f-Information Measures in Medical Image Registration

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

***Abstract*—A measure for registration of medical images that currently draws much attention is mutual information. The measure originates from information theory, but has been shown to be successful for image registration as well. Information theory, however, offers many more measures that may be suitable for image registration. These all measure the divergence of the joint distribution of the images' grey values from the joint distribution that would have been found had the images been completely independent. This paper compares the performance of mutual information as a registration measure with that of other f-information measures.**

The measures are applied to rigid registration of positron emission tomography(PET)/magnetic resonance (MR) and MR/computed tomography (CT) images, for 35 and 41 image pairs, respectively. An accurate gold standard transformation is available for the images, based on implanted markers. The registration performance, robustness and accuracy of the measures are studied.

Some of the measures are shown to perform poorly on all aspects. The majority of measures produces results similar to those of mutual information. An important finding, however, is that several measures, although slightly more difficult to optimize, can potentially yield significantly more accurate results than mutual information.

***Index Terms*—f-information, multimodal image registration, mutual information.**

I. Introduction

In the mid 1990s, mutual information made its entrance into the field of medical image registration [1], [2]. Since then it has been adopted by a large number of researchers for a large number of applications [3]–[12]. Because registration based on mutual information is successful for a variety of image modalities, can be fully automated and yields good results, interest in the measure grew rapidly. Research into the measure currently takes up a substantial part of medical image registration research. Apart from mutual information, information theory comprises many more information measures, which could be considered for image registration. An example is I_α-information (explained in Section II), which defines a class of measures of order α. Mutual information is actually a member of this class; it is the I_α measure of order 1. The question logically arises whether I_1 is the optimal registration measure of the class of I_α measures and what the influence of the order α is on the registration results.

Information theory measures for image registration other than joint entropy and mutual information have received little attention. Some of the results in this paper have been presented in [13]. Zhu [14] has studied the performance of cross entropy for multimodal registration. Wachowiak *et al.* [15] compare information measures based on Rényi and Havrda-Charvát entropies and I_α-information. Exclusive f-informations are registration measures introduced by Rougon *et al.* [16]. Finally, both He *et al.* [17] and Bardera *et al.* [18] employ Jensen divergence for image registration.

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

$$I(A,B) = \sum_{a,b} p(a,b) \log \frac{p(a,b)}{p(a)p(b)}. \tag{1}$$

Although usually defined for two images, as we have done here, mutual information is actually computed on probability distributions, viz. the marginal distributions $[p(a), p(b)]$ and joint distribution $[p(a,b)]$ of grey value pairs (a,b) of corresponding image grey values. Mutual information measures the dependence of the images by determining the distance of their joint distribution $p(a,b)$ to the joint distribution in case of complete independence, $p(a)p(b)$. Maximal dependence is assumed to occur when the images are aligned. Registration is, therefore, achieved by finding the geometrical transformation that yields the highest mutual information value.

Measures of the distance between a joint probability distribution and the product of the marginal distributions are *information* measures. Information measures constitute a subclass of the *divergence* measures, which are measures of the distance between two arbitrary distributions. A specific class of information (divergence) measures, of which mutual information is a member, is formed by the f-information (f-divergence) measures. In this paper we compare mutual information with several other f-information measures by applying them to the registration of clinical magnetic resonance (MR), positron emission tomography (PET), and computed tomography (CT) images.

II. f-Divergence and f-Information of Probability Distributions

The extent to which two probability distributions differ can be expressed by a so-called *measure of divergence*. Such a measure will reach a minimum value when the two probability distributions are identical and the value increases with increasing

Manuscript received March 3, 2004; revised August 13, 2004. This work was supported by the Netherlands Organization for Scientific Research (NWO). The Associate Editor responsible for coordinating the review of this paper was M. Unser. *Asterisk indicates corresponding author.*

*J. P. W. Pluim is with the Image Sciences Institute, University Medical Center Utrecht, Room E01.335, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands (e-mail: josien@isi.uu.nl).

M. A. Viergever is with the Image Sciences Institute, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands (e-mail: max@isi.uu.nl).

J. B. A. Maintz is with Utrecht University, Institute of Information and Computing Sciences, 3508 TB Utrecht, The Netherlands (e-mail: twan@cs.uu.nl).

Digital Object Identifier 10.1109/TMI.2004.836872

disparity between the two distributions. A specific class of divergence measures is the set of f-*divergence* measures [19]. For two discrete probability distributions $P = \{p_i | i = 1, \ldots, n\}$ and $Q = \{q_i | i = 1, \ldots, n\}$, the f-divergence is defined as

$$f(P\|Q) = \sum_i q_i f\left(\frac{p_i}{q_i}\right). \tag{2}$$

The demands on the function f are that 1) $f : [0, \infty) \mapsto (-\infty, \infty]$, 2) f is continuous and convex on $[0, \infty)$, 3) finite on $(0, \infty)$, and 4) strictly convex at some point $x \in (0, \infty)$.

The following definition completes the definition of f-divergence for the two cases for which (2) is not defined:

$$q_i f\left(\frac{p_i}{q_i}\right) = \begin{cases} 0, & \text{if } p_i = q_i = 0 \\ p_i \lim_{x \uparrow \infty} \frac{f(x)}{x}, & \text{if } p_i > 0, q_i = 0 \end{cases}. \tag{3}$$

An example of an f-divergence is the I_α-divergence [19], which is formed by substituting the following function $I_\alpha(x)$:

$$I_\alpha(x) = \frac{x^\alpha - \alpha x + \alpha - 1}{\alpha(\alpha - 1)}, \quad \alpha \neq 0, \quad \alpha \neq 1$$

for f in the definition of f-divergence (2). The resulting divergence measure reads

$$\begin{aligned} I_\alpha(P\|Q) &= \frac{1}{\alpha(\alpha-1)} \sum_i \left(\frac{p_i^\alpha}{q_i^{\alpha-1}} - \alpha p_i + \alpha q_i - q_i \right) \\ &= \frac{1}{\alpha(\alpha-1)} \left(\sum_i \left(\frac{p_i^\alpha}{q_i^{\alpha-1}} \right) - 1 \right) \end{aligned}$$

for $\alpha \neq 0$, $\alpha \neq 1$ and using the completeness of the distributions, i.e., $\sum_i p_i = \sum_i q_i = 1$.

Taking the limit for $\alpha \to 1$, one finds that $I_1(P\|Q) = \sum_i p_i \log(p_i/q_i)$, which is known as the Kullback–Leibler distance [20] or relative entropy [21] (see the Appendix).

A special case of f-divergence are the f-*information* measures. These are defined similarly to f-divergence measures, but apply only to specific probability distributions; namely, the joint probability P of two variables and their marginal probabilities' product $Q = P_1 \times P_2$. f-Information is a measure of *dependence*: it measures the distance between a given joint probability p_{ij} and the joint probability when the variables are independent $(p_i p_j)$.

