Multi-modality Image Registration By Maximization of Mutual Information

Frederik Maes, André Collignon, Dirk Vandermeulen, Guy Marchal, Paul Suetens
KU Leuven, Laboratory for Medical Imaging Research
Radiologie-ESAT, UZ Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium
Frederik.Maes@uz.kuleuven.ac.be

Abstract

Mutual information of image intensities has been proposed as a new matching criterion for automated multi-modality image registration. In this paper, we give experimental evidence of the power and the generality of the mutual information criterion by showing results for various applications involving CT, MR and PET images. Our results illustrate the large applicability of the approach and demonstrate its high suitability for routine use in clinical practice.

1. Introduction

Voxel similarity based (VSB) registration methods have recently captured a lot of interest for multi-modality image registration and have been shown to allow for robust retrospective registration by an automated algorithm without interaction or segmentation requirement [21]. VSB methods optimize a functional measuring the similarity of all geometrically corresponding voxel pairs for some feature. Feature calculation is straightforward or even absent when only grey-values are used. The registration accuracy is not limited by segmentation errors as with surface based registration methods and does not depend on the accurate indication of corresponding landmarks as with point landmark based methods.

For intra-modality registration multiple VSB methods have been proposed that rely on the assumption that the intensities of the two images are linearly correlated, which is generally not satisfied in the case of inter-modality registration. Cross-correlation of feature images derived from the original image data has been applied to CT/MR matching using geometrical features such as edges [13] and ridges [18] or using especially designed intensity transformations [19]. But feature extraction may introduce geometrical errors and requires extra calculation time, while 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 non-edge or non-ridge voxels correlate equally well. A multi-resolution optimization strategy is therefore required.

In the approach of Woods et al. [22] and Hill et al. [8, 9] 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 propose are largely heuristic and require segmentation of the images or delineation of specific histogram regions to make the method work [15]. Moreover, 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 multi-modality combinations (PET/MR).

The use of the much more general notion of Mutual Information (MI) or relative entropy [4, 16] to describe the dispersive behavior of the 2-D histogram has been proposed independently by Collignon et al. [3, 11] and by Viola et al. [20]. Mutual information 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 states that the mutual information of the image intensity values of corresponding voxel pairs is maximal if the images are geometrically aligned. Because no limiting constraints are imposed on the nature of the relation between the intensities in the images to be registered and no assumptions are made regarding the image content of the modalities involved, the mutual information criterion is very general and powerful. It allows for robust and completely automated registration of multi-modal images without prior segmentation, feature extraction or other pre-processing steps, which makes this method very well suited for clinical applications.

In this paper we give experimental evidence for the power of the mutual information approach by showing re-
results for various clinical applications. In sections 2 and 3, we first describe the algorithm and discuss some implementation issues. Section 4.1 presents results for CT/MR and PET/MR registration of brain images, validating the sub-voxel accuracy of the method. The same method has been used successfully for the rigid body registration of MR brain images of different patients to correlate functional MRI data (section 4.2) and for the registration of spiral CT images of a hardware phantom to its geometrical description to assess the accuracy of spiral CT imaging (section 4.3). Section 5 discusses our current findings, while section 6 gives some directions for further work.

2. The mutual information matching criterion

The image intensity values \( a \) and \( b \) of a pair of corresponding voxels in the two images that are to be registered can be considered to be random variables \( A \) and \( B \) respectively. \( A \) and \( B \) are related through the geometric transformation \( T_\alpha \) defined by the set of registration parameters \( \alpha \). 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 image intensity histograms of the overlapping parts of both images.

