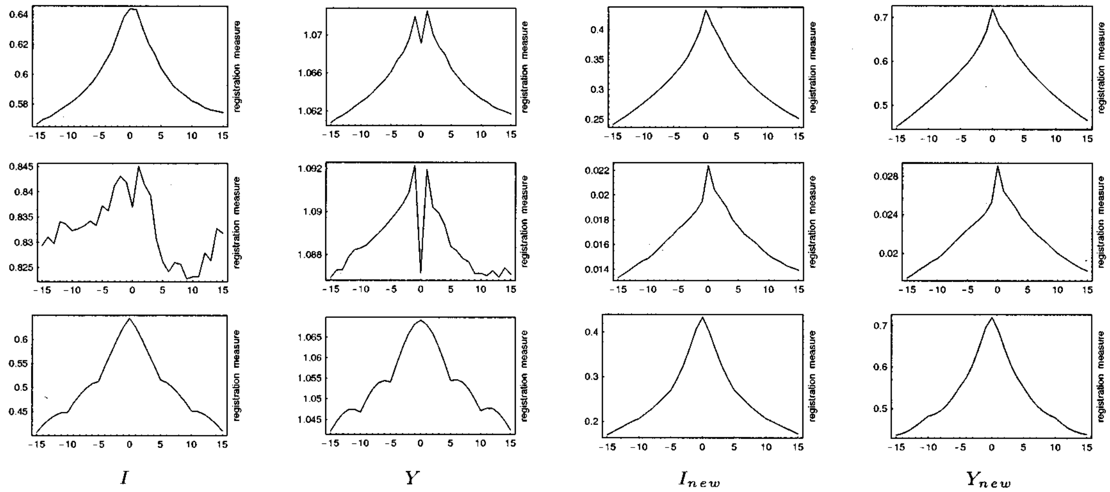

Fig. 3. Registration functions for MR-T1 and T2 matching. From top to bottom: (i) rotation around an in-plane axis, (ii) *idem*, images subsampled by a factor of three in each dimension, and (iii) translation along an in-plane axis.

bination measures for registration of MR/CT and MR/PET image pairs by comparison to bone marker-based solutions.

In the computation of the gradients, our choice of scale was motivated by past research on edge-based measures for image registration [13], which demonstrated the best performance of edge-based measures at smaller scales. Searching for a tradeoff between small scale and image resolution, we have opted for a σ of 1.5 mm for all images.

A. Registration Functions

1) MR-T1 and MR-T2: Registration of MR-T1 and MR-T2 weighted images is probably the best test case for a gradient-based method. Although the images are multimodal in the sense that different tissue characteristics are imaged, the structure in the images is similar and we would expect to find gradients at corresponding locations. Fig. 3 contains registration functions for matching of an MR-T1 and an MR-T2 image, using mutual information (I), normalized mutual information (Y), and both proposed combination measures (I_{new} and Y_{new}). We assume no transformation is required to align the images. The top row shows the functions for rotation around an in-plane axis. Clearly, the original mutual information function can hardly be improved on, and the functions for the combined measures are nearly identical. The normalized mutual information function has a local minimum at the position of correct alignment, which is probably a result of interpolation [15].

As mutual information is based on estimating probability distributions, the registration function is generally less smooth when the number of samples is small, as, for example, in multiresolution methods. In the middle row of Fig. 3, we show the registration functions for rotation around an in-plane axis after the images have been equidistantly subsampled by a factor of three in each dimension. The smoothness of the mutual information function has significantly decreased; worse yet, the optimum has moved. In the normalized mutual information function, the local minimum has become more pronounced and the function is less smooth. In contrast, the registration functions for the combined measures remain smooth and the optimum did not shift. Although the smoothness of the mutual information registration functions can be improved on to some extent by more advanced downscaling methods, these examples illustrate that the combined measures are less sensitive to the number of samples.

In the bottom row of Fig. 3, the behavior of the measures for translation in the in-plane 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. Noise increases the dispersion of a probability distribution and, hence, its entropy. The local minima in the function correspond to translations that align the image grids (the images have equal voxel sizes). For such translations, interpolation is not applied, resulting in a higher joint entropy because of the presence of noise and a decreased mutual information value (see [15] for a more detailed explanation). In the registration functions of the combined measures, the artifacts are nearly eliminated.

