Voxel Similarity Measures for 3-D Serial MR Brain Image Registration

Mark Holden, Derek L. G. Hill*, Erika R. E. Denton, Jo M. Jarosz, Tim C. S. Cox, Torsten Rohlfing, Joanne Goodey, and David J. Hawkes

***Abstract*—We have evaluated eight different similarity measures used for rigid body registration of serial magnetic resonance (MR) brain scans. To assess their accuracy we used 33 clinical three-dimensional (3-D) serial MR images, with deformable extradural tissue excluded by manual segmentation and simulated 3-D MR images with added intensity distortion. For each measure we determined the consistency of registration transformations for both sets of segmented and unsegmented data. We have shown that of the eight measures tested, the ones based on joint entropy produced the best consistency. In particular, these measures seemed to be least sensitive to the presence of extradural tissue. For these data the difference in accuracy of these joint entropy measures, with or without brain segmentation, was within the threshold of visually detectable change in the difference images.**

***Index Terms*—Consistency of transformation estimates, evaluation with clinical data, serial MR registration, voxel-based similarity measures.**

I. Introduction

THE registration of serial magnetic resonance images (serial MR) of the brain has recently become a widely used research tool [1], [2] with the potential to enter routine clinical use. The technique involves accurately aligning images of the same subject acquired at different times. The resulting aligned images can be subtracted and visually inspected [5], [6] or further processed to provide a quantitative measures of change [3], [7]. It has been proposed that this approach can increase the sensitivity of the observer to disease progression, response to therapy, or uptake of contrast material compared to conventional viewing [3], [7].

Several different similarity measures have been proposed in the literature [8], [6], [2] and, while these have been validated separately, there has been no rigorous comparison of these different measures on a single set of test data. There is no consensus on the benefit of presegmentation to remove deformable extradural material [8], [9] or the impact that presegmentation has on registration accuracy with different similarity measures. It is important that this is resolved as the need to presegment images is an obstacle to routine clinical use, due to the unreliability of segmentation algorithms.

There is agreement in the literature about the need for subvoxel registration accuracy in serial MR. Lemieux has estimated a registration accuracy of the order of 0.05 voxels (typically 50 μm) is required to avoid a misregistration artifact [6]. The most accurate registration gold standard for clinical data is obtainable with bone-implanted markers and has an accuracy of around 500 μm [10]. This is insufficient for assessing serial MR techniques. In light of this limitation, authors have attempted to quantify accuracy by assessing the consistency of transformations. For example, Freeborough [2] used the ratio of image uniformity (RIU) similarity measure that was proposed by Woods [11] for registration and a sinc kernel for transformation and measured the number of voxels that were outside an intensity window in the difference image [12]. Lemieux used the Pearson product-moment correlation coefficient to register scalp segmented serial images of three patients [6]. Each patient had three scans which were registered first to second, second to third, third to first, and in the opposite sense. The registration consistency of the combined transformations was measured by calculating the mean of rms displacement of 1000 uniformly distributed points in the brain (Δ_{rms} estimate). For three subjects the mean Δ_{rms} was 60 μm [6]. In a recent experiment with images acquired within a few minutes of each other, Woods compared his RIU and the least mean squared difference similarity measures and found typical inconsistency of 75–100 μm and a maximum of less than 500 μm [9]. However, it has not yet been demonstrated what accuracy is required to maximize the sensitivity and specificity of the technique in a clinical situation. It is, therefore, not clear whether a technique with an accuracy of 10 μm is more clinically useful than one with an accuracy of 250 μm.

In this paper, we report the results of a systematic comparison of eight similarity measures, three of which have previously been used for serial MR registration, four for related applications such as multimodal registration, and one that is a novel variant of the RIU measure. We measure the accuracy of the measures using simulated MR brain images [13] and quantify consistency using images from 11 subjects from a clinical growth hormone study [14]. We compare the performance of the

Manuscript received May 24, 1999; revised December 14, 1999. The work of T. Rohlfing was supported in part by Deutsche Forschungsgemeinschaft (DFG), Graduate School 331-1/97. The Associate Editor responsible for the review of this paper and recommending its publication was C. Meyer. *Asterisk indicates corresponding author*.

M. Holden, *D. L. G. Hill, J. Goodey, and D. J. Hawkes are with Radiological Sciences and Medical Engineering, Guy's, King's and St Thomas' School of Medicine, King's College London, London SE1 9RT, U.K..

E. R. E. Denton is with Radiology Department, King's College Hospital, London SE5 9RS, U.K..

J. M. Jarosz is with Neuro-Imaging Department, King's College Hospital, London SE5 9RS, U.K..

T. C. S. Cox is with the Institute of Neurology, UCL, Queen's Square, London WC1N 3BG, U.K..

T. Rohlfing is with Charité, Medical Faculty of Humboldt-University Berlin, Campus Virchow-Hospital, Department of Radiology, D-13353 Berlin, Germany.

Publisher Item Identifier S 0278-0062(00)02297-7.

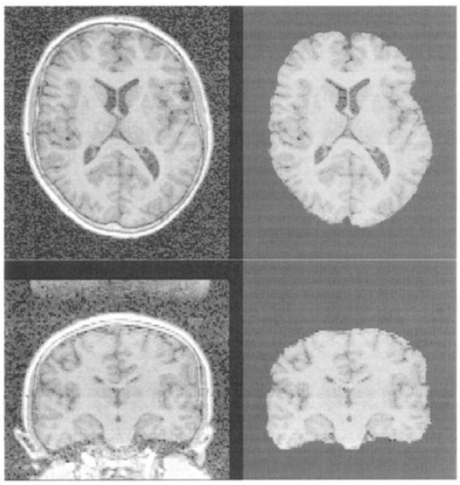

Fig. 1. Axial (top) coronal (bottom) planes through a clinical image, non-segmented (left) segmented (right).

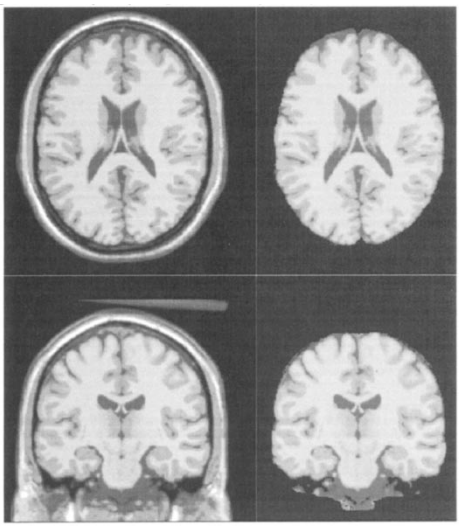

Fig. 2. Axial (top) coronal (bottom) planes through a McGill brain image without added noise or RF inhomogeneity. Nonsegmented (left) segmented (right).

similarity measures on the clinical data, with and without segmentation of extradural tissue. We interpret these results in the context of the threshold of perceptible change [14].

