

# Multimodality Image Registration by Maximization of Mutual Information

Frederik Maes,\* André Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens, *Member, IEEE*

**Abstract**—A new approach to the problem of multimodality medical image registration is proposed, using a basic concept from information theory, mutual information (MI), or relative entropy, as a new matching criterion. The method presented in this paper applies MI to measure the statistical dependence or information redundancy between the image intensities of corresponding voxels in both images, which is assumed to be maximal if the images are geometrically aligned. Maximization of MI is a very general and powerful criterion, because no assumptions are made regarding the nature of this dependence and no limiting constraints are imposed on the image content of the modalities involved. The accuracy of the MI criterion is validated for rigid body registration of computed tomography (CT), magnetic resonance (MR), and photon emission tomography (PET) images by comparison with the stereotactic registration solution, while robustness is evaluated with respect to implementation issues, such as interpolation and optimization, and image content, including partial overlap and image degradation. Our results demonstrate that subvoxel accuracy with respect to the stereotactic reference solution can be achieved completely automatically and without any prior segmentation, feature extraction, or other preprocessing steps which makes this method very well suited for clinical applications.

**Index Terms**—Matching criterion, multimodality images, mutual information, registration.

## I. INTRODUCTION

THE geometric alignment or registration of multimodality images is a fundamental task in numerous applications in three-dimensional (3-D) medical image processing. Medical diagnosis, for instance, often benefits from the complementarity of the information in images of different modalities. In radiotherapy planning, dose calculation is based on the computed tomography (CT) data, while tumor outlining is often better performed in the corresponding magnetic resonance (MR) scan. For brain function analysis, MR images provide anatomical information while functional information may be

obtained from positron emission tomography (PET) images, etc.

The bulk of registration algorithms in medical imaging (see [3], [16], and [23] for an overview) can be classified as being either frame based, point landmark based, surface based, or voxel based. *Stereotactic frame-based* registration is very accurate, but inconvenient, and cannot be applied retrospectively, as with any *external point landmark-based* method, while *anatomical point landmark-based* methods are usually labor-intensive and their accuracy depends on the accurate indication of corresponding landmarks in all modalities. *Surface-based* registration requires delineation of corresponding surfaces in each of the images separately. But surface segmentation algorithms are generally highly data and application dependent and surfaces are not easily identified in functional modalities such as PET. *Voxel-based* (VSB) registration methods optimize a functional measuring the similarity of all geometrically corresponding voxel pairs for some feature. The main advantage of VSB methods is that feature calculation is straightforward or even absent when only grey-values are used, such that the accuracy of these methods is not limited by segmentation errors as in surface based methods.

For *intramodality registration* multiple VSB methods have been proposed that optimize some global measure of the absolute difference between image intensities of corresponding voxels within overlapping parts or in a region of interest (ROI) [5], [11], [19], [26]. These criteria all rely on the assumption that the intensities of the two images are linearly correlated, which is generally not satisfied in the case of *intermodality registration*. Crosscorrelation of feature images derived from the original image data has been applied to CT/MR matching using geometrical features such as edges [15] and ridges [24] or using especially designed intensity transformations [25]. But feature extraction may introduce new geometrical errors and requires extra calculation time. Furthermore, correlation of sparse features like edges and ridges may have a very peaked optimum at the registration solution, but at the same time be rather insensitive to misregistration at larger distances, as all nonedge or nonridge voxels correlate equally well. A multiresolution optimization strategy is therefore required, which is not necessarily a disadvantage, as it can be computationally attractive.

In the approach of Woods *et al.* [30] and Hill *et al.* [12], [13], misregistration is measured by the dispersion of the two-dimensional (2-D) histogram of the image intensities of corresponding voxel pairs, which is assumed to be minimal in the registered position. But the dispersion measures they

Manuscript received February 21, 1996; revised July 23, 1996. This work was supported in part by IBM Belgium (Academic Joint Study) and by the Belgian National Fund for Scientific Research (NFWO) under Grants FGWO 3.0115.92, 9.0033.93 and G.3115.92. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was N. Ayache. *Asterisk indicates corresponding author.*

\*F. Maes is with the Laboratory for Medical Imaging Research, Katholieke Universiteit Leuven, ESAT/ Radiologie, Universitair Ziekenhuis Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium. He is an Aspirant of the Belgian National Fund for Scientific Research (NFWO) (e-mail: Frederik.Maes@uz.kuleuven.ac.be).

A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens are with the Laboratory for Medical Imaging Research, Katholieke Universiteit Leuven, ESAT/Radiologie, Universitair Ziekenhuis Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.

Publisher Item Identifier S 0278-0062(97)02397-5.

propose are largely heuristic. Hill's criterion requires segmentation of the images or delineation of specific histogram regions to make the method work [20], while Woods' criterion is based on additional assumptions concerning the relationship between the grey-values in the different modalities, which reduces its applicability to some very specific multimodality combinations (PET/MR).

In this paper, we propose to use the much more general notion of *mutual information* (MI) or relative entropy [8], [22] to describe the dispersive behavior of the 2-D histogram. MI is a basic concept from information theory, measuring the statistical dependence between two random variables or the amount of information that one variable contains about the other. The MI registration criterion presented here states that the MI of the image intensity values of corresponding voxel pairs is maximal if the images are geometrically aligned. Because no assumptions are made regarding the nature of the relation between the image intensities in both modalities, this criterion is very general and powerful and can be applied automatically without prior segmentation on a large variety of applications.

This paper expands on the ideas first presented by Collignon *et al.* [7]. Related work in this area includes the work by Viola and Wells *et al.* [27], [28] and by Studholme *et al.* [21]. The theoretical concept of MI is presented in Section II, while the implementation of the registration algorithm is described in Section III. In Sections IV, V, and VI we evaluate the accuracy and the robustness of the MI matching criterion for rigid body CT/MR and PET/MR registration. Section VII summarizes our current findings, while Section VIII gives some directions for further work. In the Appendixes, we discuss the relationship of the MI registration criterion to other multimodality VSB criteria.

## II. THEORY

Two random variables,  $A$  and  $B$ , with marginal probability distributions,  $p_A(a)$  and  $p_B(b)$ , and joint probability distribution,  $p_{AB}(a, b)$ , are statistically independent if  $p_{AB}(a, b) = p_A(a) \cdot p_B(b)$ , while they are maximally dependent if they are related by a one-to-one mapping  $T$ :  $p_A(a) = p_B(T(a)) = p_{AB}(a, T(a))$ . MI,  $I(A, B)$ , measures the degree of dependence of  $A$  and  $B$  by measuring the distance between the joint distribution  $p_{AB}(a, b)$  and the distribution associated to the case of complete independence  $p_A(a) \cdot p_B(b)$ , by means of the Kullback-Leibler measure [22], i.e.,

$$I(A, B) = \sum_{a,b} p_{AB}(a, b) \log \frac{p_{AB}(a, b)}{p_A(a) \cdot p_B(b)}. \quad (1)$$

MI is related to entropy by the equations

$$I(A, B) = H(A) + H(B) - H(A, B) \quad (2)$$

$$= H(A) - H(A | B) \quad (3)$$

$$= H(B) - H(B | A) \quad (4)$$

with  $H(A)$  and  $H(B)$  being the entropy of  $A$  and  $B$ , respectively,  $H(A, B)$  their joint entropy, and  $H(A | B)$  and

TABLE I  
SOME PROPERTIES OF MUTUAL INFORMATION

|                   |  |
|-------------------|--|
| Non-negativity:   | $I(A, B) \geq 0$   |
| Independence:     | $I(A, B) = 0 \Leftrightarrow p_{AB}(a, b) = p_A(a) \cdot p_B(b)$ |
| Symmetry:         | $I(A, B) = I(B, A)$  |
| Self information: | $I(A, A) = H(A)$   |
| Boundedness:      | $I(A, B) \leq \min(H(A), H(B))$                                  |
|                   | $\leq (H(A) + H(B))/2$   |
|                   | $\leq \max(H(A), H(B))$  |
|                   | $\leq H(A, B)$   |
|                   | $\leq H(A) + H(B)$   |
| Data processing:  | $I(A, B) \geq I(A, T(B))$  |

$H(B | A)$  the conditional entropy of  $A$  given  $B$  and of  $B$  given  $A$ , respectively

$$H(A) = - \sum_a p_A(a) \log p_A(a) \quad (5)$$

$$H(A, B) = - \sum_{a,b} p_{AB}(a, b) \log p_{AB}(a, b) \quad (6)$$

$$H(A | B) = - \sum_{a,b} p_{AB}(a, b) \log p_{A|B}(a | b). \quad (7)$$

The entropy  $H(A)$  is known to be a measure of the amount of uncertainty about the random variable  $A$ , while  $H(A | B)$  is the amount of uncertainty left in  $A$  when knowing  $B$ . Hence, from (3),  $I(A, B)$  is the reduction in the uncertainty of the random variable  $A$  by the knowledge of another random variable  $B$ , or, equivalently, the amount of information that  $B$  contains about  $A$ . Some properties of MI are summarized in Table I (see [22] for their proof).