2) 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. Fig. 4 shows some examples of MR-CT registration functions, around the marker-based gold standard solution. Again, 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. When subsampling the images by a factor of three (middle row), the mutual information function deteriorates rapidly, the normalized mutual information function is less affected, though it also loses smoothness, while the functions for the combined measures are virtually unchanged. The final example plots the measures as a function of translation in the slice direction. The images have an equal slice thickness, which explains the occurrence of artifacts.

3) MR-T1 and PET: Registration of MR and PET images is a considerably more difficult problem than the previous two, both because of the fewer similarities between the image contents and because of the lower intrinsic resolution of PET images. Experience has taught that mutual information-based matching of these images can often result in misregistration. Problems occur usually not because of a lack of smoothness, but because the function is ill-defined, having prominent maxima away from the true optimum. Examples of ill-defined mutual information functions can be found in Fig. 5, top two rows. The figure shows registration functions of an MR-T1 and a PET image for which registration with both mutual information and normalized mutual information resulted in mismatches. The zero position in the functions presented corresponds to the marker-based solution. From top to bottom, registration functions are given for each of the six rigid transformation parameters.

The cause of the misregistrations is clearly visibly in the top two rows: for both out-of-plane rotations, the standard and normalized

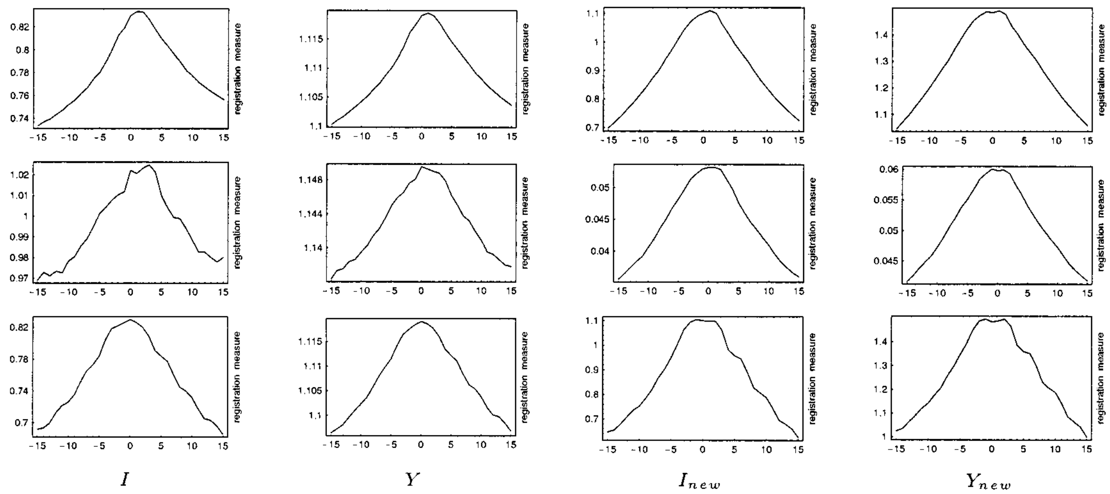

Fig. 4. Registration functions for MR-T1 and CT matching. From top to bottom: (i) rotation around an in-plane axis, (ii) *idem*, images subsampled by a factor of three in each dimension, and (iii) translation in slice direction.