Our approach compares all the serial MR registration similarity measures that, to our knowledge, have been used, and does this on both simulated and clinical data. Our tests of the consistency of registration transformations extends the work of Woods on similarity measures [9] and applies them to clinical data serially acquired over a six-month period and to simulated MR brain data. It builds on the work of Freeborough [2] and Lemieux [6] by comparing several cost functions with consistency measures and relating these to blinded visual assessment of misregistration in difference images. We have shown a significant difference in the performance of measures and propose an explanation for these differences.

II. Methods and Materials

A. Acquisition and Segmentation of Clinical Data

Our clinical data are from a recent study of five adult growth hormone deficient patients undergoing therapy and six normal subjects (see [14]). Each subject was scanned three times at three monthly intervals with a fixed MR protocol. Subjects were positioned such that the whole brain and brain stem were imaged. To reduce the effect of wraparound the field of view was set so that ten slices superior to the head contained only air. All acquisitions were with a 1-T Siemens Impact Expert with a three-dimensional (3-D) flash spoiled gradient echo sequence: 256 $\times$ 192 $\times$ 169-mm FOV; 256 $\times$ 256 $\times$ 94, 1 $\times$ 1 $\times$ 1.8-mm voxels; axial slices; 20/6/2/60 (TR/TE/NEX/flip). The readout gradient was in the posterior/anterior direction with a field strength of 4.587 mT/m. As part of the preamble of the 3-D flash sequence, a scout pulse is transmitted and the returned signal used to normalize the intensity of the whole volume. As part of the scanning protocol, a scaling phantom was also scanned so that scaling errors could be measured [15]. Images were segmented by a radiologist to eliminate potentially deformable tissue, such as the scalp and muscles of the head and neck, using the interactive tracing facility available in the Analyze software package (Mayo Clinic, Rochester, MN). As a result, voxels corresponding to tissue external to the dura were excluded. Fig. 1 (left) shows the mid axial and coronal planes of a typical scan of a growth hormone patient, the extradural segmentation of this image is also shown (right).

B. Simulated MR Brain Image with Added Noise and Distortion

Simulated data were derived from the McGill full anatomical MR brain model image, 181 $\times$ 217 $\times$ 181, 1 $\times$ 1 $\times$ 1-mm voxels [13]. Two different noiseless images were used, one without any intensity distortion, the other with 40% RF inhomogeneity. To create a segmented brain comparable to the ones from the clinical data, the models of cerebrospinal fluid, grey matter, white matter, and glial matter tissue classes were combined. Fig. 2 (left) shows axial and coronal planes through the image of the brain model without added noise or RF distortion the segmented image is also shown (right).

During the acquisition of an MR image, noise arises in both real and imaginary signals. The two noise signals are independent and Gaussian distributed, so noise in the magnitude image is Rician distributed [16]. For a correctly calibrated quadrature detector the real and imaginary noise have zero mean and the same variance [17]. To simulate Rician noise a numerical complex random variable was generated and added to each voxel of the noiseless (real) image, then the modulus was taken to produce a magnitude image. The complex random variable was based on two independent Gaussian distributed random deviates as described by Press [18]. The mean and standard deviation of the simulated noise distribution were estimated from measurements of voxel intensities of an artifact-free region of a clinical image corresponding to air, as described by Henkelman [17].

C. Voxel Similarity Measures

The purpose of a similarity measure is to return a value indicating how well two images match. Ideally, the function would have one optimum at the point of registration and, to simplify

optimization, be quadratic. In the case of two images each with Gaussian noise and whose intensity transformation is the identity, it has been shown that the sum of the square intensity differences measure is optimal [19]. If there is a linear intensity transformation, then the Pearson product moment measure is optimal [19]. In cases where the intensity transformation is not known Viola [19] proposed that a measure based on joint entropy would be optimal. For serial MR, neither the Gaussian noise distribution or linear intensity transformation are valid assumptions so neither the sum of square intensity differences nor the Pearson product moment measures are optimal. A Gaussian noise distribution is present in complex MR images, but for magnitude images the noise distribution is Rician [16]. Although the Rician distribution is approximately Gaussian for high-intensity signals, for low intensities (typically 40% of voxels in an MR head image correspond to air) it is approximately Rayleigh distributed [16]. The Rayleigh probability density function can be expressed as $P_M(M) = (M/\sigma^2)\ \exp(-M^2/2\sigma^2))$ where M is the actual intensity and σ is the standard deviation of the Gaussian noise in the real and imaginary images [16]. In addition to noise, there is intensity distortion in typical clinical data, such as the bias field due to RF inhomogeneity and motion artifacts due to blood and CSF flow. Furthermore, although the brain is rigidly attached to the skull, the scalp and facial tissue may deform between serial acquisitions and so violate the rigid-body hypothesis. Typically, about 20% of voxels in an MR head image correspond to tissue types that are potentially deformable, e.g., skin, muscle, and fat. Although these tissues are attached to rigid structures, e.g., the skull, they can deform by several millimetres. Deformations are caused by forces acting upon soft tissue while the subject is in the scanner. Typical physical forces are gravity and friction which acts on a region at the back of the head in contact with the bed. These forces are likely to lead to large deformations in a particular direction. Forces arising from biological processes such as muscle movements around the eyes and forehead are likely to cause smaller deformations in many different directions. If the deformation is substantially different between acquisitions then the optimum of the similarity measure could correspond to a significantly different transformation [8]. Similarity measures may differ in their sensitivity to these changes. The complicating factors of Rician noise, distortion, deformable tissue, and potential anatomical change make the problem difficult to resolve by theoretical analysis and suggest that a systematic experimental approach is most appropriate.

We have implemented eight similarity measures: 1) mean square difference of intensities (MSD) [8]; 2) entropy of the difference image (EDI) [20], [21]; 3) mutual information (MI) [22], [19], [23], [24]; 4) Pearson product-moment cross correlation (NCC) [6]; 5) normalized mutual information (NMI) [25], [26]; 6) pattern intensity, radius one, $\sigma = 10$, (PI) [27]; 7) ratio image uniformity (RIU); and 8) we have also devised a novel variant of the RIU measure (MRIU).

These measures are defined mathematically in Section II-D below. Three of these measures have been widely used by researchers in serial MR (MSD, NCC, and RIU). With the exception of the new measure MRIU, all the others have been used in other image registration applications. The measures can be split into two groups: a) those based on entropy: MI, NMI, and EDI and b) those based on correlation: MSD, NCC, PI, RIU, and MRIU. The entropy-based measures are one-dimensional (1-D) in the case of entropy of the difference image and two-dimensional (2-D) (joint entropy) in the case of the MI measures. The MSD and NCC measures are variants of standard statistical correlation. The PI measure applies a mean squared difference type correlation measure to each local neighborhood thus allowing for local spatial changes in intensity between the two images. The ratio image uniformity measure is robust to linear changes in mean intensity because these are factored out by the division.