Considering the image intensity values,  $a$  and  $b$ , of a pair of corresponding voxels in the two images that are to be registered to be random variables  $A$  and  $B$ , respectively, estimations for the joint and marginal distributions  $p_{AB}(a, b)$ ,  $p_A(a)$ , and  $p_B(b)$  can be obtained by simple normalization of the joint and marginal histograms of the overlapping parts of both images. Intensities  $a$  and  $b$  are related through the geometric transformation  $T_\alpha$  defined by the registration parameter  $\alpha$ . The MI registration criterion states that the images are geometrically aligned by the transformation  $T_{\alpha^*}$  for which  $I(A, B)$  is maximal. This is illustrated in Fig. 1 for a CT and an MR image of the brain, showing the 2-D histogram of the image intensity values in a nonregistered and in the registered position. The high-intensity values in the histogram of the CT image originating from the bone of the skull are most likely to be mapped on low-intensity values in the histogram of the MR image if the images are properly aligned, resulting in a peak in the 2-D histogram. The uncertainty about the MR voxel intensity is thus largely reduced if the corresponding CT voxel is known to be of high intensity. This correspondence is lost in case of misregistration. However, the MI criterion does not make limiting assumptions regarding the relation between image intensities of corresponding voxels in the different modalities, which is highly data dependent, and no constraints are imposed on the image content of the modalities involved.

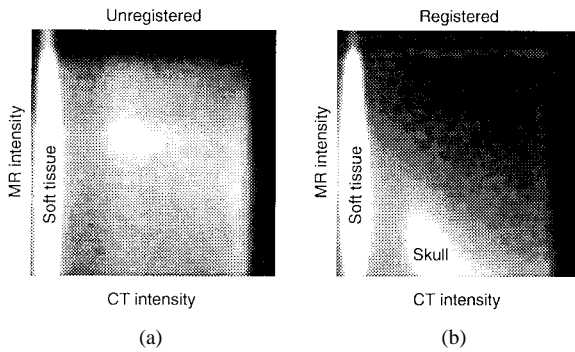


Fig. 1. Joint histogram of the overlapping volume of the CT and MR brain images of dataset A in Tables II and III: (a) Initial position:  $I(CT, MR) = 0.46$ , (b) registered position:  $I(CT, MR) = 0.89$ . Misregistration was about 20 mm and  $10^\circ$  (see the parameters in Table III).

If both marginal distributions  $p_A(a)$  and  $p_B(b)$  can be considered to be independent of the registration parameters  $\alpha$ , the MI criterion reduces to minimizing the joint entropy  $H_{AB}(A, B)$  [6]. If either  $p_A(a)$  or  $p_B(b)$  is independent of  $\alpha$ , which is the case if one of the images is always completely contained in the other, the MI criterion reduces to minimizing the conditional entropy  $H(A | B)$  or  $H(B | A)$ . However, if both images only partially overlap, which is very likely during optimization, the volume of overlap will change when  $\alpha$  is varied and  $p_A(a)$  and  $p_B(b)$  and also  $H(A)$  and  $H(B)$  will generally depend on  $\alpha$ . The MI criterion takes this into account explicitly, as becomes clear in (2), which can be interpreted as follows [27]: “maximizing MI will tend to find as much as possible of the complexity that is in the separate datasets (maximizing the first two terms) so that at the same time they explain each other well (minimizing the last term).”

For  $I(A, B)$  to be useful as a registration criterion and well behaved with respect to optimization,  $I(A, B)$  should vary smoothly as a function of misregistration  $|\alpha - \alpha^*|$ . This requires  $p_A(a)$ ,  $p_B(b)$  and  $p_{AB}(a, b)$  to change smoothly when  $\alpha$  is varied, which will be the case if the image intensity values are spatially correlated. This is illustrated by the graphs in Fig. 2, showing the behavior of  $I(A, B)$  as a function of misregistration between an image and itself rotated around the image center. The trace on the left is obtained from an original MR image and shows a single sharp optimum with a rather broad attraction basin. The trace on the right is obtained from the same image after having reduced the spatial correlation of the image intensity by repeatedly swapping pairs of randomly selected pixels. This curve shows many local maxima and the attraction basin of the global maximum is also much smaller, which deteriorates the optimization robustness. Thus, although the formulation of the MI criterion suggests that spatial dependence of image intensity values is not taken into account, such dependence is in fact essential for the criterion to be well behaved around the registration solution.

### III. ALGORITHM

#### A. Transformation

Each of the images is associated an image coordinate frame with its origin positioned in a corner of the image, with the

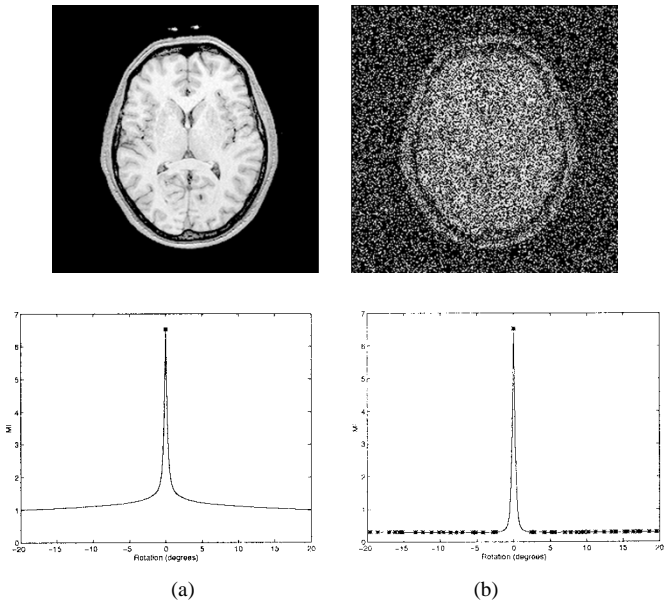


Fig. 2. Spatial correlation of image intensity values increases MI registration robustness. Top: (a) original  $256 \times 256$  2-D MR image and (b) image of (a) shuffled by swapping 30 000 randomly selected pixel pairs. Both images have the same image content. Bottom: MI registration traces obtained using partial volume distribution (PV) interpolation for in-plane rotation of each image over itself. Local maxima are marked with “\*”.

$x$  axis along the row direction, the  $y$  axis along the column direction, and the  $z$  axis along the plane direction.

One of the images is selected to be the *floating* image,  $\mathbf{F}$ , from which samples  $s \in S$  are taken and transformed into the *reference* image,  $\mathbf{R}$ .  $S$  can be the set of grid points of  $\mathbf{F}$  or a sub- or superset thereof. Subsampling of the floating image might be used to increase speed performance, while supersampling aims at increasing accuracy. For each value of the registration parameter  $\alpha$  only those values  $s \in S_\alpha \subset S$  are retained for which  $T_\alpha s$  falls inside the volume of  $\mathbf{R}$ .

In this paper, we have restricted the transformation  $T_\alpha$  to rigid-body transformations only, although it is clear that the MI criterion can be applied to more general transformations as well. The rigid-body transformation is a superposition of a 3-D rotation and a 3-D translation and the registration parameter  $\alpha$  is a six-component vector consisting of three rotation angles  $\phi_x, \phi_y, \phi_z$  (measured in degrees) and three translation distances  $t_x, t_y, t_z$  (measured in millimeters). Transformation of image coordinates  $P_F$  to  $P_R$  from the image  $\mathbf{F}$  to image  $\mathbf{R}$  is given by

$$V_R \cdot (P_R - C_R) = R_x(\phi_x) \cdot R_y(\phi_y) \cdot R_z(\phi_z) \cdot V_F \cdot (P_F - C_F) + t(t_x, t_y, t_z) \quad (8)$$

with  $V_F$  and  $V_R$  being  $3 \times 3$  diagonal matrixes representing the voxel sizes of images  $\mathbf{F}$  and  $\mathbf{R}$ , respectively (in millimeters),  $C_F$  and  $C_R$  the image coordinates of the centers of the images,  $R = R_x \cdot R_y \cdot R_z$  the  $3 \times 3$  rotation matrix, with the matrixes  $R_x$ ,  $R_y$ , and  $R_z$  representing rotations around the  $x$ -,  $y$ -, and  $z$ -axis, respectively, and  $t$  the translation vector.

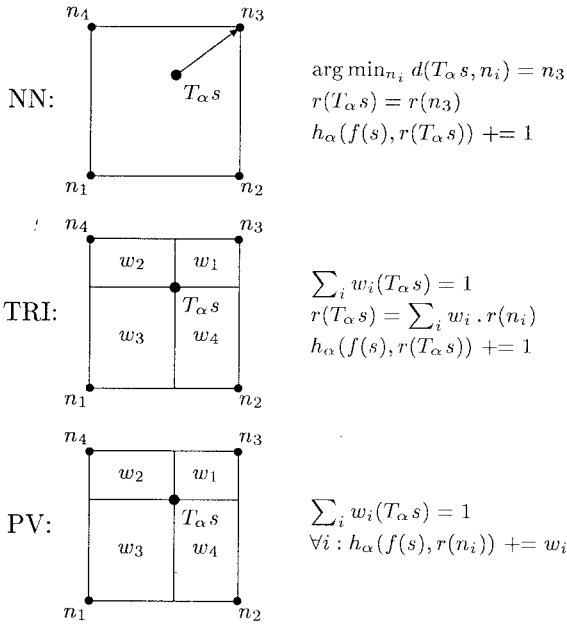


Fig. 3. Graphical illustration of NN, TRI, and PV interpolation in 2-D. NN and TRI interpolation find the reference image intensity value at position  $T_\alpha s$  and update the corresponding joint histogram entry, while PV interpolation distributes the contribution of this sample over multiple histogram entries defined by its NN intensities, using the same weights as for TRI interpolation.