Mutual information (MI) \( I(A,B) \) of two random variables \( A \) and \( B \) measures the degree of dependence between \( A \) and \( B \) by the Kullback-Leibler distance [16] between the joint distribution \( p_{AB}(a,b) \) and the distribution associated to the case of complete independence \( p_A(a) \cdot p_B(b) \):

\[
I(A,B) = \sum_{a,b} p_{AB}(a,b) \log \frac{p_{AB}(a,b)}{p_A(a) \cdot p_B(b)} \tag{1}
\]

\[
= H(A) - H(A|B) \tag{2}
\]

\[
= H(B) - H(B|A) \tag{3}
\]

\[
= H(A) + H(B) - H(A,B) \tag{4}
\]

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 \( H(B|A) \) the conditional entropy of \( A \) given \( B \) and of \( B \) given \( A \) respectively. 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, \( I(A,B) \) 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 \( A \). If \( A \) and \( B \) are statistically independent, then \( p_{AB}(a,b) = p_A(a) \cdot p_B(b) \) and \( I(A,B) = 0 \), while if \( A \) and \( B \) are maximally dependent, they are related by a one-to-one mapping \( T \) such that \( p_A(a) = p_B(T(a)) = p_{AB}(a,T(a)) \) and \( I(A,B) = H(A) - H(B) = H(A,B) \).

The mutual information registration criterion states that the images are geometrically aligned by the transformation \( T_\alpha \), for which \( I(A,B) \) is maximal. This criterion assumes that the amount of information that \( A \) contains about \( B \) is maximal in the registered position. When applied to images of the brain, for instance, the skull is high intense in CT images and low intense in MR images. If both images are correctly aligned, the uncertainty about the MR voxel intensity is therefore largely reduced if the corresponding CT voxel is known to be high intense and thus likely to be part of the skull. This correspondence is lost in case of misregistration. However, the mutual information criterion does not rely on limiting assumptions regarding the nature of the relation between corresponding voxel intensities in different modalities, which is highly data-dependent, and no constraints are imposed on the image content of the modalities involved. From equation 4, the mutual information matching criterion can be interpreted as follows [20]: “maximizing mutual information 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)”.

3. Implementation issues

Let \( (f(s), r(T_\alpha s)) \) be a pair of image intensity values in the images \( F \) and \( R \) to be registered at corresponding sites \( s \) in \( F \) and \( T_\alpha s \) in \( R \) with \( T_\alpha \) a geometric transformation with parameters \( \alpha \). Let \( p_F(f), p_R(r) \) and \( p_{FR}(f,r) \) be the marginal and joint probability distributions of \( f(s) \) and \( r(T_\alpha s) \). The mutual information registration criterion can then be summarized by the following equations:

\[
I(\alpha) = \sum_{f,r} p_{FR}(f,r) \log \frac{p_{FR}(f,r)}{p_F(f) \cdot p_R(r)} \tag{5}
\]

\[
\alpha^* = \arg \max_\alpha I(\alpha) \tag{6}
\]

with \( I(\alpha) \) the mutual information of \( F \) and \( R \) at registration position \( \alpha \) and \( \alpha^* \) the position at which \( I(\alpha) \) is maximal.

One of the images is selected to be the floating image \( F \) from which samples \( s \in S \) with intensity \( f(s) \) are taken and transformed into the reference image \( R \). The set of sample points \( S \) usually is the set of grid points of \( F \), although subsampling of the floating image can be used to increase speed performance. Each sample taken from \( F \) is transformed into \( R \) by the transformation \( T_\alpha \). 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 the reference image \( R \).

The joint image intensity histogram \( h(f, r) \) is computed by binning the image intensity pairs \( (f(s), r(T_\alpha s)) \). In or-
order to do this efficiently, the floating and the reference image 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 $n_F = n_R = 256$. Normalization of $h(f, r)$ yields estimations for the marginal and joint image intensity distributions $p_F(f), p_R(r)$ and $p_{FR}(f, r)$ of the overlapping volume of the images:

$$p_{FR}(f, r) = \frac{h(f, r)}{\sum_f h(f, r)} \quad (7)$$

$$p_F(f) = \sum_r p_{FR}(f, r) \quad (8)$$

$$p_R(r) = \sum_f p_{FR}(f, r) \quad (9)$$

In general, $T_\alpha s$ will not coincide with a grid point of $R$ and interpolation of the reference image is needed to obtain the image intensity value $r(T_\alpha s)$ at the position $T_\alpha s$ in the reference image. Nearest neighbor (NN) interpolation of $R$ is generally insufficient to guarantee subvoxel accuracy, as it is insensitive to translations up to 1 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 $p_R(r)$ for small variations of $\alpha$. To avoid this problem, we have proposed 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 $R$, the contribution of the image intensity $f(s)$ of the sample $s$ of $F$ to the joint histogram is distributed over the intensity values of all 8 nearest neighbors of $T_\alpha s$ on the grid of $R$, using the same weights as for trilinear interpolation. Each entry in the joint histogram is then the sum of smoothly varying fractions of 1, such that the histogram changes smoothly as $\alpha$ is varied.