D. Mathematical Definition of Measures

All the measures are applied to overlapping voxels in the intersecting region of the reference (target) and transformed images. This region, denoted by $\mathbf{X_0}$, is the subset of N_0 voxel locations of the reference image, $\mathbf{X}$, whose corresponding locations are in the space of the transformed image $\mathbf{X'}$, or, more precisely, $\mathbf{X_0} = \{\mathbf{x}_0 \colon \mathbf{x}_0 \in \mathbf{X} \cap \mathbf{T}(\mathbf{X'})\}$ where $\mathbf{T}$ is the trial rigid body transformation (see later). The intensity of a voxel located at $\mathbf{x}_0$, in the reference image is denoted by $f(\mathbf{x}_0)$ and the corresponding one in the transformed image by $g(\mathbf{x}_0)$. The sets of intensities of the overlapping voxels of the reference and transformed images are referred to as $F(\mathbf{X_0})$ and $G(\mathbf{X}_0)$, respectively. The difference image $D(\mathbf{X_0})$ is defined as the set of differences of overlapping voxel intensities, i.e., $d(\mathbf{x}_0) = f(\mathbf{x}_0) - g(\mathbf{x}_0)$. $P(f_a)$, $P(g_b)$ refer to the probability of observing (frequency of occurrence) intensity f_a in $F(\mathbf{X_0})$ and g_b in $G(\mathbf{X}_0)$, respectively. $P(f_a, g_b)$ is the joint probability of observing the pair of intensities (f_a, g_b) at corresponding voxel locations.

1) MSD: we seek the minimum of $\mathrm{msd}(F, G)$ where

$$\mathrm{msd}(F,G) = \frac{1}{N_0} \sum_{\mathbf{x}_0 \in \mathbf{X}_0} (f(\mathbf{x}_0) - g(\mathbf{x}_0))^2.$$

2) (EDI: we seek the minimum of $H(D(\mathbf{X_0}))$ where

$$H(D(\mathbf{X_0})) = -\sum_{d_a \in D(\mathbf{Z_0})} P(d_a)\ \log(P(d_a))$$

3) MI: we seek the maximum of $I(F, G)$ where

$$I(F,G) = \sum_{(f_a,g_b) \in F(\mathbf{X}_0) \times G(\mathbf{X}_0)} P(f_a, g_b) \cdot \log\left(\frac{P(f_a, g_b)}{P(f_a)P(g_b)}\right)$$

and where $F(\mathbf{X}_0) \times G(\mathbf{X}_0)$ is the space of joint intensities.

4) NCC: we seek the maximum of $\mathrm{NCC}(F, G)$ where

$$\mathrm{NCC}(F,G) = F(\mathbf{X_0}) \star G(\mathbf{X}_0) = \frac{1}{N_0^2} \frac{\sum_{\mathbf{x}_0 \in \mathbf{X}_0} \left(f(\mathbf{x}_0) - \overline{f}\right)\left(g(\mathbf{x}_0) - \overline{g}\right)}{\sigma_f \sigma_g}$$

where $\overline{f}$ and $\overline{g}$ are the mean intensities of $F(\mathbf{X}_0)$ and $G(\mathbf{X}_0)$ and

$$\sigma_f = \frac{1}{N_0}\sqrt{\sum_{\mathbf{x}_0 \in \mathbf{X}_0} \left(f(\mathbf{x}_0) - \overline{f}\right)^2}$$

$$\sigma_g = \frac{1}{N_0}\sqrt{\sum_{\mathbf{x}_0 \in \mathbf{X}_0} \left(g(\mathbf{x}_0) - \overline{g}\right)^2}$$

5) NMI: we seek the maximum of $Y(F,G)$ where the equation at the bottom of this page holds.
6) PI: this measure is applied locally in a sphere (in discrete space), $S_r(\mathbf{x}_0)$, radius, r, center, $\mathbf{x}_0$, in the difference image, $D(\mathbf{X}_0)$. We seek the maximum of $P_{r,\sigma}\left(D(\mathbf{X}_0)\right)$ where

$$P_{r,\sigma}\left(D(\mathbf{X}_0)\right) = \frac{1}{N_0}\sum_{\mathbf{x}_0 \in \mathbf{X}_0}\ \sum_{\mathbf{u}:\ |\mathbf{x}_0 - \mathbf{u}|^2 \le r^2} \cdot \left(\frac{\sigma^2}{\sigma^2 + (d(\mathbf{x}_0) - d(\mathbf{u}))^2}\right)$$

and where $\mathbf{u} = (u, v, w)$ are the locations of voxels in $S_r(\mathbf{x}_0)$ and $|\mathbf{x}_0 - \mathbf{u}|$ is the Euclidean distance from center of the sphere and $\mathbf{x}_0 = (x_0, y_0, z_0)$ and σ^2 is an arbitrary constant.
7) RIU: we seek the minimum of $\mathrm{RIU}(F,G)$ where

$$\mathrm{RIU}(F,G) = \frac{1}{\mu\left(\frac{g}{f}\right)}\ \frac{1}{N_0}\sum_{\mathbf{x}_0 \in \mathbf{X}_0}\left(\mu\left(\frac{g}{f}\right) - \frac{g(\mathbf{x}_0)}{f(\mathbf{x}_0)}\right)^2$$

and

$$\mu\left(\frac{g}{f}\right) = \frac{1}{N_0}\sum_{\mathbf{x}_0 \in \mathbf{X}_0}\frac{g(\mathbf{x}_0)}{f(\mathbf{x}_0)}.$$

8) MRIU: the standard RIU measure assigns an uneven weighting to off-diagonal intensity pairs. The weighting for pairs for which the numerator (transformed voxel intensity) is high and the denominator (target voxel intensity) is low gives a high weighting and the converse gives a low weighting. This bias can be reduced if a constant c (we used the maximum intensity of the target image) is added to both the numerator and denominator, hence, we seek the minimum of $\mathrm{MRIU}(F,G)$ where

$$\mathrm{MRIU}(F,G) = \frac{1}{\mu\left(\frac{g}{f}\right)}\ \frac{1}{N_0}\sum_{\mathbf{x}_0 \in \mathbf{X}_0} \cdot \left(\mu\left(\frac{g+c}{f+c}\right) - \frac{g(\mathbf{x}_0)+c}{f(\mathbf{x}_0)+c}\right)^2.$$

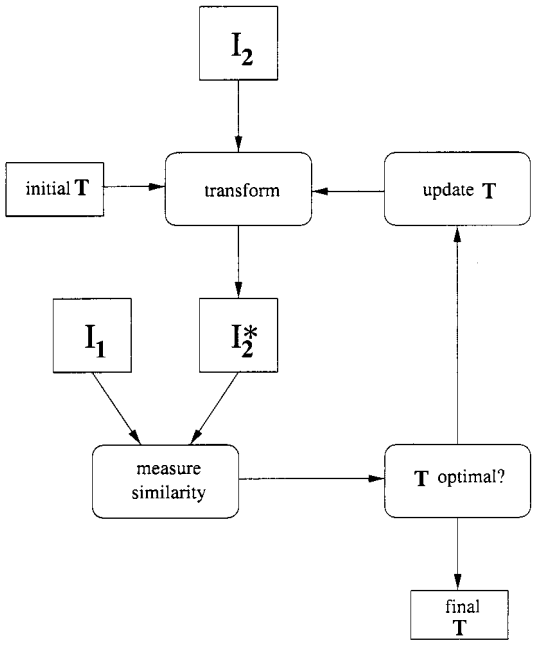

Fig. 3. Schematic representation of flow of control of registration algorithm for a single resolution level. Measure similarity, transform and update $\mathbf{T}$ refer to the matching, transformation, and optimization stages, respectively. I_1 (reference image), I_2 (transformed image), I_2^* (I_2 spatially transformed by $\mathbf{T}$). The initial transformation $\mathbf{T}$ is iteratively modified to maximize the similarity of I_1 and I_2^*.