Using the same example as before, the I_α-*information* is defined as

$$I_\alpha(P\|P_1 \times P_2) = \frac{1}{\alpha(\alpha-1)} \left(\sum_{i,j} \frac{p_{ij}^\alpha}{(p_i p_j)^{\alpha-1}} - 1 \right)$$

for $\alpha \neq 0$, $\alpha \neq 1$.

For $\alpha \to 1$, $I_1(P\|P_1 \times P_2) = \sum_{i,j} p_{ij} \log(p_{ij}/p_i p_j)$, which is identical to the definition of mutual information given in (1).

III. Measures of f-Information

Because the list of functions f that can be used to form f-information measures is a long one, we will limit ourselves to measures that are frequently encountered in the literature. The following is an overview of those measures, accompanied by an explanation of our choice of measures for this comparison study.

We assume all probability distributions are complete, i.e., $\sum_i p_i = 1$.

A. V-Information

One of the simplest measures of dependence is obtained using the function $V = |x - 1|$, which results in the V-information [19]

$$V(P\|P_1 \times P_2) = \sum_{i,j} |p_{ij} - p_i p_j|. \tag{4}$$

The absolute distance of two variables is a frequently used measure of similarity. In statistics, the Kolmogorov–Smirnov measure tests the similarity of an observed and a hypothesized distribution using absolute differences. When applied directly to the grey values of images, instead of their probability distributions, the sum of absolute differences is a well-known registration measure for monomodal images (for example, see [22], [23]).

B. I_α-Information

A measure already mentioned is the I_α-information [19]

$$I_\alpha(P\|P_1 \times P_2) = \frac{1}{\alpha(\alpha-1)} \left(\sum_{i,j} \frac{p_{ij}^\alpha}{(p_i p_j)^{\alpha-1}} - 1 \right) \tag{5}$$

for $\alpha \neq 0$, $\alpha \neq 1$.

The class of I_α-information includes mutual information, which equals I_α for the limit $\alpha \to 1$.

C. M_α-Information

Matusita [19] defined the function

$$M_\alpha(x) = |x^\alpha - 1|^{\frac{1}{\alpha}}, \quad 0 < \alpha \leq 1.$$

When applying this function in the definition of an f-information measure, the resulting M_α-information measures are

$$M_\alpha(P\|P_1 \times P_2) = \sum_{i,j} \left| p_{ij}^\alpha - (p_i p_j)^\alpha \right|^{\frac{1}{\alpha}} \tag{6}$$

for $0 < \alpha \leq 1$.

These constitute a generalized version of V-information; M_α-information is identical to V-information for $\alpha = 1$.

D. χ^α-Information

The class of χ^α-information measures is given by Liese [24]. The function f to construct the information measures is

$$\chi^\alpha(x) = \begin{cases} |1 - x^\alpha|^{\frac{1}{\alpha}}, & \text{for } 0 < \alpha \leq 1 \\ |1 - x|^\alpha, & \text{for } \alpha > 1 \end{cases}.$$

For $0 < \alpha \leq 1$, this function equals the M_α function. The χ^α and M_α information measures are, therefore, also identical for $0 < \alpha \leq 1$. For $\alpha > 1$, the χ^α-information reads

$$\chi^\alpha(P\|P_1 \times P_2) = \sum_{i,j} \frac{|p_{ij} - p_i p_j|^\alpha}{(p_i p_j)^{\alpha-1}}. \tag{7}$$

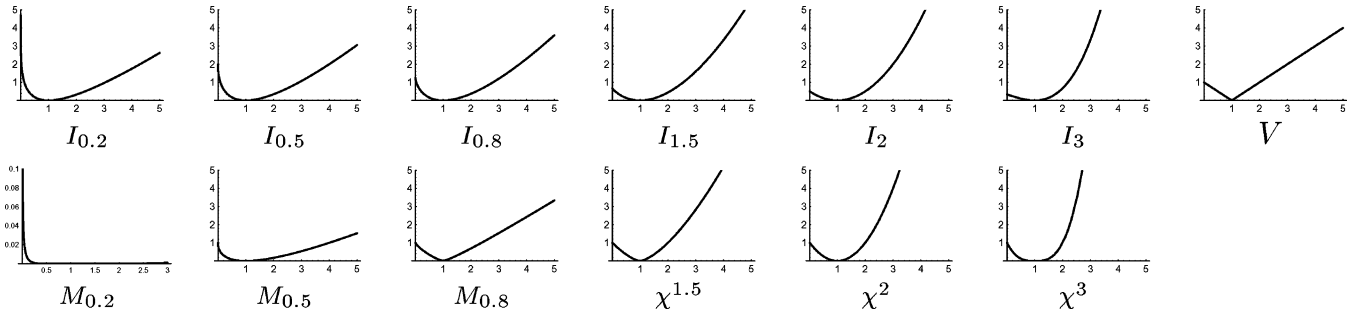

Fig. 1. Examples of different functions f for f-divergence and f-information measures (note the different extent of the axes for $M_{0.2}$).

The χ^2-divergence $\sum_{i,j}((p_i - q_i)^2/q_i)$ is known as the Pearson statistic, which is used in statistical hypothesis testing.

Some examples of the functions f, which can be used to form f-divergence or f-information measures, are given in Fig. 1. These examples illustrate the influence of the parameter α on the behavior of the function f. All functions $f(x)$ have a minimum at $x = 1$.

The following measures do not fall in the class of f-divergence measures, because they do not satisfy the definition of f-divergence as given at the beginning of the previous section. They *are*, however, divergence measures (they measure the distance between two distributions) and they are directly related to I_α-divergence.

The *Hellinger integral* of order α (given in [19] for continuous distributions, hence, the term "integral") is defined by

$$H_\alpha(P\|Q) = 1 + \alpha(\alpha-1)I_\alpha(P\|Q) \tag{8}$$

for $\alpha \neq 0$, $\alpha \neq 1$.

Hellinger proposed $H_{(1/2)}(P\|Q) = \sum_i (p_i q_i)^{(1/2)}$ to measure the distance between probability distributions. Although defined as a divergence measure, it can be transformed to an information measure by replacing P and Q with a joint probability distribution and the product of the marginal distributions, respectively. It is, however, directly related to the I_α-divergence and we have not included it in this comparison study.

E. Rényi Distance

Based on the Hellinger integral is the *Rényi distance* [19]

$$R_{\alpha,1}(P\|Q) = \frac{1}{\alpha(\alpha-1)} \log H_\alpha(P\|Q) \tag{9}$$

for $\alpha \neq 0$, $\alpha \neq 1$.