### B. Criterion

Let  $f(s)$  denote the image intensity in the floating image  $\mathbf{F}$  at position  $s$  and  $r(T_\alpha s)$  the intensity at the transformed position in the reference image  $\mathbf{R}$ . The joint image intensity histogram  $h_\alpha(f, r)$  of the overlapping volume of both images at position  $\alpha$  is computed by binning the image intensity pairs  $(f(s), r(T_\alpha s))$  for all  $s \in S_\alpha$ . In order to do this efficiently, the floating and the reference image intensities are first linearly rescaled to the range  $[0, n_F - 1]$  and  $[0, n_R - 1]$ , respectively,  $n_F \times n_R$  being the total number of bins in the joint histogram. Typically, we use  $n_F = n_R = 256$ .

In general,  $T_\alpha s$  will not coincide with a grid point of  $\mathbf{R}$  and interpolation of the reference image is needed to obtain the image intensity value  $r(T_\alpha s)$ . *Nearest neighbor* (NN) interpolation of  $\mathbf{R}$  is generally insufficient to guarantee subvoxel accuracy, as it is insensitive to translations up to one voxel. Other interpolation methods, such as *trilinear* (TRI) interpolation, may introduce new intensity values which are originally not present in the reference image, leading to unpredictable changes in the marginal distribution  $p_{R,\alpha}(r)$  of the reference image for small variations of  $\alpha$ . To avoid this problem, we propose to use trilinear partial volume distribution (PV) interpolation to update the joint histogram for each voxel pair  $(s, T_\alpha s)$ . Instead of interpolating new intensity values in  $\mathbf{R}$ , the contribution of the image intensity  $f(s)$  of the sample  $s$  of  $\mathbf{F}$  to the joint histogram is distributed over the intensity values of all eight NN's of  $T_\alpha s$  on the grid of  $\mathbf{R}$ , using the same weights as for TRI interpolation (Fig. 3). Each entry in the joint histogram is then the sum of smoothly varying fractions of one, such that the histogram changes smoothly as  $\alpha$  is varied.

Estimations for the marginal and joint image intensity distributions  $p_{F,\alpha}(f)$ ,  $p_{R,\alpha}(r)$ , and  $p_{FR,\alpha}(f, r)$  are obtained by normalization of  $h_\alpha(f, r)$

$$p_{FR,\alpha}(f, r) = \frac{h_\alpha(f, r)}{\sum_{f,r} h_\alpha(f, r)} \quad (9)$$

$$p_{F,\alpha}(f) = \sum_r p_{FR,\alpha}(f, r) \quad (10)$$

$$p_{R,\alpha}(r) = \sum_f p_{FR,\alpha}(f, r). \quad (11)$$

The MI registration criterion  $I(\alpha)$  is then evaluated by

$$I(\alpha) = \sum_{f,r} p_{FR,\alpha}(f, r) \log_2 \frac{p_{FR,\alpha}(f, r)}{p_{F,\alpha}(f) p_{R,\alpha}(r)} \quad (12)$$

and the optimal registration parameter  $\alpha^*$  is found from

$$\alpha^* = \arg \max_{\alpha} I(\alpha). \quad (13)$$

### C. Search

The images are initially positioned such that their centers coincide and that the corresponding scan axes of both images are aligned and have the same orientation. Powell's multidimensional direction set method is then used to maximize  $I(\alpha)$ , using Brent's one-dimensional optimization algorithm for the line minimizations [18]. The direction matrix is initialized with unit vectors in each of the parameter directions. An appropriate choice for the order in which the parameters are optimized needs to be specified, as this may influence optimization robustness. For instance, when matching images of the brain, the horizontal translation and the rotation around the vertical axis are more constrained by the shape of the head than the pitching rotation around the left-to-right horizontal axis. Therefore, first aligning the images in the horizontal plane by first optimizing the in-plane parameters  $(t_x, t_y, \phi_z)$  may facilitate the optimization of the out-of-plane parameters  $(\phi_x, \phi_y, t_z)$ . However, as the optimization proceeds, the Powell algorithm may introduce other optimization directions and change the order in which these are considered.

### D. Complexity

The algorithm was implemented on an IBM RS/6000 workstation (AIX 4.1.3, 58 MHz, 185 SPECfp92; source code is available on request). The computation time required for one evaluation of the MI criterion varies linearly with the number of samples taken from the floating image. While TRI and PV interpolation have nearly the same complexity (1.4 s per million samples), NN interpolation is about three times as efficient (0.5 s per million samples). The number of criterion evaluations performed during optimization typically varies between 200 and 600, depending on the initial position of the images, on the order in which the parameters are optimized, and on the convergence parameters specified for the Brent and Powell algorithm.

TABLE II  
DATASETS USED IN THE EXPERIMENTS DISCUSSED IN SECTIONS V AND VI

| Set | Image | Size               | Voxels (mm)          | Range            |
|-----|-------|--------------------|----------------------|------------------|
| A   | MR    | $256^2 \times 180$ | $0.98^2 \times 1.00$ | $0 \times 4094$  |
|     | CT    | $256^2 \times 100$ | $0.94^2 \times 1.55$ | $0 \times 4093$  |
| B   | MR    | $200^2 \times 45$  | $1.25^2 \times 4.00$ | $38 \times 2940$ |
|     | CT    | $192^2 \times 39$  | $1.25^2 \times 4.00$ | $0 \times 2713$  |
| C   | MR    | $256^2 \times 24$  | $1.25^2 \times 4.00$ | $2 \times 2087$  |
|     | CT    | $512^2 \times 29$  | $0.65^2 \times 4.00$ | $0 \times 2960$  |
|     | PET   | $128^2 \times 15$  | $2.59^2 \times 8.00$ | $0 \times 683$   |
| D   | MR    | $256^2 \times 30$  | $1.33^2 \times 4.00$ | $2 \times 3359$  |

#### IV. EXPERIMENTS

The performance of the MI registration criterion was evaluated for rigid-body registration of MR, CT, and PET images of the brain of the same patient. The rigid-body assumption is well satisfied inside the skull in 3-D scans of the head if patient related changes (due to for instance interscanning operations) can be neglected, provided that scanner calibration problems and problems of geometric distortions have been minimized by careful calibration and scan parameter selection, respectively. Registration accuracy is evaluated in Section V by comparison with external marker-based registration results and other retrospective registration methods, while the robustness of the method is evaluated in Section VI with respect to implementation issues, such as sampling, interpolation and optimization, and image content, including image degradations, such as noise, intensity inhomogeneities and distortion, and partial image overlap. Four different datasets are used in the experiments described below (Table II). Dataset A<sup>1</sup> contains high-resolution MR and CT images, while dataset B was obtained by smoothing and subsampling the images of dataset A to simulate lower resolution data. Dataset C<sup>2</sup> contains stereotactically acquired MR, CT, and PET images, which have been edited to remove stereotactic markers. Dataset D contains an MR image only and is used to illustrate the effect of various image degradations on the registration criterion. All images consist of axial slices and in all cases the  $x$  axis is directed horizontally right to left, the  $y$  axis is directed horizontally front to back, and the  $z$  axis is directed vertically up, such that the image resolution is lowest in the  $z$  direction. In all experiments, the joint histogram size is  $256 \times 256$ , while the fractional precision convergence parameters for the Brent and Powell optimization algorithm are set to  $10^{-3}$  and  $10^{-5}$ , respectively [18].

#### V. ACCURACY

The images of datasets A, B, and C were registered using the MI registration criterion with different choices of the floating image and using different interpolation schemes. In each case the same optimization strategy was used, starting from all parameters initially equal to zero and optimizing the parameters in the order  $(t_x, t_y, \phi_z, \phi_x, \phi_y, t_z)$ . The results are summarized in Table III by the parameters of the transformation that

<sup>1</sup>Data provided by van den Elsen [25].

<sup>2</sup>Data provided by Fitzpatrick [10].

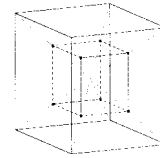


Fig. 4. The bounding box of the central eighth of the floating image defines eight points near the brain surface at which the difference between different registration transforms is evaluated.

takes the MR image as the reference image. Optimization required 300 to 500 evaluations of the MI criterion, which was performed on an IBM RS6000/3AT workstation using PV interpolation in about 20 min for CT to MR matching of dataset A (40 min for MR to CT matching) and in less than 2 min for PET to MR matching of dataset C.