As well as reducing the bias, another advantage this measure has over RIU is that low intensities need not be excluded (by thresholding) and so statistics can be based on a larger number of voxels.

E. Registration Algorithm

Our algorithm has been designed in four distinct modules so that each can be changed individually to allow the impact of different strategies on registration accuracy to be investigated. Fig. 3 shows schematically the flow of control. Image I_1 is the baseline (target) image and I_2 is the transformed image. The algorithm consists of four main stages: 1) a data preparation stage that creates a set of multiscale images and prescribes an initial transformation $\mathbf{T}$; 2) a function that measures the similarity of a pair of images; 3) a function that spatially transforms, using trilinear interpolation, the I_2 image into I_2^* according to a prescribed rigid body transformation, $\mathbf{T}$; and 4) an optimization stage that searches for the optimal transformation and prescribes a new search interval and position in search space, $\mathbf{T}$.

The optimization stage is similar to that described by Studholme [28]. It evaluates the similarity measure within a local neighborhood of the current position $\mathbf{T}$ and tests for convergence. The neighborhood is the set of points at $\pm\delta T$

$$Y(F,G) = \frac{\sum_{f_a \in F(\mathbf{X}_0)} P(f_a)\log\left(P(f_a)\right) + \sum_{g_b \in G(\mathbf{X}_0)} P(g_b)\log\left(P(g_b)\right)}{\sum_{(f_a,g_b)\in F(\mathbf{X}_0)\times G(\mathbf{X}_0)} P(f_a,g_b)\log\left(P(f_a,g_b)\right)}$$

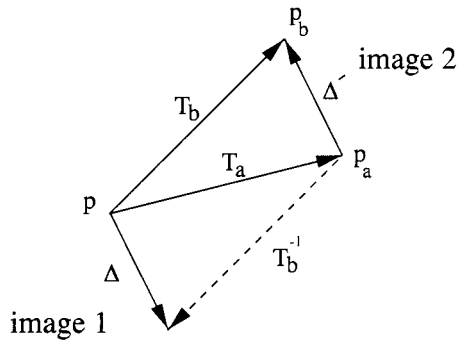

Fig. 4. Comparing two transformation estimates.

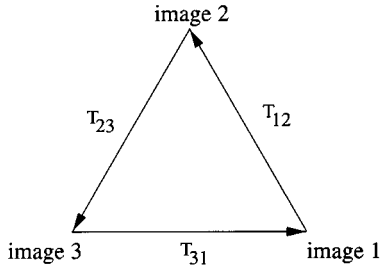

Fig. 5. Composition of three perfect transformations.

where δT is the search interval in each of the six directions of rigid body motion. The search interval is gradually decreased, and $\mathbf{T}$ updated as better matches are achieved. The convergence criterion is that the current transformation estimate is the best one in the local neighborhood for the smallest search interval. The function measure similarity provides an estimate of how well the images I_1 and I_2^* match.

1) Multiresolution Search: This is a refinement to the optimization strategy aimed at reducing the number of iterations needed to find the optimal transformation estimate and decreasing the incidence of local optima far from the optimal solution [29]–[31]. A pyramid of images, each of successively lower resolution than the original (base resolution) is created for both reference and transformed images in a similar way to that described by Studholme [28]. The image is sub-sampled to an isotropic voxel dimension with a Gaussian kernel to create the first lower (base) resolution image. The base voxel dimension is chosen to be the next power of two greater than the largest original dimension (in mm). So 1× 1 × 1.8-mm images are resampled to 2 × 2 × 2 mm. Each low resolution level is produced by calculating the eight neighborhood mean intensity of the previous level. This strategy is a simple and a computationally inexpensive way of constructing the lower resolution levels. In the multiresolution search strategy, transformation estimates $\mathbf{T}(r)$ from the matching level r are propagated to the next higher resolution level, $r-1$.

F. Measures of Registration Consistency

We term the combination or product of two transformations [six-dimensional (6-D) vectors] a composition, so for two transformations $\mathbf{T}_1$ and $\mathbf{T}_2$ we write $\mathbf{T}_2 \circ \mathbf{T}_1$ for the transformation that applies first $\mathbf{T}_1$ then $\mathbf{T}_2$. Given two estimates of a transformation $\mathbf{T}_a$ and $\mathbf{T}_b$ that map points in image one, e.g., p, to points p_a and p_b in image two, then a measure of their vector difference in image one can be derived by applying the composition of $\mathbf{T}_a$ with the inverse of $\mathbf{T}_b$ to p, i.e., $\Delta(p) = \mathbf{T}_b^{-1} \circ \mathbf{T}_a(p)$ (see Fig. 4).

1) Consistency of Two Transformations: To measure the shift of a voxel location p resulting from the transformation Δ, we calculate the length of the displacement vector $\Delta(p)$, i.e., $|\Delta(p)|$. For a set of voxel locations $\{p(i)\}$ in an image I, we can determine a mean displacement $\langle dp \rangle$ given by $\langle dp \rangle = (1/N) \sum_{\forall i \in I} |\Delta(p(i))|$ where N is the number of voxels in the image and i is a voxel index. Similarly, if we wish to determine the mean displacement in a region of interest, for example, just those voxels that correspond to brain tissue, then the measure is applied only to the region of interest $\langle dp \rangle = (1/N_B) \sum_{\forall i \in I_B} |\Delta(p(i))|$ where I_B is the region of interest containing N_B voxels. Some authors argue that an rms measure is less biased [6]. This can be calculated as follows: $dp_{\mathrm{rms}} = (1/N) \sqrt{(\sum_{\forall i \in I} |\Delta(p(i))|^2)}$.

2) Consistency of Three Transformations: For N images the number of possible transformations is the number of two permutations of N, i.e, $\mathcal{P}(N,2) = (N!/(N-2)!) = N(N-1)$. So, for three images one, two, three, there are $3 \times 2 = 6$ possible transformations between image pairs. If we consider three transformations $\mathbf{T}_{12}, \mathbf{T}_{23}, \mathbf{T}_{31}$ between image pairs such that $\mathbf{T}_{12}$ transforms image one into the coordinate frame of image two, etc., then in the absence of error their composition $\mathbf{T}_{31} \circ \mathbf{T}_{23} \circ \mathbf{T}_{12}$ maps image one back into it's own coordinate system, forming a closed loop (circuit) and is, therefore, equivalent to the identity as shown schematically in Fig. 5.