Rényi himself, on the other hand, also defined a measure of information of order α [25]. It is given by

$$R_{\alpha,2}(P\|Q) = \frac{1}{\alpha-1} \log \sum_i \frac{p_i^\alpha}{q_i^{\alpha-1}} \tag{10}$$

for $\alpha \geq 0$, $\alpha \neq 1$, with $q_i = p(x_i)$ the probability of x_i happening and $p_i = p(x_i|Y)$ the conditional probability of x_i happening given an event Y. At first sight, it may appear to be a measure of divergence, but the use of conditional probabilities renders it a measure of information. The conditional probability in the above definition is for a single event Y. When it depends on a range of events y_j, p_i^α becomes $\sum_j p(y_j)p(x_i|y_j)^\alpha$. Extending (10) to a range of events gives

$$\begin{aligned} R_{\alpha,2}(P\|Q) &= \frac{1}{\alpha-1} \log \sum_i \sum_j \frac{p(y_j)p(x_i|y_j)^\alpha}{p(x_i)^{\alpha-1}} \\ &= \frac{1}{\alpha-1} \log \sum_{i,j} \frac{p(x_i,y_j)^\alpha}{(p(x_i)p(y_j))^{\alpha-1}} \end{aligned}$$

which is a measure of information.[1] It reaches its minimum value when $p(x_i, y_j)$ and $p(x_i)p(y_j)$ are identical, in which case the summation reduces to $\sum_{i,j} p(x_i, y_j)$. Because we assume complete distributions, the sum is 1 and the minimum value of the measure is, therefore, equal to zero. The limit of Rényi's measure for α approaching 1 equals $I_1(P\|Q)$ and, therefore, mutual information. The measure is equal to $(1/(\alpha-1))\log H_\alpha(P\|P_1 \times P_2)$, which means it differs from the Rényi distance (9) only by a factor $1/\alpha$. In this paper, we have chosen to include the measure defined by Rényi himself, $R_{\alpha,2}$, henceforth referred to as R_α.

The following relations between the different measures hold, amongst others:

$$\begin{aligned} M_{\frac{1}{2}}(P\|Q) &= \frac{1}{2} I_{\frac{1}{2}}(P\|Q) = 2\left[1 - H_{\frac{1}{2}}(P\|Q)\right] \\ \chi^2(P\|Q) &= 2I_2(P\|Q) \\ M_1(P\|Q) &= V(P\|Q) \\ \chi^\alpha(P\|Q) &= M_\alpha(P\|Q) \quad 0 < \alpha \leq 1. \end{aligned}$$

IV. Experiments and Results

The comparison of the different information measures is based on rigid registration of clinical PET, CT, and MR (T1, T2, and PD weighted) images. We have used the data from the Retrospective Registration Evaluation Project (RREP), a comparison study of numerous registration methods. The data consists of CT and/or PET images of nine patients, together with MR T1, T2, and PD weighted images. The patients wore stereotactic frames and had markers implanted for the purpose of neurosurgery. All MR images were corrected for geometric and intensity distortion resulting from static field inhomogeneity (by combining the original image and an image acquired with reversed readout gradients [26]) and for scale distortion [27] (using the frame as a reference object of known

[1] In (10), the multiplication with the probability of the event, $p(Y)$, is absent, because the event is given, hence, $p(Y) = 1$.

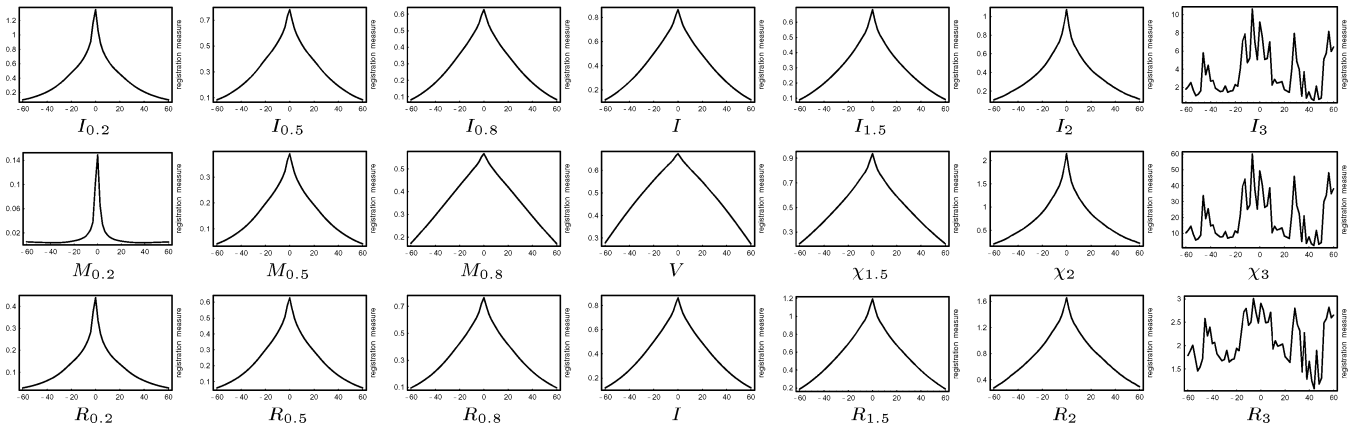

Fig. 2. Examples of the various registration measures as functions of translation along an in-plane axis (in millimeters), for registration of an MR-T1 and a CT image.

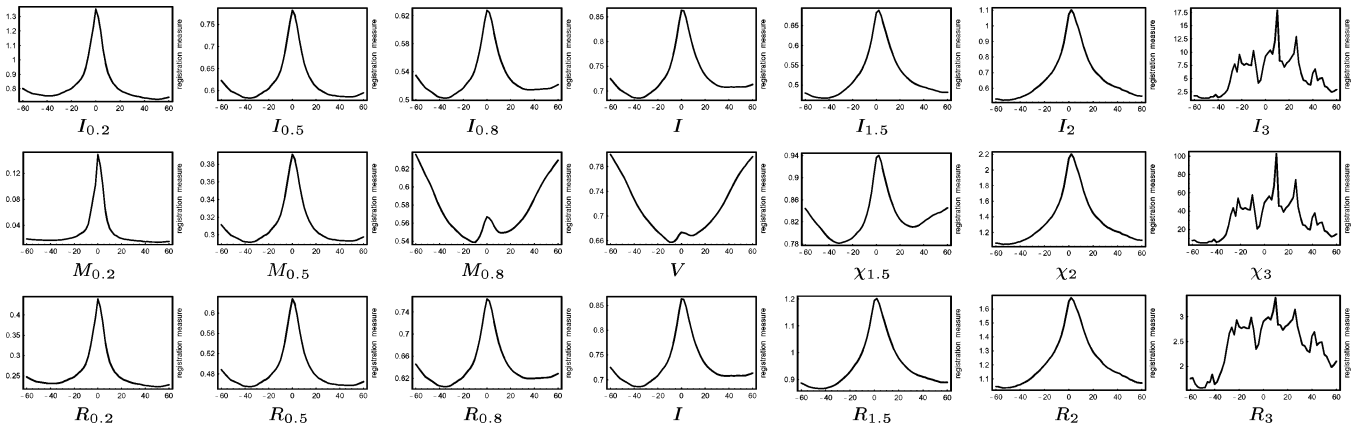

Fig. 3. Examples of the various registration measures as functions of rotation around an in-plane axis (in degrees), for registration of an MR-T1 and a CT image.