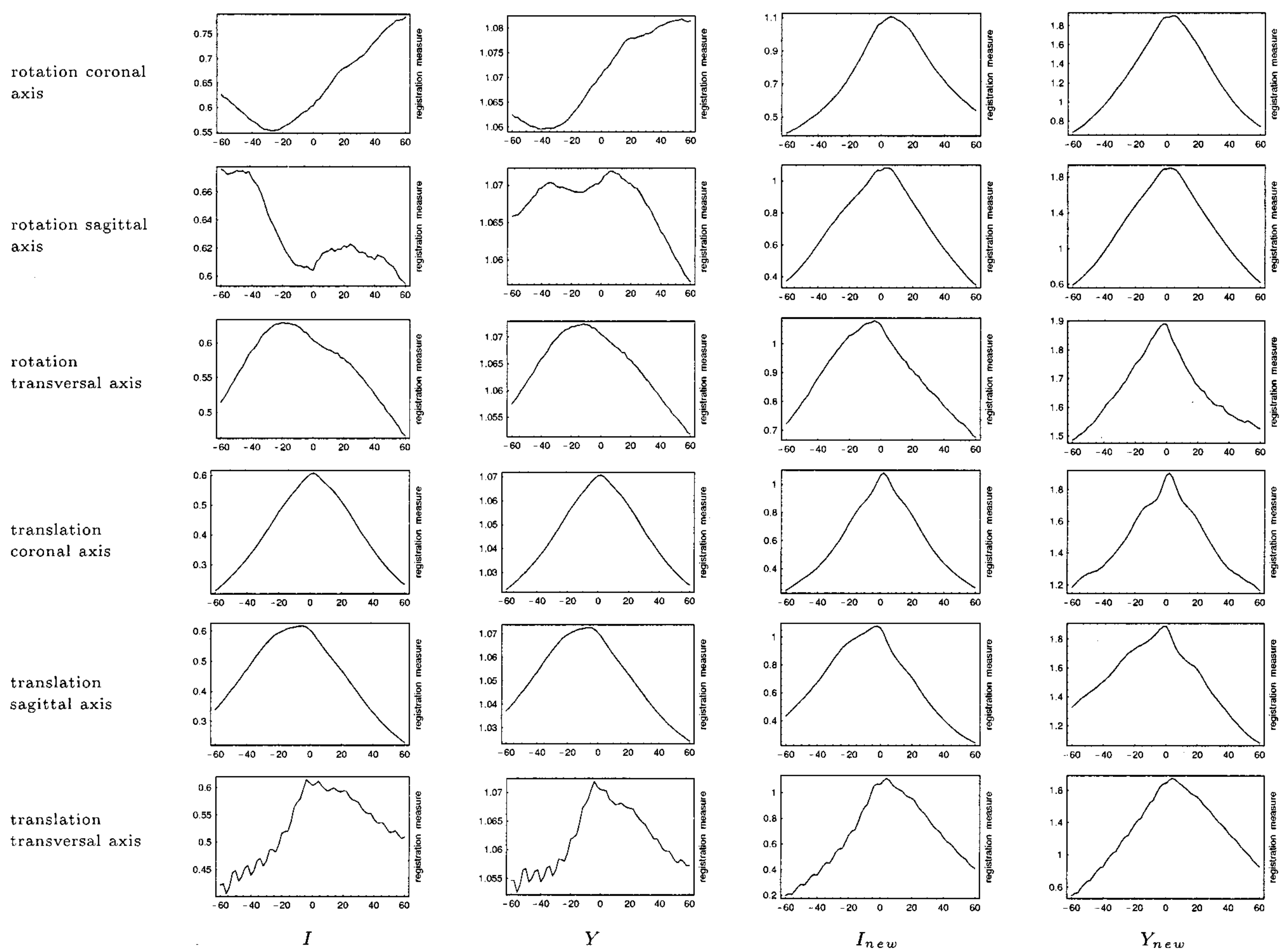

Fig. 5. Registration functions for MR-T1 and PET matching.

mutual information functions are ill defined (a result of large slice thickness and a small number of slices). By attempting several starting positions, a reasonable match could be found for these images, because the functions contain small local maxima in the vicinity of the true optimum. However, 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). The functions now show a global optimum close to the marker-based solution.

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 optima of standard and normalized mutual

TABLE I
REGISTRATION RESULTS FOR MRI AND CT IMAGE PAIRS (IN mm)

	median				maximum				
	I	Y	I_{new}	Y_{new}	I	Y	I_{new}	Y_{new}	n
T1	1.07	0.84	1.21	1.37	2.23	2.08	2.23	2.91	7
PD	1.42	1.43	1.53	1.73	3.20	3.95	2.90	3.08	7
T2	1.48	1.51	1.29	1.63	8.79	3.02	3.16	3.01	7
T1rect	0.71	0.61	0.78	0.80	1.69	1.00	1.62	2.15	6
PDrect	0.68	0.71	0.69	0.87	1.49	1.19	1.67	2.40	7
T2rect	0.72	0.63	0.89	1.04	3.54	2.20	1.83	2.99	7

TABLE II
REGISTRATION RESULTS FOR MRI AND PET IMAGE PAIRS (IN mm)

	median				maximum				
	I	Y	I_{new}	Y_{new}	I	Y	I_{new}	Y_{new}	n
T1	2.37	1.90	2.46	2.43	38.91	74.74	6.65	8.78	7
PD	2.33	2.24	3.04	2.67	5.13	6.55	7.55	6.82	7
T2	2.51	2.40	3.13	2.27	7.24	5.38	6.05	7.91	7
T1rect	1.82	1.34	2.12	1.29	3.51	3.56	3.62	4.29	4
PDrect	2.72	2.88	2.07	2.97	6.00	3.78	6.29	6.44	5
T2rect	2.74	2.67	2.24	1.62	6.95	5.25	3.27	5.42	5

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), since the slice thicknesses of the images have a common factor. The inclusion of gradient information reduces the artefacts, as can be seen in the rightmost two functions.