If the transformations are estimated by a registration algorithm each will, inevitably, have some error and hence their composition will differ from the identity by an amount that depends on the error in each transformation. It is possible, though extremely improbable, that errors may combine and cancel out.

For the case of 3-D images, the rigid body transformations can be represented as 4 × 4 homogeneous matrices, so we can replace the explicit ∘ symbol for composition with implicit matrix multiplication and write $\mathbf{T}_{31}\mathbf{T}_{23}\mathbf{T}_{12} = \mathbf{I} + \boldsymbol{\Delta}\mathbf{T}$ where $\mathbf{I}$ is the identity transformation, so $\boldsymbol{\Delta}\mathbf{T} = \mathbf{I} - \mathbf{T}_{31}\mathbf{T}_{23}\mathbf{T}_{12}$ is the error resulting from each individual transformation. Applying the error transformation to each voxel location $p(i)$ and taking the modulus, as above, we can calculate the mean error over the image $(1/N) \sum_{\forall i \in I} |\boldsymbol{\Delta}\mathbf{T}(p(i))|$.

G. Registration of Clinical Data and Measurement of Consistency

All registrations were rigid body with six degrees of freedom, four lower resolution levels and with the identity as the starting transformation. The search interval (step size) ranged from 4 mm, 4° to 0.01 mm, 0.01°. The step size was always reduced by a factor of $1/\sqrt{2}$. For the lower resolution levels four, three, two, and one there were four, two, two, and four reductions in the step size, respectively and for the highest resolution level, original image, there were 12 reductions. For each of the 11 subjects, who were imaged 3 times, the first image was registered to the second, the second to the third, and the third to the first. For each of the eight similarity measures there were three registrations for each of the 11 subjects, with

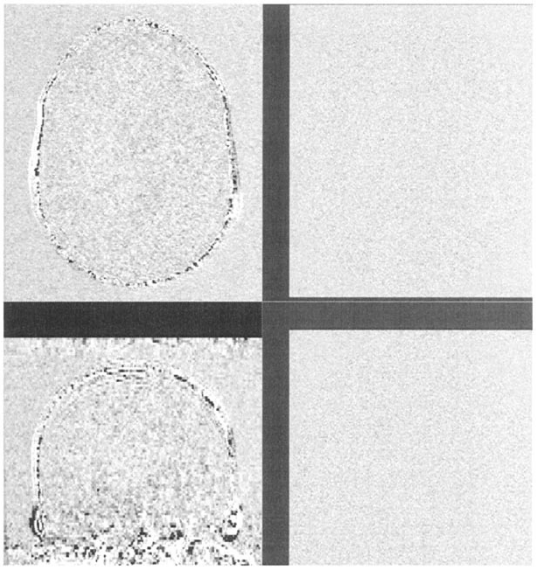

Fig. 6. Axial (top) coronal (bottom) planes through two difference images, clinical data (left) McGill data with simulated Rician noise (right).

both unsegmented and segmented data. This made a total of 66 registrations for each similarity measure, 33 with unsegmented data and 33 with segmented data. Both consistency measures described in Section II-F were applied to each set of transformations for each of the 11 subjects. This produced a total of 33 measurements of the consistency of corresponding pairs of transformation estimates derived from registration of segmented data and non segmented data; 11 measurements of internal (triangular) consistency for registrations with segmented data and 11 for registrations without. For each of the consistency measurements the mean, rms, and maximum voxel shifts over the segmented brain region were calculated.

H. Registration of Simulated Data

We used the noiseless brain model image and the noiseless brain model with 40% RF inhomogeneity from the McConnell Brain Imaging Centre, McGill University [32] to create four image pairs: 1) two noiseless replica images; 2) two noiseless images with different simulated noise; 3) two noiseless images, one with and one without RF inhomogeneity; and 4) two images with simulated noise, one with and one without RF inhomogeneity. So that the effects of interpolation could be avoided, images were not transformed before registration. Instead the registration algorithm was given starting estimates that represented typical misregistration. This was done by choosing six transformations from the registrations (with NMI) of patient data in the clinical study. A perfect solution is the identity transformation and so the inconsistency between the estimate obtained and the identity was measured over voxels corresponding to the brain. Fig. 6 shows resulting difference images from subtracting two serial clinical scans (left) and two noisy simulated model images (right). This demonstrates the similarity of the simulated and actual noise distributions. For the simulated data exact registration is achieved so the difference signal is spatially uncorrelated. For the clinical data there are spatially correlated signals near the scalp because of non-rigid tissue deformation.

TABLE I
MEAN (STANDARD DEVIATION) OF 33 MEASUREMENTS OF THE CONSISTENCY BETWEEN TRANSFORMATION ESTIMATES (SEE FIG. 4) OBTAINED BY REGISTRATION WITH AND WITHOUT PRIOR SEGMENTATION

measure	mean(μm)		RMS (μm)		max (μm)	
MSD	175	(76)	185	(81)	331	(148)
EDI	8429	(5316)	8587	(5298)	12033	(6139)
MI	122	(46)	128	(50)	223	(96)
NCC	164	(74)	174	(78)	316	(143)
NMI	121	(48)	127	(51)	222	(96)
PI	700	(1503)	756	(1629)	1510	(3234)
RIU	880	(609)	942	(651)	1797	(1228)
MRIU	171	(69)	181	(72)	322	(134)

TABLE II
MEAN (STANDARD DEVIATION) OF 11 MEASUREMENTS OF THE CONSISTENCY OF THE COMPOSITION OF TRANSFORMATION ESTIMATES (SEE FIG. 5) FROM REGISTRATION WITHOUT PRIOR SEGMENTATION

measure	mean (μm)		RMS (μm)		max (μm)	
MSD	91	(28)	96	(31)	165	(55)
EDI	1757	(1148)	1780	(1198)	2160	(2072)
MI	88	(23)	94	(26)	168	(53)
NCC	87	(37)	93	(40)	168	(73)
NMI	86	(32)	92	(35)	162	(70)
PI	1565	(2204)	1690	(2391)	3348	(4809)
RIU	1221	(553)	1276	(549)	2100	(748)
MRIU	92	(40)	99	(43)	176	(78)

TABLE III
MEAN (STANDARD DEVIATION) OF 11 MEASUREMENTS OF THE CONSISTENCY OF THE COMPOSITION OF TRANSFORMATION ESTIMATES (SEE FIG. 5) OBTAINED BY REGISTRATION WITH PRIOR SEGMENTATION

measure	mean (μm)		RMS (μm)		max (μm)	
MSD	99	(31)	104	(33)	169	(62)
EDI	66	(35)	69	(36)	117	(55)
MI	78	(29)	82	(31)	139	(59)
NCC	97	(25)	101	(27)	168	(63)
NMI	78	(29)	81	(30)	133	(57)
PI	145	(94)	154	(100)	278	(181)
RIU	258	(92)	276	(101)	531	(224)
MRIU	85	(33)	90	(35)	161	(74)

III. RESULTS

We present consistency measurements for registration with all eight measures for both clinical (segmented and not segmented) and simulated data. All measurements are expressed as the mean, rms, and maximum voxel displacements over the segmented brain region and are given in microns, rounded to the nearest micron. To assist in the interpretation of the results, we plot some example joint probability distributions for simulated data and use these to explain why some measures perform better than others.