size). Evidence of the stereotactic frames and markers was carefully removed from the images in such a manner to avoid any bias, to rule out any influence of these objects on the registration performance. Including both noncorrected and corrected MR images, the dataset contains 41 MR-CT images pairs and 35 MR-PET pairs. All image volumes consist of transversal slices. A CT image generally contains $512 \times 512 \times 30$ voxels of dimensions $0.65 \times 0.65 \times 4\,\mathrm{mm}$, an MR image $256 \times 256 \times 26$ voxels of dimensions $1.25 \times 1.25 \times 4\,\mathrm{mm}$ and a PET image $128 \times 128 \times 15$ voxels of dimensions $2.59 \times 2.59 \times 8\,\mathrm{mm}$. All CT-MR and PET-MR image pairs had been registered using the markers, yielding a gold standard with which we compare the registrations based on information measures. The estimated accuracy of the gold standard is 0.39 mm for CT to MR and 1.65 mm for PET to MR. For more details we refer the reader to the paper on this study [28].

Based on the argumentation given in the previous section, we have chosen to include the following information measures in the study:

I	mutual information;
V	
I_α	for $\alpha \neq 0, \alpha \neq 1$;
M_α	for $0 < \alpha \leq 1$;
χ^α	for $\alpha > 1$;
R_α	for $\alpha \geq 0, \alpha \neq 1$.

These measures were applied to rigid registration of the CT, PET and MR images, which means that the measures were to be optimized for six parameters (three rotations and three translations). The probability distributions of the grey values were estimated from grey value histograms, using 256 bins. Linear interpolation was applied to approximate grey values at nongrid positions. Optimization of the measures was achieved using Powell's method, as described in [29]. The values of α we have investigated were 0.2, 0.5, 0.8, 1.5, 2.0, and 3.0. From visualization of registration functions, as in Figs. 2 and 3 for example, it was clear that values larger than 3.0 were unreasonable. Values close to 1.0 were also excluded, either because the measures resemble mutual information for such values (I_α, R_α) or because they resemble another measure (M_1 and χ^1 equal V). The values 0.5 and 2 may seem an inappropriate choice, because some measures are linearly related for these α ($I_{0.5}$ and $M_{0.5}$, I_2 and χ^2). However, for values of α close to 0.5 and 2, the results will not differ much and for values further from 0.5 and 2, the results will be similar to those for other choices of α. Even though some measures yield identical results for $\alpha = 0.5$ or 2, these values are interesting to explore.

To test statistical significance of the difference in performance of the measures, we have used the nonparametric Wilcoxon matched pairs signed rank sum test. This method is better adapted to possibly skewed data than a method that expects the results to be normally distributed. The null hypoth-

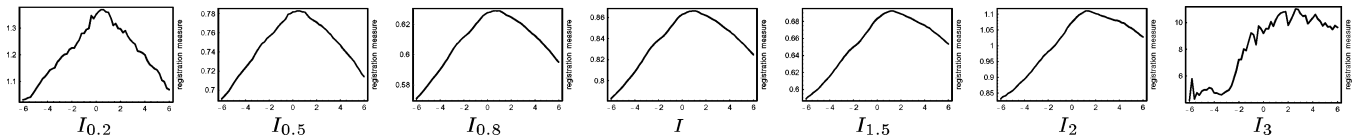

Fig. 4. Examples of the smoothness of the I_α registration function as α varies. The functions show the information measure of an MR-T1 and a CT image, for rotation around an in-plane axis (in degrees).

esis for all tests assumes there is no difference in performance between mutual information and the measure under scrutiny. For each combination of modalities (PET-MR or MR-CT), the statistical testing was performed on the rectified and nonrectified images separately, because of the dependence between an image and its rectified version.

As a first illustration of the performance of the different information measures, registration functions are given in Figs. 2 and 3. These show the measures as a function of one of the transformation parameters. The position of zero translation (Fig. 2) or zero rotation (Fig. 3) corresponds to the gold standard solution.

In Fig. 2, the behavior of the measures for translation along an in-plane axis (left-right axis) is displayed. All measures show well-behaved functions with an optimum near the gold standard, except for the generalized measures of order 3. The order α causes the functions to be more peaked for values of α larger than 1 and, particularly, as α approaches 0. The behavior described is typical, both across patients and across modalities.

Fig. 3 illustrates the sensitivity of the measures to the size of the overlapping volume of the images. The measures are plotted as functions of rotation around an in-plane axis (left-right axis). Because the images have few slices and a large slice thickness in comparison to the in-plane voxel sizes, the number of samples in the overlapping volume decreases relatively fast for rotation around an in-plane axis (compared to translation along an in-plane axis). This can lead to erratic behavior of the functions, as is exhibited, for example, by $\chi^{1.5}$ and $M_{0.8}$. The sensitivity to overlap appears to be influenced by α. For α around 1, the function values sometimes increase for large rotations ($\chi^{1.5}$ for instance). When α either increases or decreases, this phenomenon declines.

The functions in Figs. 2 and 3 are not sampled densely. They show the global behavior of the measures. By viewing densely sampled functions on a smaller scale, we found that the smoothness of the registration functions varied with α. Fig. 4 contains close-ups of the optima of the functions in Fig. 3, for mutual information and I_α-information. The other measures of order α behaved similarly to I_α. The same behavior was found for other patients and PET-MR data.

In the following sections, we first compare the registration measures for the overall method, i.e., registering the images as one would in clinical practice. We then study the accuracy of the measures in more detail, by using the gold standard as an initial estimate for optimization. Finally, we focus on the smoothness of the registration functions for varying α.

A. Registration

A first method to evaluate the performance of the different registration measures is to register all image pairs assuming no prior knowledge about the registration solution is known. All translation and rotation parameters are initialized to zero. The center of each image was taken as the origin of its coordinate system. As a result, the centers of two images were aligned at the beginning of the registration process.

When the optimization method had converged, we computed a measure of registration error with respect to the marker-based gold standard solution. This measure was defined as the maximum deviation between the transformation found and the gold standard transformation, calculated on a sphere with its origin located at the image center and having a radius of 10 cm.