The images of dataset A have been registered by van den Elsen [25] using a correlation-based VSB registration method. Visual inspection showed this result to be more accurate than skin marker-based registration and we use it as a reference to validate registration accuracy of the MI criterion for datasets A and B. For dataset C, we compare our results with the stereotactic registration solution provided by Fitzpatrick [10]. The difference between the reference and each of the MI registration solutions was evaluated at eight points near the brain surface (Fig. 4). The reference solutions and the mean and the maximal absolute transformed coordinate differences measured at these points are included in Table III.

The solutions obtained for dataset A and for dataset B using different interpolation schemes or for a different choice of the floating image are all very similar. For dataset A the largest differences with the reference solutions occur for rotation around the  $x$  axis ( $0.7^\circ$ ), but these are all subvoxel. For dataset B the differences are somewhat larger, especially in the  $y$  direction due to an offset in the  $y$  translation parameter (0.8 mm). However, these translational differences may have been caused by interpolation and subsampling artifacts introduced when creating the images of dataset B.

For dataset C, CT to MR registration using TRI interpolation did not converge to the reference solution. In this case, CT to MR registration performs clearly worse than MR to CT registration, for which all differences are subvoxel, the largest being 1.2 mm in the  $y$  direction for the solution obtained using PV interpolation due to a  $1^\circ$  offset for the  $x$  rotation parameter. For MR to PET as well as for PET to MR registration, PV interpolation yields the smallest differences with the stereotactic reference solution, especially in the  $z$  direction, which are all subvoxel with respect to the voxelsizes of the PET image in case of MR to PET registration. Relatively large differences occur in the  $y$  direction due to offsets in the  $y$  translation parameter of about 1 to 2 mm.

#### VI. ROBUSTNESS

##### A. Interpolation and Optimization

The robustness of the MI registration criterion with respect to interpolation and optimization was evaluated for dataset A. The images were registered using either the CT or the MR

TABLE III  
REFERENCE AND MI REGISTRATION PARAMETERS FOR DATASETS A, B, AND C AND THE MEAN  
AND MAXIMAL ABSOLUTE DIFFERENCE EVALUATED AT EIGHT POINTS NEAR THE BRAIN SURFACE

| Set    | F/R            | Rotation (degrees) |       |       | Translation (mm) |       |        | Difference (mm) |             |             |             |
|--------|----------------|--------------------|-------|-------|------------------|-------|--------|-----------------|-------------|-------------|-------------|
|        |                | x                  | y     | z     | x                | y     | z      | x               | y           | z           |             |
| A      | Reference [25] | 9.62               | -3.13 | 2.01  | 7.00             | 1.14  | 18.15  |                 |             |             |             |
|        | CT/MR          | NN                 | 10.23 | -3.23 | 2.10             | 6.98  | 1.00   | 18.24           | 0.09 (0.18) | 0.40 (0.79) | 0.63 (0.84) |
|        |                | TRI                | 10.24 | -3.21 | 2.08             | 6.97  | 1.05   | 18.22           | 0.08 (0.16) | 0.40 (0.72) | 0.63 (0.80) |
|        |                | PV                 | 10.36 | -3.17 | 2.09             | 6.94  | 1.15   | 18.20           | 0.08 (0.17) | 0.48 (0.76) | 0.76 (0.89) |
|        | MR/CT          | NN                 | 10.24 | -3.17 | 2.09             | 6.95  | 1.04   | 18.18           | 0.08 (0.16) | 0.41 (0.74) | 0.64 (0.74) |
|        |                | TRI                | 10.24 | -3.15 | 2.07             | 6.92  | 1.00   | 18.23           | 0.08 (0.15) | 0.41 (0.76) | 0.64 (0.80) |
|        |                | PV                 | 10.39 | -3.14 | 2.09             | 6.90  | 1.15   | 18.18           | 0.10 (0.18) | 0.51 (0.77) | 0.79 (0.94) |
|        | B              | Reference [25]     | 9.62  | -3.13 | 2.01             | 7.00  | 1.14   | 18.15           |             |             |             |
|        |                | CT/MR              | NN    | 10.02 | -3.42            | 2.25  | 6.63   | 0.34            | 18.28       | 0.40 (0.83) | 0.80 (1.45) |
| TRI    |                |                    | 10.27 | -3.11 | 2.05             | 6.53  | 0.54   | 18.34           | 0.48 (0.54) | 0.61 (1.22) | 0.67 (0.99) |
| PV     |                |                    | 10.57 | -3.17 | 2.11             | 6.60  | 0.62   | 18.36           | 0.40 (0.53) | 0.68 (1.47) | 0.97 (1.32) |
| MR/CT  |                | NN                 | 10.17 | -3.06 | 2.25             | 6.47  | 0.30   | 17.90           | 0.54 (0.84) | 0.84 (1.57) | 0.57 (1.03) |
|        |                | TRI                | 10.03 | -3.05 | 2.22             | 6.44  | 0.37   | 18.19           | 0.56 (0.84) | 0.77 (1.34) | 0.42 (0.64) |
|        |                | PV                 | 10.29 | -3.16 | 2.08             | 6.48  | 0.33   | 17.95           | 0.52 (0.61) | 0.81 (1.48) | 0.69 (0.98) |
| C      |                | Reference [10]     | -0.63 | 0.05  | 4.74             | 26.15 | -41.08 | -12.35          |             |             |             |
|        |                | CT/MR              | NN    | 0.87  | 0.05             | 4.84  | 26.70  | -40.67          | -9.92       | 0.54 (0.70) | 0.74 (1.33) |
|        | TRI            |                    | 1.21  | -1.94 | 3.67             | 29.51 | -39.78 | 43.61           | -           | -           | -           |
|        | PV             |                    | -0.00 | 0.00  | 4.95             | 26.57 | -40.72 | -10.00          | 0.41 (0.77) | 0.49 (1.00) | 2.35 (3.28) |
|        | MR/CT          | NN                 | -0.21 | 0.00  | 4.95             | 26.56 | -41.27 | -12.01          | 0.41 (0.76) | 0.35 (0.71) | 0.62 (0.98) |
|        |                | TRI                | -0.51 | 0.25  | 5.03             | 26.35 | -40.80 | -11.84          | 0.42 (0.75) | 0.43 (0.79) | 0.51 (0.95) |
|        |                | PV                 | -1.58 | 0.13  | 4.97             | 26.48 | -41.39 | -12.18          | 0.35 (0.73) | 0.56 (1.18) | 1.38 (1.57) |
|        | C              | Reference [10]     | 1.52  | -1.17 | 4.22             | 27.62 | -2.60  | -4.46           |             |             |             |
|        |                | PET/MR             | NN    | 0.70  | 0.26             | 5.20  | 27.57  | -0.74           | -5.08       | 1.40 (2.28) | 1.82 (3.66) |
| TRI    |                |                    | 0.38  | 0.01  | 5.25             | 27.50 | -1.29  | -1.37           | 1.47 (2.31) | 1.62 (3.34) | 3.22 (6.46) |
| PV     |                |                    | 1.63  | 0.18  | 4.98             | 27.65 | -0.46  | -4.94           | 1.09 (1.83) | 2.14 (3.32) | 1.97 (2.46) |
| MR/PET |                | NN                 | 0.42  | 0.14  | 5.04             | 27.93 | -1.28  | -5.03           | 1.17 (2.16) | 1.47 (3.00) | 2.00 (4.03) |
|        |                | TRI                | 0.16  | -0.11 | 4.90             | 27.99 | -1.60  | -4.27           | 0.98 (1.90) | 1.27 (2.59) | 2.05 (3.66) |
|        |                | PV                 | 1.46  | -0.34 | 4.71             | 27.94 | -0.85  | -4.49           | 0.72 (1.44) | 1.74 (2.49) | 1.19 (1.37) |

volume as the floating image and using different interpolation methods. For each combination, various optimization strategies were tried by changing the order in which the parameters were optimized, each starting from the same initial position with all parameters set to zero.

The results are summarized in Fig. 5. These scatter plots compare each of the solutions found (represented by their registration parameters  $\alpha$ ) with the one for which the MI registration measure was maximal (denoted by  $\alpha^*$ ) for each of the interpolation methods separately, using either the CT or the MR image as the floating image. Different solutions are classified by the norm of the registration parameter difference vector  $|\alpha - \alpha^*|$  on the horizontal axis (using mm and degrees for the translation and rotation parameters, respectively) and by the difference in the value of the MI criterion ( $MI(\alpha^*) - MI(\alpha)$ ) on the vertical axis. Although the differences are small for each of the interpolation methods used, MR to CT registration seems to be somewhat more robust than CT to MR registration. More importantly, the solutions obtained using PV interpolation are much more clustered than those obtained using NN or TRI interpolation, indicating that the use of PV interpolation results in a much smoother behavior of the registration criterion. This is also apparent from traces in registration space computed around the optimal solution for NN, TRI, and PV interpolation (Fig. 6). These traces look very similar when a large parameter range is considered, but in the neighborhood of the registration solution, traces obtained with NN and TRI interpolation are noisy and show many

local maxima, while traces obtained with PV interpolation are almost quadratic around the optimum. Remark that the MI values obtained using TRI interpolation are larger than those obtained using NN or PV interpolation, which can be interpreted according to (2): The TRI averaging and noise reduction of the reference image intensities resulted in a larger reduction of the complexity of the joint histogram than the corresponding reduction in the complexity of the reference image histogram itself.