A. Registration Consistency for the Eight Similarity Measures

Shown in Tables I, II, and III are the averages of consistency measurements for transformation estimates derived from rigid body registration of the clinical data, with and without segmentation, with each of the eight measures. The statistics given, for each measure, are the mean (standard deviation) of the consistency measurements of 33 transformation estimates obtained from three registrations of 11 subjects. Measurements of the

TABLE IV
MEAN (STANDARD DEVIATION) OF REGISTRATION ACCURACY USING SIX DIFFERENT STARTING TRANSFORMATIONS. TWO IDENTICAL IMAGES

measure	mean (μm)		RMS (μm)		max (μm)	
MSD	5	(1)	5	(1)	9	(1)
EDI	5	(1)	6	(1)	10	(2)
MI	7	(1)	7	(1)	10	(1)
NCC	5	(2)	5	(1)	9	(3)
NMI	5	(1)	5	(1)	9	(2)
PI	4	(0)	4	(1)	7	(1)
RIU	4	(1)	4	(1)	8	(1)
MRIU	4	(1)	4	(1)	8	(2)

TABLE V
MEAN (STANDARD DEVIATION) OF REGISTRATION ACCURACY USING SIX DIFFERENT STARTING TRANSFORMATIONS. TWO IDENTICAL IMAGES WITH DIFFERENT SIMULATED NOISE

measure	mean (μm)		RMS (μm)		max (μm)	
MSD	127	(2)	137	(2)	243	(3)
EDI	135	(3)	141	(6)	242	(16)
MI	121	(2)	129	(5)	221	(13)
NCC	126	(2)	137	(2)	243	(5)
NMI	137	(3)	148	(3)	260	(5)
PI	343	(23)	358	(14)	582	(18)
RIU	416	(74)	432	(61)	687	(23)
MRIU	94	(1)	99	(3)	164	(29)

consistency between transformation estimates obtained by registrations of the same data, with and without baseline image segmentation, are given in Table I. Tables II and III show the mean (standard deviation) for consistency measurements for the composition of three transformation estimates $\mathbf{T}_{31} \circ \mathbf{T}_{23} \circ \mathbf{T}_{12}$ for the 11 subjects. Table II refers to transformation estimates derived from registration without segmentation and Table III to transformation estimates derived from registration with segmentation.

B. Registration Consistency with Simulated Images

Measurements of the consistency of transformation estimates derived from the four image pairs described in Section II-H are shown in Tables IV–VII. The four pairs correspond to two noiseless replica images (Table IV), two images with different Rician noise (Table V), two noiseless images, one with RF inhomogeneity and the other without (Table VI), and one image with simulated noise and the other with RF inhomogeneity and simulated noise (Table VII).

C. Analysis of Results

We tested whether the differences in samples of the mean voxel shift measurements were significant for each of the measures. The sample derived from the MI was assigned the baseline and each of the other measure's samples were tested against it. We used a nonparametric statistical sign test [33] (Matlab, Mathworks, MA) to test for significant differences in median between samples. For registration consistency (with and without prior segmentation) based on samples of 33 consistency measurements of registration solutions (the summary statistics are given in Table I) there was no significant difference for NMI ($p = 0.86$). For both MSD and NCC the difference was significant at the 5% level ($p = 0.013$). For the other 4 measures there was a significant difference at the 1% level: EDI ($p = 0.0001$), PI ($p = 0.00006$), MRIU ($p = 0.000113$),

TABLE VI
MEAN (STANDARD DEVIATION) OF REGISTRATION ACCURACY USING SIX DIFFERENT STARTING TRANSFORMATIONS. ONE IMAGE WITH RF INHOMOGENEITY THE OTHER WITHOUT. THE ASTERISK (*) INDICATES A FAILURE THAT WAS OMITTED

measure	mean (μm)		RMS (μm)		max (μm)	
MSD	8	(2)	8	(2)	12	(3)
EDI	39	(14)	41	(15)	68	(27)
MI	51	(1)	55	(1)	99	(3)
NCC	25	(5)	25	(5)	39	(8)
NMI	43	(6)	44	(7)	63	(13)
PI	12	(1)	13	(1)	22	(3)
RIU*	30	(7)	32	(9)	62	(21)
MRIU	18	(9)	18	(10)	29	(16)

TABLE VII
MEAN (STANDARD DEVIATION) OF REGISTRATION ACCURACY USING SIX DIFFERENT STARTING TRANSFORMATIONS. ONE IMAGE WITH RF INHOMOGENEITY AND NOISE THE OTHER WITH DIFFERENT NOISE

measure	mean (μm)		RMS (μm)		max (μm)	
MSD	203	(3)	213	(3)	372	(33)
EDI	253	(2)	270	(3)	488	(6)
MI	163	(5)	174	(6)	307	(15)
NCC	133	(2)	141	(2)	260	(5)
NMI	194	(7)	208	(9)	367	(21)
PI	391	(26)	401	(19)	622	(27)
RIU	402	(57)	413	(50)	639	(9)
MRIU	149	(2)	157	(3)	288	(6)

and RIU ($p < 0.000001$). For the 11 measurements of the internal consistency for unsegmented data (see Table II) there were significant differences at the 1% level between MI and EDI, RIU, and PI ($p = 0.00097$). For segmented data (see Table III) there were significant differences for the RIU measure at the 5% level ($p = 0.011$). We also tested whether there were significant differences in internal consistency for registration solutions with segmented and with unsegmented data for each measure. The Wilcoxon rank sum test was applied to the two sets of 11 measurements for each measure. The results showed that there were significant differences for the two types of data for EDI ($p = 0.000071$), RIU ($p = 0.000093$), and PI ($p = 0.0038$). None of the other measures, CHI, NCC, MI, NMI, and MRIU, showed any significant differences at the 5% level for the two types of data.