B. Accuracy

The accuracy of the proposed measures has been tested by comparison of the registration results against a 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. Bone marker-based solutions were available for these registration cases. Registration errors between a transformation and the gold standard were computed at the centers of several volumes of interest in the data sets, as in the RREP study (see [7] for more detail).

To minimize the dependence of the results on the optimization method, the starting position of all registration experiments was the marker-based gold standard transformation. Tables I and II summarize the registration results for all four measures. The median error, maximum error, and number of data sets are given.[1]

Statistical testing of the median errors (two-tailed paired Student t test, $p = 0.05$) found no significant difference in the results of the mutual information measures versus the corresponding combination measures. Neither when the significance tests were performed on all data sets, nor when the results were divided into four categories (by CT/PET, rectification/no rectification).

Registration using standard mutual information resulted in two misregistrations, a case of CT-T2 and a case of PET-T1, where a result was considered a mismatch when the average error over the volumes of interest was larger than the largest voxel dimension (4 mm for CT and 8 mm for PET). Normalized mutual information failed once (PET-T1). The corresponding maximum errors are an indication of the extent of the misregistration. The combined measures yielded satisfactory results in all cases.

IV. 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 decrease interpolation-induced local minima.

In cases in which standard mutual information performs well, the registration functions of the combined measures are similar and the global optimum does not alter significantly.

The accuracy of the combined measures was shown to be similar to that of standard and normalized mutual information. Registration using the combined measures is likely to be more robust because of the better defined registration functions. No distinct differences were found between combined measures based on standard or normalized mutual information.

Several issues of the method can be improved on or should be investigated further. The robustness of the combined measures should be researched more extensively. The robustness of standard and normalized mutual information was studied by Studholme *et al.* [11], 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). Another important topic is the dependence of the method on the scaling parameter in the gradient computation. Directly linked is 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 on by taking this difference into account.

[1]Please note that these errors are not directly comparable to those of other methods, because the gold standard transformation was used as the starting position.

ACKNOWLEDGMENT

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 Dr. J. M. Fitzpatrick, Vanderbilt University, Nashville, TN. The authors are indebted to Dr. J. West for his assistance. They also thank the Laboratory for Medical Imaging Research in Leuven (especially, Dr. F. Maes) for kindly supplying them with their software for mutual information-based registration.

REFERENCES

[1] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.*, vol. 2, no. 1, pp. 1–36, 1998.
[2] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–51, 1996.
[3] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, 1997.
[4] 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," *Med. Image Anal.*, vol. 1, no. 3, pp. 195–206, 1997.
[5] 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," *Med. Phys.*, vol. 24, no. 1, pp. 25–35, 1997.

[6] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *Int. J. Comput. Vision*, vol. 24, no. 2, pp. 137–154, 1997.
[7] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, Jr., R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J . Hawkes, C. Studholme, J. B. A. Maintz, M. A. Viergever, G. Malandain, X. Pennec, M. E. Noz, G. Q. Maguire, Jr., M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *J. Comput. Assisted Tomogr.*, vol. 21, no. 4, pp. 554–556, 1997.
[8] P. Thévenaz and M. Unser, "Spline pyramids for inter-modal image registration using mutual information," in *Wavelet Applications in Signal and Image Processing*, A. Aldroubi, A. F. Laine, and M. A. Unser, Eds. Bellingham, WA: SPIE Press, 1997.
[9] C. E. Rodríguez-Carranza and M. H. Loew, "A weighted and deterministic entropy measure for image registration using mutual information," in *Medical Imaging: Image Processing*, K. M. Hanson, Ed. Bellingham, WA: SPIE Press, 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 Trans. Med. Imag.*, vol. 17, no. 4, pp. 586–595, 1999.
[11] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, 1999.
[12] 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*, K. M. Hanson, Ed. Bellingham, WA: SPIE Press, 2000.
[13] 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," *Med. Image Anal.*, vol. 1, no. 2, pp. 151–161, 1996.
[14] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*. Cambridge, U.K.: Cambridge Univ. Press, 1992.
[15] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Interpolation artefacts in mutual information-based image registration," *Comput. Vision Image Understanding*, vol. 77, no. 2, pp. 211–232, 2000.