The optimal registration parameters $\alpha^*$ are found by maximization of $I(\alpha)$ using Powell’s direction set method [14]. The images are positioned initially such that their centers coincide and that the floating and reference image coordinate axes corresponding to the same patient axis (i.e., right to left, anterior to posterior, and inferior to superior) are properly aligned, which assumes that the orientation of each of the images with respect to the patient is known.

The algorithm has been implemented\(^1\) on an IBM RS6000/3AT workstation (58 MHz, 185 SPECfp92, AIX 4.1.3). The complexity of one evaluation of the MI registration criterion was found to vary linearly with the number of samples taken from the floating image. While trilinear or partial volume interpolation have about the same complexity, nearest neighbor interpolation is about twice as efficient (1.36 vs. 0.60 CPU seconds per million samples for one evaluation of the criterion). The number of evaluations performed during optimization depends on the initial orientation of the images and on the convergence parameters specified for the Powell algorithm. All experiments described in this paper typically required between 300 and 500 criterion evaluations.

4. Results

In this section, we show registration results obtained by using the mutual information criterion for various applications. These include the registration of CT, MR and PET brain images of the same patient, the registration of MR brain images of different patients and the registration of a mathematical model of a hardware phantom to its CT image. For all experiments discussed in this paper, we have restricted the geometric transformation $T_\alpha$ to rigid body transformations only, although it is clear that the mutual information criterion can be applied to more general transformations as well. The rigid body transformation $T_\alpha$ is a superposition of a 3-D rotation and a 3-D translation and the registration parameter $\alpha$ is a 6-component vector consisting of 3 rotation angles and 3 translation distances.

4.1. Registration of CT, MR and PET brain images

The performance of the MI registration criterion has been evaluated extensively 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 abstraction is made of scanner calibration problems and problems of geometric distortions, both of which can be minimized by careful calibration and scan parameter selection respectively.

The accuracy of the matching criterion was validated within the framework of the Retrospective Registration Evaluation Project (RREP) conducted by Fitzpatrick et al. at Vanderbilt University [6]. In total, 76 CT/MR and PET/MR registrations were performed on 9 patient datasets provided by Vanderbilt. Each dataset consisted of axial CT, MR and PET images which were acquired stereotactically using the same protocol for each patient, but the images were edited to remove the stereotactic markers (figure 1). 6 different MR images were acquired: proton density (PD), T1- and T2-weighted images and the corresponding images corrected for geometric distortion (PDr, T1r, T2r) [21]. The slice thickness was 4 mm for the CT and MR images and 8 mm for the PET images (table 1).

Registration of these images using the mutual information criterion was performed fully automatically on an IBM

\(1\) Object-oriented C++ source code available on request.
Figure 1. a, b, c: Typical axial slices of the CT, MR and PET images used in the experiments of section 4.1. d: The bounding box of the innermost eighth of the floating image defines 8 points near the brain surface at which registration accuracy is evaluated.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Voxelsizes (in mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR</td>
<td>256×256×25, 1.25×1.25×4.00</td>
</tr>
<tr>
<td>CT</td>
<td>512×512×30, 0.65×0.65×4.00</td>
</tr>
<tr>
<td>PET</td>
<td>128×128×15, 2.59×2.59×8.00</td>
</tr>
</tbody>
</table>

Table 1. Typical dimensions and voxelsizes of the images used in the experiments of section 4.1.

RS6000/3AT workstation using PV interpolation in about 20 minutes for MR to CT matching and in only about 2 minutes for PET to MR matching. The recovered rotational transformation parameters were generally smaller than 5 degrees, while the translational parameters varied up to 30 millimeter.