D. Relationship of Corresponding Voxel Intensities

It is clear from inspection of Tables I–VI and from the statistical analysis in Section III-C that some of the similarity measures that we have compared perform significantly better than others for the clinical data and show the same pattern for the simulated data. In particular, the measures based on joint entropy (MI, NMI) perform best. The joint probability distribution $P(f_a, g_b)$ of corresponding voxel intensities f_a and g_b can be displayed as images and provide a useful tool for visualizing how the intensities are related (see earlier work in [34]). We use these joint distributions to better understand the effect of noise and RF inhomogeneity. The joint probability distribution of three pairs of brain model images are shown in Figs. 7–9. Fig. 7 shows the joint distribution for two identical images at registration and with small amounts of misregistration. At registration there are only diagonal entries in the distribution. As mis-

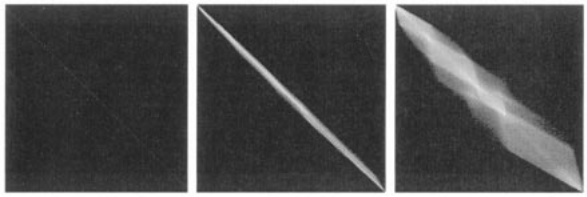

Fig. 7. Joint probability distributions of voxel intensities for replicas of the noiseless brain model image. At registration (left) misregistered by an x translation of 0.1 mm (center) and 0.5 mm (right). The axes are the image intensities (origin top left corner, range 0–255 grey levels). The intensity shown is the logarithm of the probability.

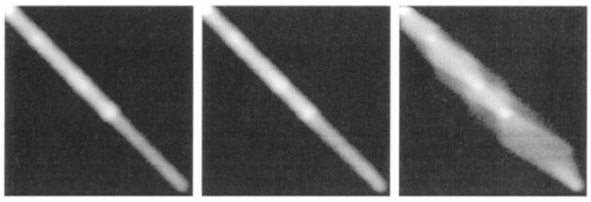

Fig. 8. Joint probability distributions of voxel intensities for brain model images with different added Rician noise. At registration (left) misregistered by an x translation of 0.1 mm (center) and 0.5 mm (right).

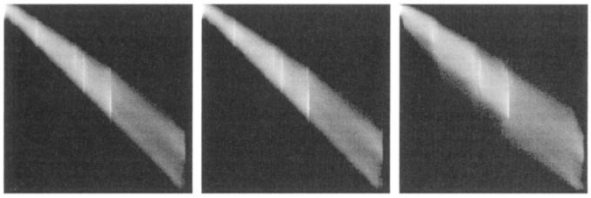

Fig. 9. Joint probability distributions for intensities of noiseless brain model image and the brain model image with added Rician noise and 40% intensity inhomogeneity. At registration (left) misregistered by an x translation of 0.1 mm (center) and 0.5 mm (right).

registration increases off-diagonal entries start to appear and the distribution becomes dispersed. Fig. 8 shows the effect of Rician noise, in both images, on the joint distribution. At registration the diagonal entries are blurred, but the blurring is symmetric and so the linear relationship is preserved. Off-diagonal entries again appear as misregistration increases. Fig. 9 shows the effect of intensity inhomogeneity, in one image, on the joint distribution. There is increased dispersion and it is intensity dependent. In general, the intensity inhomogeneity will change for serial acquisitions, this will lead to nonsymmetric dispersion and a nonlinear relationship between intensities.

IV. Discussion and Conclusion

In this paper we have carried out a comparison of eight similarity measures for serial MR registration with the aim of identifying the most appropriate one for use with clinical images containing subtle amounts of anatomical change. We have quantified the accuracy of the measures using simulated data, and their consistency using clinical images. In the simulation, we investigated the effect of noise and RF inhomogeneity on their accuracy and using the clinical images, we compared their performance on images with, and without prior segmentation of extradural tissue. In order to interpret the accuracy and consistency of values that we measured, we compare them with the threshold of discernable misregistration which was shown to correspond to a mean voxel shift of about 200 μm for similar data [14]. We use this threshold to demarcate acceptable accuracy for the simulated data and acceptable consistency for the clinical data. We found that all eight measures were accurate to within the threshold of acceptable accuracy for the noiseless simulated images with or without RF inhomogeneity. When using simulated images differing only by noise, six of the measures (MSD, EDI, MI, NCC, NMI, and MRIU) were accurate to within the threshold. When both noise and RF inhomogeneity were included, four of the measures had acceptable accuracy (MI, NCC, NMI, and MRIU), although both MSD and EDI were borderline. For the clinical data, there was a similar ranking of the measures. For the segmented images all measures except RIU could register the images with acceptable consistency and RIU was borderline (Table II). Without segmentation, however, only five measures had acceptable consistency (MSD, MI, NCC, NMI, and MRIU). These same five measures produced solutions, with or without segmentation, that agreed within the threshold of visual detectability of misregistration (see Table I). This is consistent with visual assessment results from the clinical study which indicated that there was no significant difference in perceived misregistration between the sets of difference images registered (with NMI), with and without segmentation [14].

Statistical analysis of Table I showed that MI performed significantly better than the measures that were not based on joint entropy. A possible reason for the relatively good performance of the MI measures can be seen by visual inspection of the joint probability distribution in Figs. 7–9.

Joint entropy $H(X,Y)$ is directly related the joint probability distribution, $P(x_i,y_j)$,, i.e., $H(X,Y) = -\sum_{x_i \in X} \sum_{y_j \in Y} P(x_i,y_j) \log P(x_i,y_j)$, and defines the statistical relationship of corresponding voxel intensities. Low joint entropy is equivalent to low dispersion of the joint histogram, i.e. a compact relationship. There is no assumption here about the form of relationship between voxel intensities, it could be nonlinear. In contrast, for a correlation measure, a linear relationship of corresponding voxel intensities is required. In general, for serial acquisitions the relationship is nonlinear because of intensity distortion, as illustrated in Fig. 9. A linear relationship between intensities (i.e., a line in the joint histogram) is required for a correlation type measure. A measure based on joint entropy does not require such a linear relationship. A low joint entropy simply corresponds to a compact joint distribution (i.e., a small number of entries in the joint distribution).

Results with the simulated images suggested that image noise had a significant effect on registration accuracy. However, the highest resolution matching was done with images at the original resolution without any filtering to reduce the impact of noise. It is possible that low-pass filtering with intensity thresholding might improve performance of some measures.

The results for both clinical and simulated data show that the modified RIU criterion performs better than the original one. This is most likely to be attributable to its increased robustness to RF inhomogeneity since high/low intensities are more equally weighted. However, the clinical results indicate that modified

RIU performs no better than either the MSD or NCC correlation measures and worse than both MI and NMI.

Our results show that the similarity measures based on MI are the most suitable to determine the six degrees of freedom (rigid body) transformation between serial MR images of the head. Using our optimization strategy with trilinear interpolation we achieve registration solutions with and without extradural segmentation that are consistent to within the threshold of observer discernibility (i.e, 200–300 μm). Our results apply to serial MR images under the conditions of typical scalp deformations and small-scale anatomical change.

ACKNOWLEDGMENT

The authors would like to thank Philips Medical Systems, EasyVision Advanced Development, for their support for this work. particular thanks go to F. Gerritsen and P. Desmedt for their encouragement and constructive criticism. Thanks are also due to the McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, for the provision of simulated brain data. The authors are grateful to D. Russel-Jones, endocrinologist, for the provision of clinical data.

REFERENCES

[1] M. A. Rutherford, J. M. Pennock, F. M. Cowan, N. Saeed, J. V. Hajnal, and G. M. Bydder, "Detection of subtle changes in the brains of infants and children via subvoxel registration and subtraction of serial MR images," *Amer. J. Neuroradiol.*, vol. 18, no. 5, pp. 829–845, 1997.