The results for these registration experiments for all 35 PET and MR (T1–, T2-, and PD-weighted; both without and with distortion correction) pairs and all 41 MR and CT image pairs are summarized in Figs. 5 and 6, respectively. For each registration measure, the 0.5 and 0.9 quantile errors over all image pairs are given. We prefer the 0.9 quantile error to the maximum value, because the maximum may represent a single outlier. When the 0.9 quantile error is high, it means that at least 10% of the distribution has a high error. Note that in the figures the error bars for the larger errors have been cut off for visualization purposes.

An important conclusion is that the registration errors of many information measures are of the same order of magnitude as the error of mutual information. Notable exceptions are the information measures of order $\alpha = 3.0$ for all experiments. This is not surprising in view of the registration functions given in Figs. 2 and 3. Furthermore, V-information shows rather large 0.9 quantile errors for both combinations of modalities, as do $M_{0.8}$ and several measures of order 0.2 and order 2.0 for MR-CT registration.

The quantile errors, however, only give two measures on an entire distribution. To better compare the complete distributions of errors, we have applied the Wilcoxon rank sum test to the results. For PET-MR registration, V, I_3, R_3, and χ^3 were considered significantly different ($p = 0.01$) from the mutual information measure I for both rectified and nonrectified data. For the MR and CT data, the measures of $\alpha = 0.2$ and 3.0 (I_α, R_α, and M_α/χ^α), as well as $M_{0.8}$ and V-information yielded significantly different results from mutual information based registration. When considering CT and nonrectified MR only, significantly different results were also found for all measures of order 0.5, 1.5, and 2.0.

Inspection of the registration errors revealed that these measures all performed significantly worse than mutual information. The results of all other measures did not differ significantly from the results of mutual information.

B. Accuracy

A drawback of the registration experiments in the previous section is that the results depend on more than just the regis-

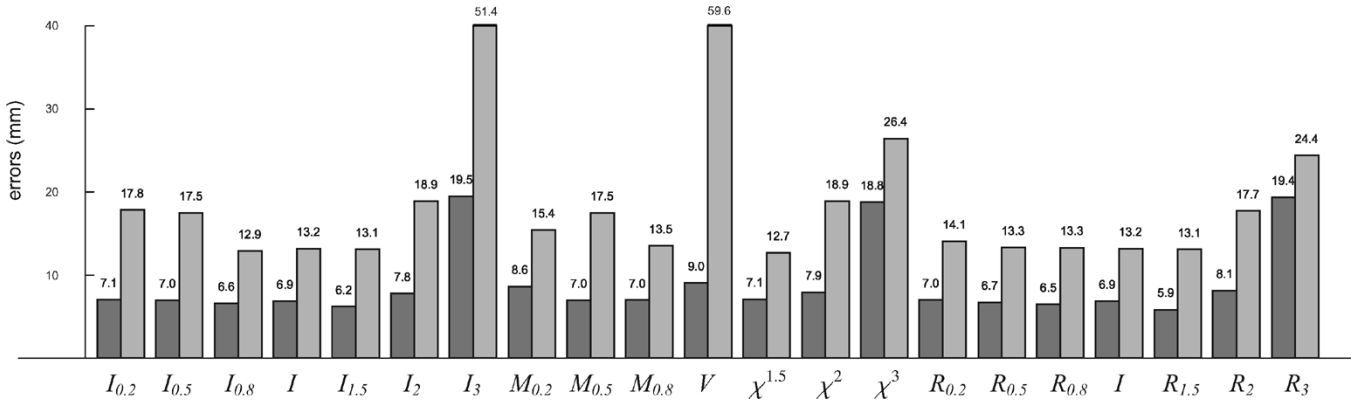

Fig. 5. 0.5 (dark grey) and 0.9 quantile errors (light grey) for PET to MR registration.

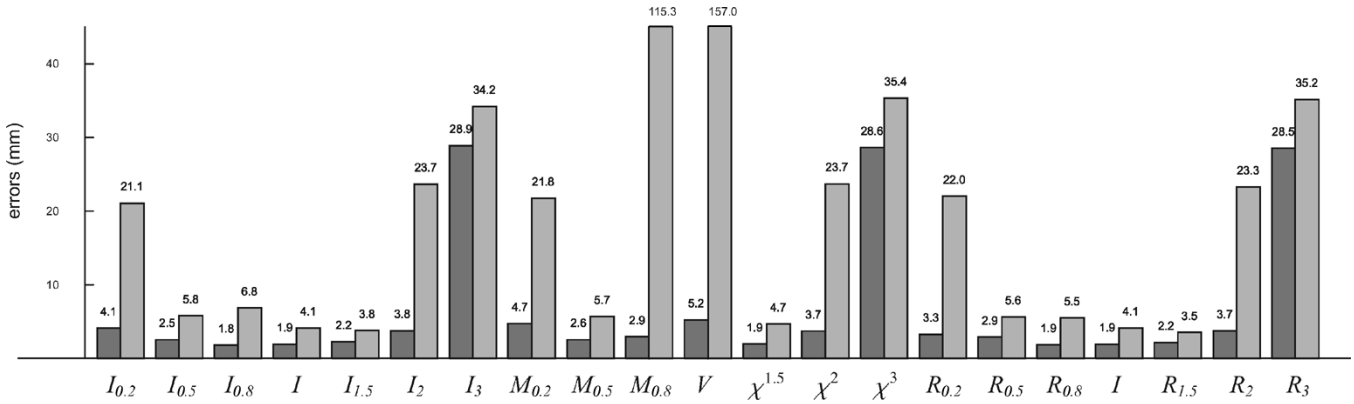

Fig. 6. 0.5 (dark grey) and 0.9 quantile errors (light grey) for MR to CT registration.

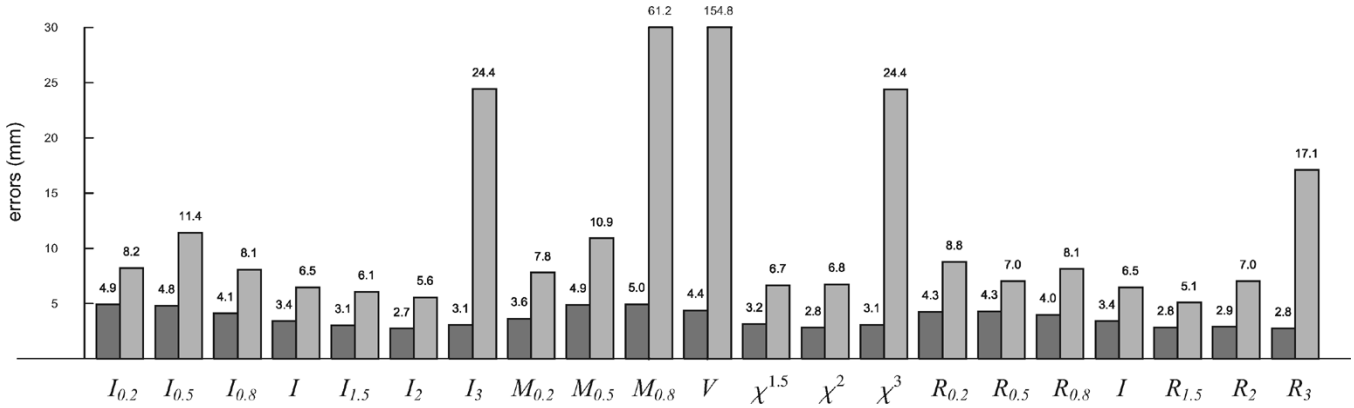

Fig. 7. 0.5 (dark grey) and 0.9 quantile errors (light grey) for PET to MR registration, starting close to the global optimum.