For each registration result separately, the difference between the stereotactic reference transformation and the registration solution as obtained with the mutual information matching criterion was evaluated by considering 8 points near the brain surface in the CT or PET image (defined as the 8 cornerpoints of the bounding box of the innermost eighth of the image, figure 1d), transforming these into the MR image using either of both transforms and taking the mean over all 8 points of the norm of the transformation difference vector (∥Δ∥) and of the absolute value of its three coordinate components separately (|Δx|, |Δy|, |Δz|). These error values are summarized in table 2 and in table 3 by the mean and the maximal value respectively over all patient datasets. The x, y and z direction correspond to the patient’s right to left, anterior to posterior and inferior to superior direction respectively.

The mean errors are all subvoxel accurate for each of the modality combinations with respect to the largest voxel-size of the images. The maximal error values show that errors are subvoxel for almost all 41 CT/MR and 35 PET/MR

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>∆x</th>
<th>∆y</th>
<th>∆z</th>
<th>∆</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT/PD</td>
<td>7</td>
<td>0.5301</td>
<td>1.3527</td>
<td>1.2952</td>
<td>2.2041</td>
</tr>
<tr>
<td>CT/T1</td>
<td>7</td>
<td>0.5104</td>
<td>1.2746</td>
<td>1.5321</td>
<td>2.2577</td>
</tr>
<tr>
<td>CT/T2</td>
<td>7</td>
<td>0.5808</td>
<td>1.1779</td>
<td>1.4569</td>
<td>2.1919</td>
</tr>
<tr>
<td>CT/PDr</td>
<td>7</td>
<td>0.3588</td>
<td>0.3748</td>
<td>0.9360</td>
<td>1.3004</td>
</tr>
<tr>
<td>CT/T1r</td>
<td>6</td>
<td>0.2977</td>
<td>0.3633</td>
<td>1.1708</td>
<td>1.4662</td>
</tr>
<tr>
<td>CT/T2r</td>
<td>7</td>
<td>0.3760</td>
<td>0.6746</td>
<td>1.3699</td>
<td>1.7496</td>
</tr>
<tr>
<td>PET/PD</td>
<td>7</td>
<td>1.5232</td>
<td>2.8007</td>
<td>3.4607</td>
<td>5.1995</td>
</tr>
<tr>
<td>PET/T1</td>
<td>7</td>
<td>0.9105</td>
<td>1.9351</td>
<td>4.6114</td>
<td>5.5750</td>
</tr>
<tr>
<td>PET/T2</td>
<td>7</td>
<td>0.8930</td>
<td>1.9682</td>
<td>2.7568</td>
<td>3.9311</td>
</tr>
<tr>
<td>PET/PDr</td>
<td>5</td>
<td>1.1752</td>
<td>1.7516</td>
<td>2.7908</td>
<td>3.9214</td>
</tr>
<tr>
<td>PET/T1r</td>
<td>4</td>
<td>0.9002</td>
<td>1.0050</td>
<td>2.4337</td>
<td>3.1128</td>
</tr>
<tr>
<td>PET/T2r</td>
<td>5</td>
<td>0.9115</td>
<td>1.7599</td>
<td>2.6957</td>
<td>3.7094</td>
</tr>
</tbody>
</table>