### B. Subsampling

The computational complexity of the MI criterion is proportional to the number of samples that is taken from the floating image to compute the joint histogram. Subsampling of the floating image can be applied to increase speed performance, as long as this does not deteriorate the optimization behavior. This was investigated for dataset A by registration of the subsampled MR image with the original CT image using PV interpolation. Subsampling was performed by taking samples on a regular grid at sample intervals of  $f_x, f_y$ , and  $f_z$  voxels in the  $x, y$ , and  $z$  direction, respectively, using NN interpolation. No averaging or smoothing of the MR image before subsampling was applied. We used  $f_x = f_y = 1, 2, 3$ , or 4, and  $f_z = 1, 2, 3$ , or 4. The same optimization strategy was used in each case. Registration solutions  $\alpha$  obtained using subsampling were compared with the solution  $\alpha^*$  found when no subsampling was applied (Fig. 7). For subsampling factors  $f = f_x \times f_y \times f_z$  up to 48 (four in the  $x$  and  $y$  direction, three

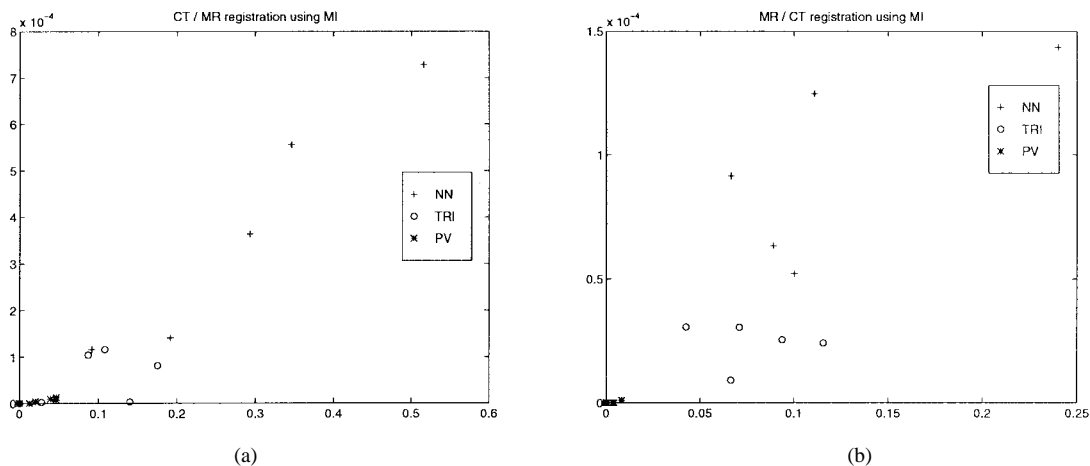


Fig. 5. Evaluation of the MI registration robustness for dataset A. Horizontal axis: norm of the difference vector  $|\alpha - \alpha^*|$  for different optimization strategies, using NN, TRI, and PV interpolation.  $\alpha^*$  corresponds to the registration solution with the best value for the registration criterion for each of the interpolation schemes applied. Vertical axis: difference in the registration criterion between each solution and the optimal one. (a) Using the CT image as the floating image. (b) Using the MR image as the floating image.

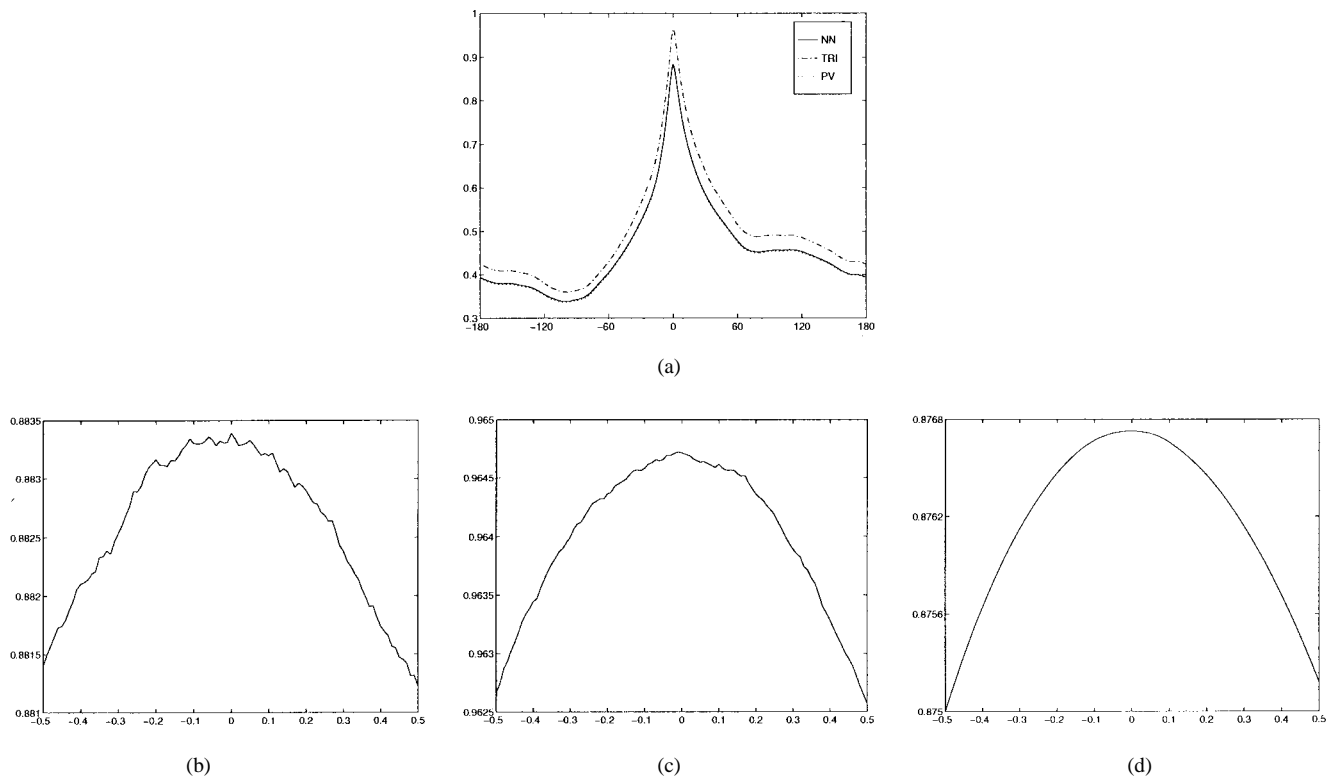


Fig. 6. MI traces around the optimal registration position for dataset A: Rotation around the  $x$  axis in the range from  $-180$  to  $+180^\circ$  (a) and from  $-0.5$  to  $+0.5^\circ$  (bottom row), using NN (b), TRI (c), and PV (d) interpolation.

in the  $z$  direction) the optimization converged in about 4 min to a solution less than  $0.2^\circ$  and  $0.2$  mm off from the solution found without subsampling.

### C. Partial Overlap

Clinically acquired images typically only partially overlap, as CT scanning is often confined to a specific region to minimize the radiation dose while MR protocols frequently image larger volumes. The influence of partial overlap on the

registration robustness was evaluated for dataset A for CT to MR registration using PV interpolation. The images were initially aligned as in the experiment in Section V and the same optimization strategy was applied, but only part of the CT data was considered when computing the MI criterion. More specifically, three 50-slice slabs were selected at the bottom (the skull basis), the middle, and the top part of the dataset. The results are summarized in Table IV and compared with the solution found using the full dataset by the mean and

TABLE IV  
INFLUENCE OF PARTIAL OVERLAP ON THE REGISTRATION ROBUSTNESS FOR CT TO MR REGISTRATION OF DATASET A

| ROI    | Slices | Rotation (degrees) |       |      | Translation (mm) |      |       | Difference (mm) |             |             |
|--------|--------|--------------------|-------|------|------------------|------|-------|-----------------|-------------|-------------|
|        |        | x                  | y     | z    | x                | y    | z     | x               | y           | z           |
| Full   | 0-99   | 10.36              | -3.17 | 2.09 | 6.94             | 1.15 | 18.20 |                 |             |             |
| Bottom | 0-49   | 10.14              | -2.91 | 2.03 | 6.67             | 1.30 | 19.46 | 0.28 (0.54)     | 0.21 (0.46) | 1.26 (1.78) |
| Middle | 25-74  | 9.46               | -2.53 | 2.13 | 6.67             | 0.71 | 17.75 | 0.43 (0.79)     | 0.62 (1.31) | 1.01 (2.14) |
| Top    | 50-99  | 9.74               | -3.05 | 2.43 | 6.86             | 0.82 | 17.59 | 0.35 (0.52)     | 0.52 (1.13) | 0.69 (1.46) |

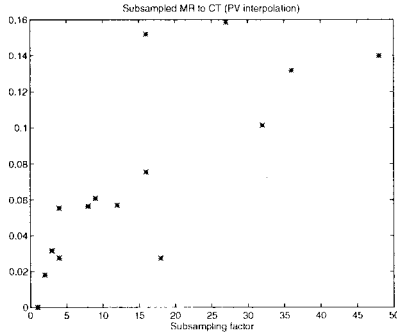


Fig. 7. Effect of subsampling the MR floating image of dataset A on the registration solution. Horizontal axis: subsampling factor  $f$ , indicating that only one out of  $f$  voxels was considered when evaluating the MI criterion. Vertical axis: norm of the difference vector  $|\alpha - \alpha^*|$ .  $\alpha^*$  corresponds to the registration solution obtained when no subsampling is applied.

maximal absolute difference evaluated over the full image at the same eight points as in Section V. The largest parameter differences occur for rotation around the  $x$  axis and translation in the  $z$  direction, resulting in maximal coordinate differences up to 1.5 CT voxel in the  $y$  and  $z$  direction, but on average all differences are subvoxel with respect to the CT voxel sizes.