tration measure. In particular, the optimization method plays an important role in the outcome. Large errors may indicate that the registration function does not have a global optimum at the position of correct alignment, but it is more likely that they are a result of a less smooth function which is more difficult to optimize. The registration experiments in the previous section were carried out to evaluate the performance of the overall method, whereas in this section we want to study the attainable accuracy of the registration measures, in other words, the position of the function's global optimum. To this end, we have repeated all experiments using the gold standard solution as an initial estimate. This means that optimization should start close to the global optimum (assuming the global optimum does not differ much from the gold standard) and the chances of finding it are vastly increased.

Registration errors with respect to the gold standard were computed as described before and the 0.5 and 0.9 quantile errors can be found in Figs. 7 and 8.

Again, V-information is not a registration measure we would recommend, at least not for multimodality registration. For the measures of order α, different patterns emerge for PET-MR and MR-CT registration. The best results are reached when α is approximately 1.5 or 2.0 for PET and MR images. For MR to CT matching, on the other hand, the best results coincide with $\alpha = 0.2$. Our assumption is that the difference in optimal α is caused by differences in the intensity distributions of the im-

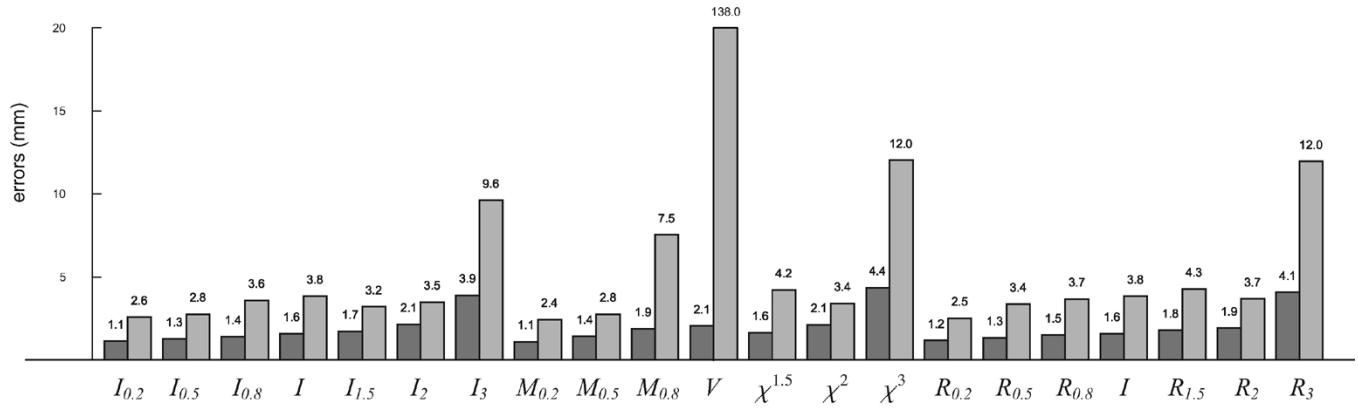

Fig. 8. 0.5 (dark grey) and 0.9 quantile errors (light grey) for MR to CT registration, starting close to the global optimum.

TABLE I
P-VALUES FOR THE NULL HYPOTHESIS THAT VARIOUS INFORMATION MEASURES ACHIEVE AN ACCURACY SIMILAR TO MUTUAL INFORMATION, WHEN STARTING FROM THE GOLD STANDARD

	$I_{0.2}$	$I_{0.5}$	$I_{0.8}$	$I_{1.5}$	$I_{2.0}$	$I_{3.0}$	V	$M_{0.2}$	$M_{0.5}$	$M_{0.8}$
PET–NR	0.1808	0.0026	0.0009	0.3570	0.2891	0.9308	0.0033	0.1219	0.0057	0.0017
PET–R	0.0843	0.1094	0.1578	0.0258	0.3305	0.1578	0.0092	0.5936	0.1981	0.0015
CT–NR	**0.0002**	**0.0010**	0.0582	0.1808	0.0129	0.0057	0.0129	**0.0037**	**0.0087**	0.2443
CT–R	0.1084	0.1005	0.9108	0.7652	0.3135	0.0304	0.0051	**0.0080**	0.2627	0.0028

	$\chi^{1.5}$	$\chi^{2.0}$	$\chi^{3.0}$	$R_{0.2}$	$R_{0.5}$	$R_{0.8}$	$R_{1.5}$	$R_{2.0}$	$R_{3.0}$
PET–NR	0.7677	0.4342	0.2046	0.1138	0.0208	0.0106	0.0250	0.8757	0.8213
PET–R	0.8261	0.1401	0.7299	0.3305	0.0962	0.1771	0.0413	0.0555	0.3627
CT–NR	0.0853	0.0157	0.0041	**0.0046**	**0.0087**	0.9032	0.0325	0.0208	0.0017
CT–R	0.1913	0.1672	0.0276	0.0187	0.1672	0.4115	0.5257	0.9108	0.1169

Values in bold denote cases where mutual information is outperformed and which could be considered significant. The data sets have been divided into two categories: nonrectified (NR) and rectified (R) MR images.

ages. By varying α, the sensitivity of the information measures to the small probabilities in the distributions is changed. Small probabilities are associated with small structures in the images, which can play an important role in the fine-tuning and, hence, the accuracy of the registration. On the other hand, small probabilities are also a result of noise in the images, which can negatively influence the registration. The optimal α for two images is, therefore, dependent on the probability distributions and it can vary for different modalities. It is likely it is affected by the number of histogram bins, noise levels and the part of the anatomy imaged. Whether an optimal choice of α can be determined *a priori* needs to be investigated further.

Some of the information measures actually seem to outperform mutual information in this accuracy test. We have used statistical significance testing again (Wilcoxon paired signed rank sum test) to further compare the results. The chances that the results for mutual information and the other measures do not differ (the null hypothesis) are given in Table I. For chances that could be considered significant ($p < 0.01$; note that no correction for multiple comparisons is included), we determined whether the measure performed better or worse than mutual information. The cases of better performance are shown in bold face.