Table 2. Registration error (in mm) evaluated at 8 points near the brain surface: mean values over all N cases.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>∆x</th>
<th>∆y</th>
<th>∆z</th>
<th>∆</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT/PD</td>
<td>7</td>
<td>1.1894</td>
<td>2.7498</td>
<td>2.8421</td>
<td>3.2472</td>
</tr>
<tr>
<td>CT/T1</td>
<td>7</td>
<td>0.8425</td>
<td>2.7716</td>
<td>3.5301</td>
<td>4.7798</td>
</tr>
<tr>
<td>CT/T2</td>
<td>7</td>
<td>1.4095</td>
<td>2.2235</td>
<td>2.7453</td>
<td>3.5607</td>
</tr>
<tr>
<td>CT/PDr</td>
<td>7</td>
<td>0.6449</td>
<td>0.8322</td>
<td>1.4986</td>
<td>1.9723</td>
</tr>
<tr>
<td>CT/T1r</td>
<td>6</td>
<td>0.4347</td>
<td>1.0065</td>
<td>2.6635</td>
<td>2.7929</td>
</tr>
<tr>
<td>CT/T2r</td>
<td>7</td>
<td>0.7475</td>
<td>1.5521</td>
<td>3.8785</td>
<td>4.5174</td>
</tr>
<tr>
<td>PET/PD</td>
<td>7</td>
<td>2.4587</td>
<td>6.2855</td>
<td>6.6968</td>
<td>10.1254</td>
</tr>
<tr>
<td>PET/T1</td>
<td>7</td>
<td>1.6536</td>
<td>3.3762</td>
<td>11.0182</td>
<td>11.1916</td>
</tr>
<tr>
<td>PET/T2</td>
<td>7</td>
<td>1.6703</td>
<td>4.7235</td>
<td>4.4445</td>
<td>6.2122</td>
</tr>
<tr>
<td>PET/PDr</td>
<td>5</td>
<td>1.4953</td>
<td>2.6593</td>
<td>4.6017</td>
<td>5.7440</td>
</tr>
<tr>
<td>PET/T1r</td>
<td>4</td>
<td>1.5704</td>
<td>1.9126</td>
<td>2.9016</td>
<td>4.0556</td>
</tr>
<tr>
<td>PET/T2r</td>
<td>5</td>
<td>1.7895</td>
<td>4.3078</td>
<td>5.4507</td>
<td>7.9527</td>
</tr>
</tbody>
</table>

Table 3. Registration error (in mm) evaluated at 8 points near the brain surface: maximal values over all N cases.
registrations, while in at least 2 cases for PET to MR registration errors were clearly larger than a PET voxel. The largest errors occur in the z direction, while we also noted that the optimization proved to be more robust if the in-plane parameters $t_x$, $t_y$ and $\phi_z$ were optimized first, before optimizing the out-of-plane parameters $\phi_x$, $\phi_y$ and $t_z$. Both observations can be explained by the lower resolution of the images in the z direction and by the fact that the registration is less well conditioned for translation in the cranio-caudal direction as these images do not include the top of the skull. When comparing the results for the MR images that were corrected for geometric distortion with those for the uncorrected images, there is a clear tendency towards lower registration errors for the corrected images [21].

4.2. Patient-to-patient MR registration

Correlation of functional MRI data obtained from different patients was done by registration of the corresponding anatomical MR images with a template image of the brain, obtained by segmentation of an original MR image\textsuperscript{2}. The template has been mapped into Talairach space, such that after registration of each of the patient images to the template, functional measurements from different patients can be compared within the Talairach frame.

The template has 2 mm cubic voxels and size $64 \times 87 \times 64$. The patient MR MPRAGE images have a voxelsize of $1 \times 1 \times 1.25$ mm and dimensions $256 \times 256 \times 128$. Registration was done from the template to the patient image using a rigid body transformation, but allowing for anisotropic scaling factors around the center of the template in the 3 axis directions. Figure 2 shows overlays of an axial, coronal and sagittal slice of the template on the registered MR image. Because of the variability of the brain topology, the rigid body assumption is not satisfied in this case, but the MI criterion succeeds at finding such a transformation that on average very well matches corresponding structures in both images.

The same method has been applied successfully on 15 other patient images. Visual inspection of the results by clinicians showed superior performance of the MI criterion compared to the registration technique that is used in the SPM package of Friston et al. [7], which is based on intensity correlation of segmented white and grey matter regions.

4.3. Image-to-model registration

To validate the geometric accuracy of various spiral CT imaging protocols for orthopedic applications [17], spiral CT images were acquired from a hardware phantom of the spine [10]. This phantom is a geometric abstraction of three vertebrae, consisting of different vertebrae components. Different components are modeled by different substitute materials, which appear with a distinct intensity in the image. A mathematical model was constructed from the geometrical description of the phantom by assigning each structure a different label (32 in total, including background). The CT images were registered with this model using the mutual information criterion on the intensity of each voxel in the CT image and its corresponding label in the model. After registration, edges segmented from the image are compared with the corresponding edges in the model, which allows to assess the acquisition accuracy, provided that registration and segmentation errors can be assumed to be small.