#### D. Image Degradation

Various MR image degradation effects, such as noise, intensity inhomogeneity, and geometric distortion, alter the intensity distribution of the image which may affect the MI registration criterion. This was evaluated for the MR image of dataset D by comparing MI registration traces obtained for the original image and itself with similar traces obtained for the original image and its degraded version (Fig. 8). Such traces computed for translation in the  $x$  direction are shown in Fig. 9.

1) *Noise*: The original MR data ranges from 2 to 3359 with mean 160. White zero-mean Gaussian noise with variance of 50, 100, and 500 was superimposed onto the original image. Fig. 9(b) shows that increasing the noise level decreases the MI between the two images, but this does not affect the MI criterion, as the position of maximal MI in traces computed for all six registration parameters is not changed when the amount of noise is increased.

2) *Intensity Inhomogeneity*: To simulate the effect of MR intensity inhomogeneities on the registration criterion, the original MR image intensity  $I$  was altered into  $I'$  using a slice-by-slice planar quadratic inhomogeneity factor

$$\log I'(x, y) = \log I(x, y) + \Delta \log I(x, y) \quad (14)$$

$$\Delta \log I(x, y) = -k((x - x_c)^2 + (y - y_c)^2) \quad (15)$$

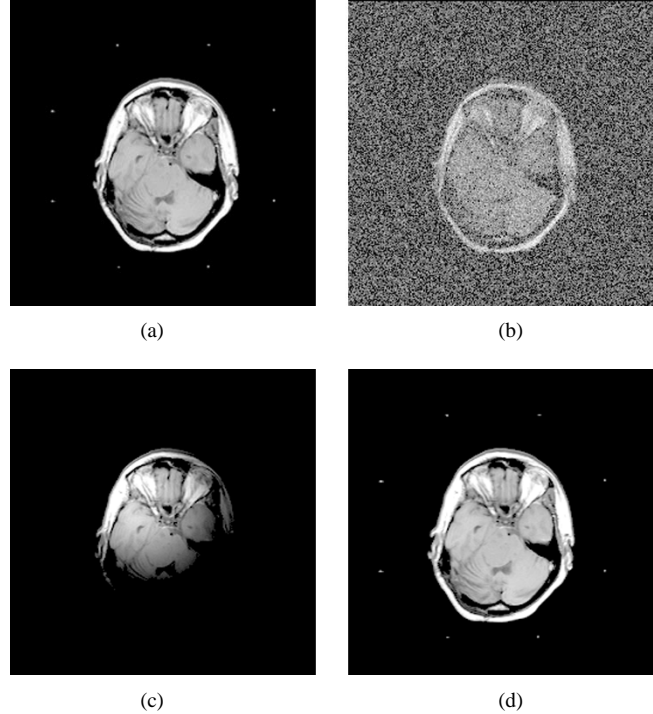


Fig. 8. (a) Slice 15 of the original MR image of dataset D, (b) zero mean noise added with variance of 500 grey-value units, (c) quadratic inhomogeneity ( $k = 0.004$ ), and (d) geometric distortion ( $k = 0.00075$ ).

with  $(x_c, y_c)$  being the image coordinates of the point around which the inhomogeneity is centered and  $k$  a scale factor. Fig. 9(c) shows MI traces for different values of  $k$  ( $k = 0.001, 0.002, 0.004; x_c = y_c = 100$ ). All traces for all parameters reach their maximum at the same position and the MI criterion is not affected by the presence of the inhomogeneity.

3) *Geometric Distortion*: Geometric distortions  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  were applied to the original MR image according to a slice-by-slice planar quadratic model of the magnetic field inhomogeneity [17]

$$\Delta x = k((x - x_c)^2 + (y - y_c)^2) \quad (16)$$

$$\Delta y = \Delta z = 0 \quad (17)$$

$$\Delta i(x, y) = |2k(x - x_c)| i(x + \Delta x, y + \Delta y) \quad (18)$$

with  $(x_c, y_c)$  the image coordinates of the center of each image plane and  $k$  a scale parameter. Fig. 9(d) shows traces of the registration criterion for various amounts of distortion ( $k = 0.0001, 0.0005, 0.00075$ ). As expected, the distortion shifts the optimum of the  $x$  translation parameter proportional to the average distortion  $\overline{\Delta x}$ . No such shift occurred for traces obtained for all other registration parameters.



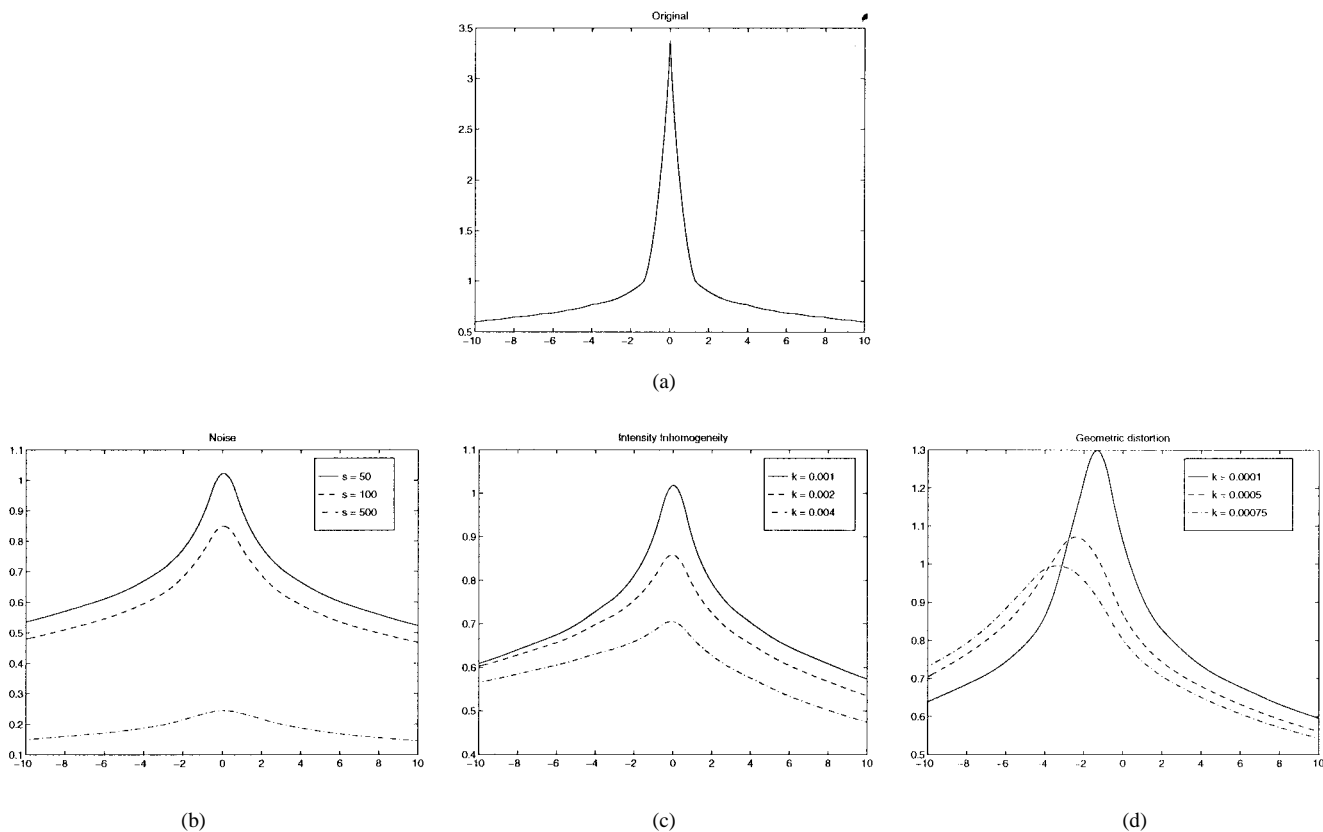


Fig. 9. MI traces using PV interpolation for translation in the  $x$  direction of the original MR image of dataset D over its degraded version in the range from  $-10$  to  $+10$  mm: (a) original, (b) noise, (c) intensity inhomogeneity, and (d) geometric distortion.