Significantly more accurate results were found for nonrectified MR and CT images when α has a value around 0.2 or 0.5 (I_α, M_α, and R_α). In a single case, these results were also significant with rectified MR images.

C. Robustness

The robustness of a method with respect to the initial misregistration depends primarily on the smoothness of the registration function. A smooth function is easy to optimize and large initial misregistrations can be corrected. The functions in Figs. 2 and 3 show varying degrees of smoothness for the different measures. These functions, however, only show one-dimensional lines through a six-dimensional (6-D) space and can, thus, only be seen as an indication of the functions' smoothness. In order to study the behavior of a higher dimensional registration function more rigorously we have employed the following method. Around the global optimum of a function we place a 6-D hypercube of a certain size. Starting in the 64 corner points of the hypercube, we perform hill-climbing optimizations, i.e., we consider all points that are a given step size away from the current position and move in the direction of the one with the highest function value until the current position has a value higher than all neighboring points. We then count the number of different maxima that the 64 optimizations ended up in. A maximum is considered different when at least one of the six transformation parameters differs by more than the hill-climbing step size from all the maxima found. The number of different maxima found

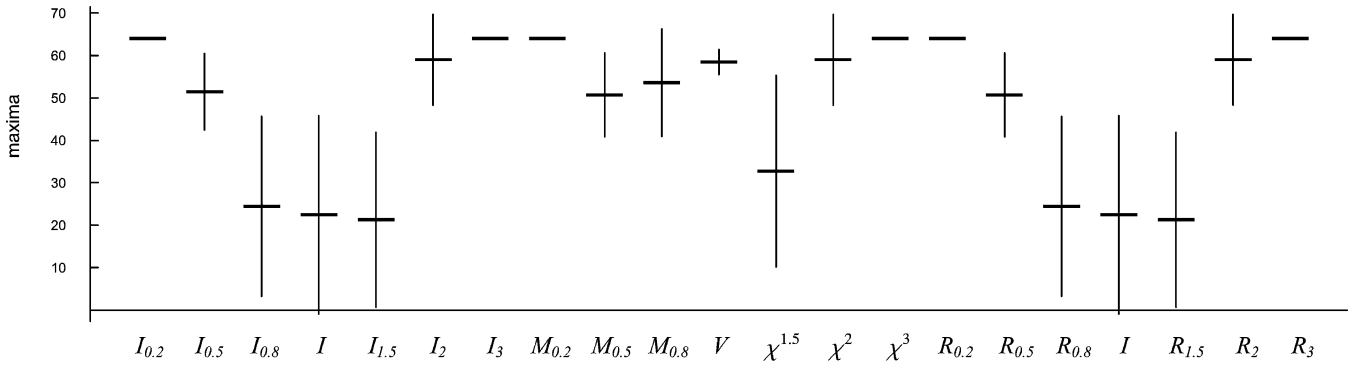

Fig. 9. Average (horizontal line) and standard deviation (vertical line) of the number of different end positions reached from 64 different starting positions, for seven MR-T1 and CT data sets.

can be viewed as a measure of the smoothness of the function, with fewer maxima denoting a smoother function.

The described method has been applied to all seven nonrectified MR-T1 images and the corresponding CT images. The center of the hypercube was placed at the position of the gold standard solution. In this manner, the parts of the search space being compared were identical for each measure. The gold standard solution is not equal to the global optimum, but the global optimum of each measure (which we assume to be the position found with the experiments in Section IV-B) was always located within the hypercube. The sides of the hypercube had a length of 10 units. Consequently, the transformation parameters at the corner points each had an offset of ± 5 millimeters or degrees from the gold standard solution. The step size for the hill-climbing optimization was chosen to be 0.2 units (millimeters or degrees). Fig. 9 shows the average and standard deviation of the number of different maxima that was found for all f-information measures under scrutiny.

Apart from the linear relationships between certain measures, as given at the end of Section III, some other relationships also hold. In particular, $R_2 = \log(\chi^2 + 1)$ and $R_\alpha = 1/(\alpha - 1)\log(1+\alpha(\alpha-1)I_\alpha)$. The hill-climbing method we used is not affected by operations like taking the logarithm. The results of the related measures are, therefore, identical for our robustness study. Other optimization methods however, such as Powell's, *are* influenced by the differences in the measures. This is the reason the results for related measures do differ in Section IV-A, in Figs. 5 and 6. Because the total performance of a similarity measure can only be viewed in the broader context of its implementation, the nonlinearly related measures are interesting to compare. For the smoothness experiments in this section, however, no information is gained from the inclusion of both I_α and R_α. They are both displayed for no other reasons than completeness and easy comparison to figures in other sections.

The results confirm the behavior found in the plots in Figs. 2 and 3: the functions are most smooth for α around 1. As α approaches 0 and also for increasing α ($\alpha \geq 2$), the average number of maxima found increases rapidly.

V. Conclusion and Discussion

We have described the results of registering clinical PET, MR (T1–, T2–, and PD-weighted; both without and with distortion correction) and CT images using different measures from information theory. The popular and extensively researched measure of mutual information was compared with V-, I_α-, and χ^α-information, as well as measures of order α defined by Matusita and Rényi. All measures denote the divergence of the joint distribution of the images' grey values from the joint distribution for complete independence of the images. Maximization of these measures—of the dependence of the grey values—is assumed to register the images.

The smoothness of the registration function (the function to be optimized in order to find the transformation that registers the images) was influenced by the value of α. The functions of the different measures were shown to loose smoothness as α approached 0 and, quite substantially, when α was larger than 2 ($\alpha = 3.0$).

The performance of the measures was evaluated both by registration of the images without using any prior knowledge about correct alignment and by registration using a marker-based gold standard to initialize optimization. The former experiments measure the performance of the complete method (including the optimization method); the latter experiments study the accuracy that can possibly be achieved for each measure. Having applied statistical significance testing to errors computed against the gold standard, V-information was found to perform poorly for PET-MR and MR-CT matching, both in overall performance and in accuracy. The Matusita measure M_α usually yielded poor results for a value of α close to 1, in accordance with the fact that M_1 equals V-information. Measures of order α usually showed deteriorating results as α increased, for $\alpha > 2$.

For the accuracy experiments, significantly better results were found for several measures, compared to mutual information. For registration of CT and nonrectified MR images, more accurate results were obtained for I_α, R_α and M_α, with α either 0.2 or 0.5. In one case, the results were also significant for the experiments with the rectified MR images.

The use of information measures other than mutual information showed promise in achieving better registration results. Although some measures yield registration functions that are less smooth and, therefore, more difficult to optimize, I_α, R_α, and M_α were shown to occasionally produce more accurate results for the registration of MR and CT images. Because I_α and R_α equal mutual information for the limit $\alpha \to 1$, it may prove beneficial to start registration with $\alpha = 1$ to take advantage of

the smoothness of the function and to adapt the value of α in subsequent iterations for better accuracy. The optimal value of α differed per modality. It could not be determined from the results found.