The result of the registration is illustrated in figure 3, showing the midsagittal plane through the model and the corresponding plane in the image after registration for a 1/1 spiral CT acquisition reconstructed at 0.5 mm slice distance. The original voxelsizes are $0.16 \times 0.16 \times 0.5$ mm, but the images of figure 3 were resampled into $0.16$ mm square voxels for the purpose of visualization. Overlay of the contours of the model on the CT image allows visual inspection of the registration accuracy. The segmentation result shown in figure 3 is obtained automatically using the method described in [12].

5. Discussion

The mutual information matching criterion is based on the assumption that the statistical dependence between corresponding voxel intensities is maximal if both images are geometrically aligned. Mutual information 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 earlier by Hill et al. [9] (third order moment of the joint histogram), Woods et al. [22] (variance of intensity ratios) and Collignon et al. [2] (entropy of the joint histogram). In appendices 1 and 2, we discuss the relationship between these measures and the MI criterion.

Because no assumptions are made regarding the nature of the dependence between corresponding voxel intensities, the MI criterion is highly data independent and allows for robust and completely automated registration of multimodality images in various applications without any prior segmentation or other pre-processing steps. The results of section 4.1 demonstrate that subvoxel registration accuracy 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. The generality of the approach is demonstrated by the results of sections 4.2 and 4.3. Although in all experiments discussed

\textsuperscript{2}Template image provided by Friston et al [7].
Figure 2. Registration of a template MR brain image to patient MR images of the head. Overlay of an axial, coronal and sagittal slice of the template on the registered patient image.

Figure 3. Registration of the mathematical model of a hardware phantom with its CT image. a) Mid-sagittal cross section through the CT image (resampled into square voxels). b) Corresponding slice through the model. c) Overlay of model contours on the registered CT image. d) Contours extracted from the CT image. e) Overlay of model and image contours allows validation of acquisition accuracy.
in this paper a rigid body transformation was used, it is clear that the MI criterion can as well be applied to more general transformations.

Partial volume interpolation was introduced to make the joint and margmal distributions and their mutual information vary smoothly for small changes in the registration parameters. Our results indicate that PV interpolation indeed behaves superior with respect to robustness compared to nearest neighbor and trilinear interpolation [11].

Evaluation of the MI criterion requires estimations of the image intensity distributions, which were obtained by simple normalization of the joint histogram. We have not evaluated the influence of the bin size, the choice of a region of interest or the application of non-linear 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 [5] on a set of samples taken from the overlapping part of both images. This approach was used by Viola et al. [20], who also use stochastic sampling of the floating image to increase speed performance.

The optimization of the MI registration criterion was implemented using Powell’s method, which proved to be very reliable and robust. To increase the speed performance of the registration method, optimization efficiency may be increased by decreasing the number of criterion evaluations, for instance by taking into account that small parameter corrections hardly influence registration accuracy. We have also noted experimentally that for high resolution data sub-sampling of the floating image can be applied without deteriorating optimization robustness. Important speed-ups can thus be realized by using a multi-resolution optimization strategy, starting with a coarsely sampled image for efficiency and increasing the resolution as the optimization proceeds for accuracy. Viola et al. [20] applied a gradient-based optimization method, using explicit expressions for the derivatives of the MI function with respect to the registration parameters. However, it is not clear how registration robustness is affected by such alternative search techniques.

Mutual information is only one of a family of measures of statistical dependence or information redundancy (see appendix 3). We have experimented with \( p(A, B) = H(A, B) - I(A, B) \), which can be shown to be a metric [4], and \( ECC(A, B) = 2I(A, B)/(H(A) + H(B)) \), the Entropy Correlation Coefficient [1]. In some cases, both measures showed superior performance compared to the original MI criterion, but we could not establish a clear preference for either of these. Furthermore, the use of mutual information for multi-modality 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.

6. Conclusion

The mutual information matching criterion allows for accurate, robust and completely automated registration of multi-modality medical images. The method is highly data independent and requires no user interaction, segmentation or pre-processing and is therefore well suited to be used in clinical practice. However, the performance of the method on clinical data needs to be further investigated and compared to that of other registration methods. Further research is also needed to better understand the influence of implementation issues, such as sampling and interpolation, on the registration criterion and to tune the optimization method towards specific applications. Finally, other registration criteria can be derived from the one presented here, using alternative information measures applied on different features.