## VII. DISCUSSION

The MI registration criterion presented in this paper assumes that the statistical dependence between corresponding voxel intensities is maximal if both images are geometrically aligned. Because no assumptions are made regarding the nature of this dependence, the MI criterion is highly data independent and allows for robust and completely automatic registration of multimodality images in various applications with minimal tuning and without any prior segmentation or other preprocessing steps. The results of Section V demonstrate that subvoxel registration differences with respect to the stereotactic registration solution can be obtained for CT/MR and PET/MR matching without using any prior knowledge about the grey-value content of both images and the correspondence between them. Additional experiments on nine other datasets similar to dataset C within the Retrospective Registration Evaluation Project by Fitzpatrick *et al.* [10] have verified these results [29], [14]. Moreover, Section VI-C demonstrated the robustness of the method with respect to partial overlap, while it was shown in Section VI-D that large image degradations, such as noise and intensity inhomogeneities, have no significant influence on the MI registration criterion.

Estimations of the image intensity distributions were obtained by simple normalization of the joint histogram. In all experiments discussed in this paper, the joint histogram was computed from the entire overlapping part of both images, using the original image data and a fixed number of bins of

$256 \times 256$ . We have not evaluated the influence of the bin size, the choice of a ROI, or the application of nonlinear image intensity transformations on the behavior of the MI registration criterion. Other schemes can be used to estimate the image intensity distributions, for instance by using Parzen windowing [9] on a set of samples taken from the overlapping part of both images. This approach was used by Viola *et al.* [27], who also use stochastic sampling of the floating image to increase speed performance.

PV interpolation was introduced to make the joint and marginal distributions and their MI vary smoothly for small changes in the registration parameters. The results of Section VI-A indicate that PV interpolation indeed improves optimization robustness compared to NN and TRI interpolation. More experiments are needed to compare this approach to the Parzen windowing method as used by Viola *et al.* [27] and the multiresolution cubic resampling approach as used by Studholme *et al.* [20].

The optimization of the MI registration criterion is performed using Powell's method. We noticed that for low-resolution images the initial order in which the parameters are optimized strongly influences optimization robustness. Generally, we obtained the best results when first optimizing the in-plane parameters  $t_x$ ,  $t_y$  and  $\phi_z$ , before optimizing the out-of-plane parameters  $\phi_x$ ,  $\phi_y$ , and  $t_z$ . For low-resolution images, the optimization often did not converge to the global optimum if a different parameter order was specified, due to the occurrence of local optima especially for the  $x$ -rotation and

the  $z$ -translation parameters. In the experiments discussed in this paper the amount of misregistration that was recovered was as large as  $10^\circ$  and 40 mm, but we have not extensively investigated the robustness of the method with respect to the initial positioning of the images, for instance by using multiple randomised starting estimates. The choice of the floating image may also influence the behavior of the registration criterion. In the experiment of Section VI-A, MR to CT matching was found to be more robust than CT to MR matching. However, it is not clear whether this was caused by sampling and interpolation issues or by the fact that the MR image is more complex than the CT image and that the spatial correlation of image intensity values is higher in the CT image than in the MR image.

We have not tuned the design of the search strategy toward specific applications. For instance, the number of criterion evaluations required may be decreased by taking the limited image resolution into account when determining convergence. Moreover, the results of Section VI-B demonstrate that for high-resolution images subsampling of the floating image can be applied without deteriorating optimization robustness. Important speed-ups can, thus, be realized by using a multiresolution optimization strategy, starting with a coarsely sampled image for efficiency and increasing the resolution as the optimization proceeds for accuracy [20]. Furthermore, the smooth behavior of the MI criterion, especially when using PV interpolation, may be exploited by using gradient-based optimization methods, as explicit formulas for the derivatives of the MI function with respect to the registration parameters can be obtained [27].

All the experiments discussed in this paper were for rigid-body registration of CT, MR, and PET images of the brain of the same patient. However, it is clear that the MI criterion can equally well be applied to other applications, using more general geometric transformations. We have used the same method successfully for patient-to-patient matching of MR brain images for correlation of functional MR data and for the registration of CT images of a hardware phantom to its geometrical description to assess the accuracy of spiral CT imaging [14].

MI measures statistical dependence by comparing the complexity of the joint distribution with that of the marginals. Both marginal distributions are taken into account explicitly, which is an important difference with the measures proposed by Hill *et al.* [13] (third-order moment of the joint histogram) and Collignon *et al.* [6] (entropy of the joint histogram), which focus on the joint histogram only. In Appendixes A and B we discuss the relationship of these criteria and of the measure of Woods *et al.* [30] (variance of intensity ratios) to the MI criterion.

MI is only one of a family of measures of statistical dependence or information redundancy (see Appendix C). We have experimented with  $\rho(A, B) = H(A, B) - I(A, B)$ , which can be shown to be a metric [8], and  $ECC(A, B) = 2I(A, B)/(H(A) + H(B))$ , the *entropy correlation coefficient* [1]. In some cases these measures performed better than the original MI criterion, but we could not establish a clear preference for either of these. Furthermore, the use of MI

for multimodality image registration is not restricted to the original image intensities only: other derived features such as edges or ridges can be used as well. Selection of appropriate features is an area for further research.

## VIII. CONCLUSION

The MI registration criterion presented in this paper allows for subvoxel accurate, highly robust, and completely automatic registration of multimodality medical images. Because the method is largely data independent and requires no user interaction or preprocessing, the method is well suited to be used in clinical practice.

Further research is needed to better understand the influence of implementation issues, such as sampling and interpolation, on the registration criterion. Furthermore, the performance of the registration method on clinical data can be improved by tuning the optimization method to specific applications, while alternative search strategies, including multiresolution and gradient-based methods, have to be investigated. Finally, other registration criteria can be derived from the one presented here, using alternative information measures applied on different features.

## APPENDIX A

We show the relationship between the multimodality registration criterion devised by Hill *et al.* [12] and the joint entropy  $H(a, b)$ . Hill *et al.* used the  $n$ th-order moment of the scatter-plot  $h$  as a measure of dispersion

$$T_n = \sum_{a,b} \left( \frac{h(a,b)}{V} \right)^n \quad (19)$$

with  $h(a, b)$  the histogram entries and  $V = \sum_{a,b} h(a, b)$  the common volume of overlap. Approximating the joint probability distribution  $p(a, b)$  by  $p(a, b) = h(a, b)/V$ , we get

$$T_n = \sum_{a,b} p(a, b)^n.$$

It turns out that  $T_n$  is one-to-one related to the joint Rényi entropy  $H_n$  of order  $n$  [22]

$$H_n = \frac{1}{1-n} \log(T_n)$$

with the following properties.

- 1)  $\lim_{n \rightarrow 1} H_n(p) = -\sum_i p_i \log p_i$ , which is the Shannon entropy.
- 2)  $n_2 > n_1 \rightarrow H_{n_2}(p) \leq H_{n_1}(p)$ .

Hence, the normalized second- or third-order moment criteria defined by Hill *et al.* are equivalent to a generalized version of the joint entropy  $H(a, b)$ .

## APPENDIX B

We show how the multimodality registration criterion devised by Woods *et al.* [30] relates to the conditional entropy  $H(a | b)$ . Denote by  $A$  and  $B$  the set of possible intensities in the two images. Denote by  $a_i$  and  $b_i$  the intensities of  $A$  and  $B$  at the common voxel position  $i$ . For each voxel  $i$  with

value  $b_i = b$  in image  $B$ , let  $a_i(b)$  be the value at voxel  $i$  in the corresponding image  $A$ . Let  $\mu_a(b)$  be the mean and  $\sigma_a(b)$  be the standard deviation of the set  $\{a_i(b) \mid \forall i : b_i = b\}$ . Let  $n_b = \#\{i \mid b_i = b\}$  and  $N = \sum_b n_b$ . The registration criterion that Woods *et al.* minimize is then defined as follows:

$$\sigma'' = \sum_b \frac{n_b \sigma_a(b)}{N \mu_a(b)} \quad (20)$$

$$= \sum_b p_b(b) \frac{\sigma_a(b)}{\mu_a(b)} \quad (21)$$

with  $p_b$  the marginal distribution function of image intensities  $B$ .

It can be shown [8] that for a given mean  $\mu_a(b)$  and standard deviation  $\sigma_a(b)$

$$H(A \mid B) = \sum_b p(b) H(A \mid B = b) \quad (22)$$

$$= - \sum_b p(b) \sum_a p(a \mid b) \cdot \log p(a \mid b) \quad (23)$$

$$\leq \sum_b p(b) \log(\sigma_a(b)) + \frac{1}{2} \log(2\pi e) \quad (24)$$

with equality if the conditional distribution  $p(a \mid b)$  of image intensities  $A$  given  $B$  is the normal distribution  $N(\mu_a(b), \sigma_a(b))$ .