APPENDIX
I_α-INFORMATION AND KULLBACK–LEIBLER DISTANCE

To proof: $\lim_{\alpha\to 1} I_\alpha(P\|Q) = \sum p\log(p/q)$ (Kullback–Leibler distance)

$$\begin{aligned}\lim_{\alpha\to 1} I_\alpha(P\|Q) &= \lim_{\alpha\to 1}\frac{1}{\alpha(\alpha-1)}\left(\sum p^\alpha q^{1-\alpha}-1\right)\\ &= \lim_{\alpha\to 1}\frac{(p^\alpha q^{1-\alpha}\log p - p^\alpha q^{1-\alpha}\log q)}{(2\alpha-1)}\\ &= \sum p\log\frac{p}{q}\end{aligned}$$

using l'Hôpital's rule in the second step: if both $\lim_{\alpha\to 1} f(x)$ and $\lim_{\alpha\to 1} g(x)$ equal zero, $\lim_{\alpha\to 1}(f(x)/g(x)) = \lim_{\alpha\to 1}(f'(x)/g'(x))$.

ACKNOWLEDGMENT

The images and the standard transformation(s) were provided as part of the project, "Retrospective Image Registration Evaluation," National Institutes of Health, Project Number 8R01EB002 124-03, Principal Investigator J. M. Fitzpatrick, Vanderbilt University, Nashville, TN. The authors would like to thank the Laboratory for Medical Imaging Research in Leuven, The Netherlands, for supplying us with their software for mutual-information-based registration.

REFERENCES

[1] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," in *Information Processing in Medical Imaging*, Y. Bizais, C. Barillot, and R. Di Paola, Eds. Dordrecht, The Netherlands: Kluwer Academic, 1995, pp. 263–274.

[2] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," in *Proc. Int. Conf. Computer Vision*, E. Grimson, S. Shafer, A. Blake, and K. Sugihara, Eds., 1995, pp. 16–23.

[3] M. Holden, D. L. G. Hill, E. R. E. Denton, J. M. Jarosz, T. C. S. Cox, T. Rohlfing, J. Goodey, and D. J. Hawkes, "Voxel similarity measures for 3-D serial MR brain image registration," *IEEE Trans. Med. Imag.*, vol. 19, pp. 94–102, Feb. 2000.

[4] B. Likar and F. Pernuš, "A hierarchical approach to elastic registration based on mutual information," *Image Vis. Computing*, vol. 19, no. 1–2, pp. 33–44, 2001.

[5] 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, pp. 187–198, Apr. 1997.

[6] C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, G. L. Lecarpentier, J. B. Fowlkes, M. A. Roubidoux, and P. L. Carson, "Semiautomatic registration of volumetric ultrasound scans," *Ultrasound Med. Biol.*, vol. 25, no. 3, pp. 339–347, 1999.

[7] 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 Trans. Med. Imag.*, vol. 18, pp. 712–721, Aug. 1999.

[8] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3 D medical image alignment," *Pattern Recogn.*, vol. 32, no. 1, pp. 71–86, 1999.

[9] P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Processing*, vol. 9, pp. 2083–2099, Dec. 2000.

[10] 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.

[11] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imag.*, vol. 22, pp. 986–1004, Aug. 2003.

[12] F. Maes, D. Vandermeulen, and P. Suetens, "Medical image registration using mutual information," *Proc. IEEE*, vol. 91, pp. 1699–1722, Oct. 2003.

[13] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "f-Information measures in medical image registration," in *Proc. SPIE Medical Imaging: Image Processing*, vol. 4322, M. Sonka and K. M. Hanson, Eds., 2001, pp. 579–587.

[14] Y.-M. Zhu, "Volume image registration by cross-entropy optimization," *IEEE Trans. Med. Imag.*, vol. 21, pp. 174–180, Feb 2002.

[15] M. P. Wachowiak, R. Smolíková, G. D. Tourassi, and A. S. Elmaghraby, "Similarity metrics based on nonadditive entropies for 2D-3D multimodal biomedical image registration," in *Proc. SPIE Medical Imaging: Image Processing*, vol. 5032, M. Sonka and J. M. Fitzpatrick, Eds., 2003, pp. 1090–1100.

[16] N. F. Rougon, C. Petitjean, and F. Prêteux, "Variational non rigid image registration using exclusive F-information," in *Proc. Int. Conf. Image Processing*, Los Alamitos, CA, 2003, pp. 703–706.

[17] Y. He, A. Ben Hamza, and H. Krim, "A generalized divergence measure for robust image registration," *IEEE Trans. Signal Processing*, vol. 51, pp. 1211–1220, May 2003.

[18] A. Bardera, M. Feixas, and I. Boada, "Normalized similarity measures for medical image registration," in *Proc. SPIE Medical Imaging: Image Processing*, vol. 5370, J. M. Fitzpatrick and M. Sonka, Eds. Bellingham, WA, 2004, pp. 108–118.

[19] I. Vajda, *Theory of Statistical Inference and Information*. Dordrecht, The Netherlands: Kluwer Academic, 1989.

[20] S. Kullback, *Information Theory and Statistics*. New York: Wiley , 1959.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[22] N. Dekker, L. S. Ploeger, and M. van Herk, "Evaluation of cost functions for gray value matching of two-dimensional images in radiotherapy," *Med. Phys.*, vol. 30, no. 5, pp. 778–784, 2003.

[23] L. Ding, A. Goshtasby, and M. Satter, "Volume image registration by template matching," *Image Vis. Computing*, vol. 19, no. 12, pp. 821–832, 2001.

[24] F. Liese and I. Vajda, "Convex statistical distances," in *Teubner-texte zur Mathematik*. Leipzig, Germany: BSB Teubner, 1987, vol. 95.

[25] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Mathematical Statistics and Probability*, vol. 1, Berkeley, CA, 1961, pp. 547–561.

[26] H. Chang and J. M. Fitzpatrick, "A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities," *IEEE Trans. Med. Imag.*, vol. 11, pp. 319–329, Mar. 1992.

[27] C. R. Maurer Jr., G. B. Aboutanos, B. M. Dawant, S. Gadamsetty, R. A. Margolin, R. J. Maciunas, and J. M. Fitzpatrick, "Effect of geometrical distortion correction in MR on image registration accuracy," *J. Comput. Assist. Tomogr.*, vol. 20, no. 4, pp. 666–679, 1996.

[28] 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. Assist. Tomogr.*, vol. 21, no. 4, pp. 554–566, 1997.

[29] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*. Cambridge, U.K: Cambridge Univ. Press, 1992.