7. Acknowledgments

This research was supported by IBM Belgium (Academic Joint Study) and by the National Fund for Scientific Research, Belgium, under grant numbers FGWO 3.0115.92, 9.0033.93 and G.3115.92. The authors would like to acknowledge the contributions to this paper of Sylvie Van Oostende and Stefan Sunaert to section 4.2 and of Johan Van Cleynenbreugel to section 4.3. The work on spiral CT accuracy measurement is part of PHIDIAS (laser PHotopolymerisation models based on medical Imaging, a Development Improving the Accuracy of Surgery) project nr. BE/5930 of the BRITE-EURAM programme of the European Commission.

A. Appendix 1

We show the relationship between the multi-modality registration criterion devised by Hill et al. [8] 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:

\[
T_n = \sum_{a,b} \left( \frac{h(a, b)}{V} \right)^n
\]  

(10)

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 \( n \) [16]:

\[
H_n = \frac{1}{1-n} \log(T_n)
\]
with the following properties:

- \( \lim_{n_1 \to 1} H_{n_1}(p) = \sum_i p_i \log p_i \), which is the Shannon entropy.

- \( n_2 > n_1 \to H_{n_2}(p) \leq H_{n_1}(p) \)

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

B. Appendix 2

We show how the multi-modality registration criterion devised by Woods et al. [22] relates to the conditional entropy \( H(a \mid b) \). 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:

\[
\sigma'' = \sum_b \frac{n_b \sigma_a(b)}{N \mu_a(b)}
\]

(11)

\[
= \sum_b p_b(b) \frac{\sigma_a(b)}{\mu_a(b)}
\]

(12)

with \( p_b \) the marginal distribution function of image intensities \( B \).

It can be shown [4] that for a given mean \( \mu_a(b) \) and standard deviation \( \sigma_a(b) \)

\[
H(A \mid B) = \sum_b p(b) H(A \mid B = b)
\]

(13)

\[
= -\sum_b p(b) \sum_a p(a \mid b) \log p(a \mid b)
\]

(14)

\[
\leq \sum_b p(b) \log(\sigma_a(b)) + \frac{1}{2} \log(2\pi e)
\]

(15)

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 [4] we get

\[
H(A \mid B) \leq \sum_b p(b) (\log(\frac{\sigma_a(b)}{\mu_a(b)}) + \log(\mu_a(b)))
\]

(16)

\[
\leq \log(\sum_b p(b) \frac{\sigma_a(b)}{\mu_a(b)}) + \log(\sum_b p(b) \mu_a(b))
\]

\[
= \log(\sigma''') + \log(\mu(a))
\]

(18)

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 et al., this assumption is approximately accomplished by editing away parts in one data set (in case 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.

C. Appendix 3

Mutual Information \( I(a; b) \) is only one example of the more general \( f \)-information measures of dependence \( f(P \mid P_1 \times P_2) \) [16] with \( P \) the set of joint probability distributions \( P(a, b) \) and \( P_1 \times P_2 \) the set of joint probability distributions \( P(a), P(b) \) assuming \( a \) and \( b \) to be independent.

\( f \)-information is derived from the concept of \( f \)-divergence, which is defined as:

\[
f(P \mid Q) = \sum_i q_i f(p_i / q_i)
\]

with \( P = \{ p_1, p_2 \ldots \} \) and \( Q = \{ q_1, q_2 \ldots \} \) 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:

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

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

- \( \chi^2 \)-information:

\[
\chi^2(P \mid P_1 \times P_2) = \sum_{i,j} \frac{(p_{ij} - p_i p_j)}{p_i p_j}
\]
Note that $I_{ij}(F_1/F_2)$ is the information-measure counterpart of the n-th order moment used by Hill et al.
for $n = \alpha = 2, 3$. Furthermore, $I_{ij}(F_1/F_2) = \sum_{p_i} \log(\frac{p_{ij}}{p_{i}p_{j}})$ which is the definition of Mutual Information used in this paper.

References