Using Jensen's inequality for concave functions [8] we get

$$H(A \mid B) \leq \sum_b p(b) \log \left( \frac{\sigma_a(b)}{\mu_a(b)} \right) + \sum_b p(b) \log(\mu_a(b)) \quad (25)$$

$$\leq \log \left( \sum_b p(b) \frac{\sigma_a(b)}{\mu_a(b)} \right) + \log \left( \sum_b p(b) \mu_a(b) \right) \quad (26)$$

$$= \log(\sigma'') + \log(\mu(a)) \quad (27)$$

with  $\mu(a) = \sum_b p(b) \mu_a(b)$  the mean intensity of image  $A$ .

If  $\mu(a)$  is constant and  $p(a \mid b)$  can be assumed to be normally distributed, minimization of  $\sigma''$  then amounts to optimizing the conditional entropy  $H(A \mid B)$ . In the approach of Woods, this assumption is approximately accomplished by editing away parts in one dataset (namely the skin in MR) for which otherwise additional modes might occur in  $p(a \mid b)$ , while Hill *et al.* have proposed to take only specifically selected regions in the joint histogram into account.

## APPENDIX C

MI  $I(A, B)$  is only one example of the more general  $f$ -information measures of dependence  $f(P \parallel P_1 \times P_2)$  [22] with  $P$  the set of joint probability distributions  $P(A, B)$  and  $P_1 \times P_2$  the set of joint probability distributions  $P(A) \cdot P(B)$  assuming  $A$  and  $B$  to be independent.

$f$ -information is derived from the concept of  $f$ -divergence, which is defined as

$$f(P \parallel Q) = \sum_i q_i \cdot f(p_i/q_i)$$

with  $P = \{p_1, p_2, \dots\}$  and  $Q = \{q_1, q_2, \dots\}$  with suitable definitions when  $q_i = 0$ .

Some examples of  $f$ -divergence are:

- $I_\alpha$ -divergence

$$I_\alpha = \frac{1}{\alpha(\alpha-1)} \left[ \sum_i \frac{p_i^\alpha}{q_i^{\alpha-1}} - 1 \right]$$

- $\chi^2$ -divergence

$$\chi^2 = \sum_i \frac{(p_i - q_i)^2}{q_i}$$

with corresponding  $f$ -informations

- $I_\alpha$ -information

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

with  $p_{ij} = P(i, j)$  and  $p_i = \sum_j p_{ij}$  and  $p_j = \sum_i p_{ij}$

- $\chi^2$ -information

$$\chi^2(P \parallel P_1 \times P_2) = \sum_{i,j} \frac{(p_{ij} - p_i \cdot p_j)^2}{p_i \cdot p_j}$$

Note that  $I_\alpha(P \parallel P_1 \times P_2)$  is the information-measure counterpart of the  $n$ th-order moment used by Hill *et al.* for  $n = \alpha = 2, 3$ . Furthermore,  $I_1(P \parallel P_1 \times P_2) = \sum_{i,j} p_{ij} \log \left( \frac{p_{ij}}{p_i \cdot p_j} \right)$  which is the definition of MI used in this paper.

## REFERENCES

- [1] J. Astola and I. Virtanen, "Entropy correlation coefficient, a measure of statistical dependence for categorized data," in *Proc. Univ. Vaasa, Discussion Papers*, Finland, 1982, no. 44.
- [2] J. A. Baddeley, "An error metric for binary images," in *Proc. IEEE Workshop on Robust Computer Vision*, Bonn, 1992, pp. 59–78.
- [3] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, Dec. 1992.
- [4] C-H. Chen, *Statistical Pattern Recognition*. Rochelle Park, N.J.: Spartan, Hayden, 1973.
- [5] J. Y. Chiang and B. J. Sullivan, "Coincident bit counting—A new criterion for image registration," *IEEE Trans. Med. Imag.*, vol. 12, no. 1, pp. 30–38, Mar. 1993.
- [6] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal, "3D multimodality medical image registration using feature space clustering," in *Proc. 1st Int. Conf. Computer Vision, Virtual Reality and Robotics in Medicine; Lecture Notes in Computer Science 905*, N. Ayache, Ed. New York: Springer-Verlag, Apr. 1995, pp. 195–204.
- [7] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multimodality medical image registration using information theory," in *Proc. 14th Int. Conf. Information Processing in Medical Imaging; Computational Imaging and Vision 3*, Y. Bizais, C. Barillot, and R. Di Paola, Eds. Boston: Kluwer, June 1995, pp. 263–274.
- [8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [9] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [10] J. M. Fitzpatrick, "Evaluation of retrospective image registration," Vanderbilt Univ., Nashville, TN, National Institutes of Health, Project Number 1 R01 NS33926-01, 1994.

- [11] P. Gerlot-Chiron and Y. Bizais, "Registration of multimodality medical images using region overlap criterion," *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 5, pp. 396–406, Sept. 1992.
- [12] D. L. G. Hill, D. J. Hawkes, N. A. Harrison, and C. F. Ruff, "A strategy for automated multimodality image registration incorporating anatomical knowledge and imager characteristics," in *Proc. 13th Int. Conf. Information Processing in Medical Imaging; Lecture Notes in Computer Science 687*, H. H. Barrett and A. F. Gmitro, Eds. New York: Springer-Verlag, June 1993, pp. 182–196.
- [13] D. L. G. Hill, C. Studholme, and D. J. Hawkes, "Voxel similarity measures for automated image registration," in *Proc. Visualization in Biomedical Computing 1994*, SPIE, 1994, vol. 2359, pp. 205–216.
- [14] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multi-modality image registration by maximization of mutual information," in *Proc. IEEE Workshop Mathematical Methods in Biomedical Image Analysis*, June 1996, pp. 14–22.
- [15] J. B. A. Maintz, P. A. van den Elsen, and M. A. Viergever, "Comparison of feature-based matching of CT and MR brain images," in *Proc. 1st Int. Conf. Computer Vision, Virtual Reality and Robotics in Medicine; Lecture Notes in Computer Science 905*, N. Ayache, Ed. New York: Springer-Verlag, Apr. 1995, pp. 219–228.
- [16] C. R. Maurer and J. M. Fitzpatrick, "A review of medical image registration," in *Interactive Image-Guided Neurosurgery*, R. J. Maciunas, Ed. Park Ridge, IL: Amer. Association of Neurological Surgeons, 1993, pp. 17–44.
- [17] J. Michiels, P. Pelgrims, H. Bosmans, D. Vandermeulen, J. Gybels, G. Marchal, and P. Suetens, "On the problem of geometric distortion in magnetic resonance images for stereotactic neurosurgery," *Magn. Reson. Imag.*, vol. 12, no. 5, pp. 749–765, 1994.
- [18] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*, 2nd ed. Cambridge, U. K.: Cambridge Univ. Press, 1992, ch. 10, pp. 412–419.
- [19] T. Radcliffe, R. Rajapakshe, and S. Shalev, "Pseudocorrelation: A fast, robust, absolute, grey-level image alignment algorithm," *Med. Phys.*, vol. 21, no. 6, pp. 761–769, June 1994.
- [20] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Multiresolution voxel similarity measures for MR-PET registration," in *Proc. 14th Int. Conf. Information Processing in Medical Imaging; Computational Imaging and Vision 3*, Y. Bizais, C. Barillot, and R. Di Paola, Eds. Boston: Kluwer, June 1995, pp. 287–298.
- [21] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Automated 3D registration of truncated MR and CT images of the head," in *Proc. British Machine Vision Conf.*, 1995, pp. 27–36.
- [22] I. Vajda, *Theory of Statistical Inference and Information*. Dordrecht, The Netherlands: Kluwer, 1989.
- [23] P. A. van den Elsen, E.-J. D. Pol, and M. A. Viergever, "Medical image matching—A review with classification," *IEEE Eng. Med. Biol.*, pp. 26–38, Mar. 1993.
- [24] P. A. van den Elsen, J. B. A. Maintz, E.-J. D. Pol, and M. A. Viergever, "Automatic registration of CT and MR brain images using correlation of geometrical features," *IEEE Trans. Med. Imag.*, vol. 14, no. 2, June 1995.
- [25] P. A. van den Elsen, E.-J. D. Pol, T. S. Sumanaweera, P. F. Hemler, S. Napel, and J. Adler, "Grey value correlation techniques used for automatic matching of CT and MR brain and spine images," in *Proc. Visualization in Biomedical Computing*, Oct. 1994, vol. 2359, pp. 227–237.
- [26] A. Venot, J. F. Lebruchec, and J. C. Roucaeyrol, "A new class of similarity measures for robust image registration," *Comput. Vision, Graphics, Image Processing*, vol. 28, no. 2, pp. 176–184, Nov. 1984.
- [27] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," in *Proc. 5th Int. Conf. Computer Vision*, June 1995, pp. 16–23.
- [28] 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, Mar. 1996.
- [29] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, Jr., R. M. Kessler, and R. J. Maciunas, *et al.*, "Comparison and evaluation of retrospective intermodality image registration techniques," in *Proc. Image Processing*, Feb. 1996, vol. 2710, pp. 332–347.
- [30] R. P. Woods, J. C. Mazziotta, and S. R. Cherry, "MRI-PET registration with automated algorithm," *J. Comput. Assist. Tomogr.*, vol. 17, no. 4, pp. 536–546, July/Aug. 1993.