[2] P. A. Freeborough, R. P. Woods, and N. C. Fox, "Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders," *J. Comp. Assisted Tomogr.*, vol. 20, no. 6, pp. 1012–1022, 1996.

[3] J. V. Hajnal, N. Saeed, A. Oatridge, E. J. Williams, I. R. Young, and G. M. Bydder, "Detection of subtle brain changes using subvoxel registration and subtraction of serial MR images," *J. Comp. Assisted Tomogr.*, vol. 19, no. 5, pp. 677–691, 1995.

[4] W. G. Bradley and G. M. Bydder, *Advanced MR Imaging Techniques*, M. Dunitz, Ed. London , U.K.: Livery House, 1997.

[5] P. A. Freeborough and N. C. Fox, "The boundary shift integral: An accurate and robust measure of cerebral volume changes from registered repeat MRI," *IEEE Trans. Med. Imag.*, vol. 16, pp. 623–629, Oct. 1997.

[6] L. Lemieux, U. C. Wieshmann, N. F. Moran, D. R. Fish, and S. D. Shoovon, "The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalisation," *Med. Image Anal.*, vol. 2, no. 3, pp. 227–242, 1998.

[7] D. Rueckert, C. Hayes, C. Studholme, P. Summers, M. Leach, and D. J. Hawkes, "Non-rigid registration of breast MR images using mutual information," in *First Int. Conf. Medical Image Computing Computer-Assisted Intervention (MICCAI '98)*, S. Delp, W. M. Wells, and A. Cochester, Eds., Cambridge, MA, USA, Oct. 1998, pp. 1144–1152.

[8] J. V. Hajnal, N. Saeed, E. J. Soar, A. Oatridge, I. R. Young, and G. M. Bydder, "A registration and interpolation procedure for subvoxel matching of serially acquired MR images," *J. Comp. Assisted Tomogr.*, vol. 19, no. 2, pp. 289–296, 1995.

[9] R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, and J. C. Mazziotta, "Automated image registration: 1. General methods and intrasubject, intramodality validation," *J. Comp. Assisted Tomogr.*, vol. 22, no. 1, pp. 139–152, 1998.

[10] C. R. Maurer, J. M. Fitzpatrick, M. Y. Wang, R. L. Galloway, R. J. Maciunas, and G. S. Allen, "Registration of head volume images using implantable fiducial markers," *IEEE Trans. Med. Imag.*, vol. 16, pp. 447–462, Aug. 1997.

[11] R. P. Woods, S. R. Cherry, and J. C. Mazziotta, "Rapid automated algorithm for aligning and reslicing PET images," *J. Comp. Assisted Tomogr.*, vol. 16, no. 4, pp. 620–633, July 1992.

[12] P. A. Freeborough and N. C. Fox, "Assessing patterns and rates of brain atrophy by serial magnetic resonance imaging: A segmentation, registration, display and quantification procedure," in *Visualization in Biomedical Computing, Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 1996, vol. 1131, pp. 419–429.

[13] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, pp. 463–468, June 1998.

[14] E. R. E. Denton, M. Holden, E. Christ, J. M. Jarosz, D. Russell-Jones, J. Goodey, T. C. S. Cox, and D. L. G. Hill, "The identification of cerebral volume changes in treated growth hormone deficient patients using serial 3-D MR image processing," *J. Comp. Assisted Tomogr.*, vol. vol. 24, no. 1, pp. 139–145, Jan. 2000.

[15] M. Holden, E. R. E. Denton, J. M. Jarosz, T. C. S. Cox, D. J. Hawkes, and D. L. G. Hill, "Detecting small anatomical change with 3D serial MR subtraction images," in *Proc. Society of Photo-Optical Instrument Engineers (SPIE)*, vol. 3661, San Diego, CA, 1999, pp. 44–55.

[16] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Resonan. Med.*, vol. 34, pp. 910–914, 1995.

[17] R. M. Henkelman, "Measurement of signal intensities in the presence of noise in MR images," *Med. Phys.*, vol. 12, no. 2, pp. 232–233, 1985.

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[19] P. A. Viola, "Alignment by maximization of mutual information," Ph.D. thesis, Massachusetts Inst. Technol., Cambridge, MA, 1995.

[20] T. M. Buzug, J. Weese, C. Fassnacht, and C. Lorenz, "Using an entropy similarity measure to enhance the quality of DSA images with an algorithm based on template matching," in *Visualization in Biomedical Computing, Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 1996, vol. 1131, pp. 235–240.

[21] ——, "Image registration: Convex weighting functions for histogram-based similarity measures," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, July 1997, vol. 1205, pp. 119–128.

[22] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Seutens, and G. Marchal, "Automated multimodality image registration using information theory," in *Information Processing in Medical Imaging*, Y. Bizais, C. Barillot, and R. Di Paola, Eds. Ile de Berder, France: Kluwer, 1995, pp. 263–274.

[23] W. M. Wells, P. Viola, H. Atsumi, and S. Nakajima, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–51, 1996.

[24] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration using mutual information," *IEEE Trans. Med. Imag.*, vol. 16, pp. 187–198, Apr. 1997.

[25] C. Studholme, "Meaures of 3D medical image alignment," Ph.D. thesis, Univ. of London, London, U.K., 1997.

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

[27] J. Weese, G. P. Penney, P. Desmedt, T. M. Buzug, D. L. G. Hill, and D. J. Hawkes, "Voxel-based 2-D/3-D registration of fluoroscopy images and CT scans for image-guided surgery," *IEEE Trans. Inform. Technol. Biomed.*, vol. 1, no. 4, pp. 284–293, 1997.

[28] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxal similarity measures," *Med. Phys.*, vol. 24, no. 1, pp. 25–35, 1997.

[29] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surveys*, vol. 24, no. 4, pp. 325–375, Dec. 1992.

[30] F. Maes, "Segmentation and registration of multi-modal medical images," Ph.D. thesis, Univ. of Leuven, 1998.

[31] J. P. Pluim, A. J. B. Maintz, and M. A. Viergever, "Multiscale approach to mutual information matching.," in *Proc. Society of Photo-Optical Instrument Engineers (SPIE)*, vol. 3338, San Diego, CA, 1998, pp. 1334–1344.

[32] R. K.-S. Kwan, A. C. Evans, and G. B. Pike, "An extensible MRI simulator for post-processing evaluation," in *Visualisation in Biomedical Computing, VBC 96, LNCS*. Berlin, Germany: Springer-Verlag, 1996, vol. 1131, pp. 135–140.

[33] M. Bland, *An Introduction to Medical Statistics*. Oxford, U.K.: Oxford Medical, 1995.

[34] D. L. G. Hill, C. Studholme, and D. J. Hawkes, "Voxel similarity measures for automated image registration," in *Proc. Visualization in Biomedical Computing (SPIE)*, vol. 2359, 1994, pp. 205–216.