

Tina Memo No. 2004-001
Short version presented at MICCAI 2004
Short version presented at MIUA 2004
Full version submitted to Medical Image Analysis

Empirical Validation of Covariance Estimates for Mutual Information Coregistration

P.A. Bromiley and N.A. Thacker

Last updated
21 / 10 / 2009

This document forms part of the **Statistics and Segmentation Series (2008-001)** available from www.tina-vision.net.

- 2007-008 Tutorial: Defining Probability for Science.
- 2001-007 Performance Characterisation in Computer Vision:
The Role of Statistics in Testing and Design.
- 2002-007 The Effects of an Arcsin Square Root Transform on a Binomial Distributed Quantity.
- 2001-010 The Effects of a Square Root Transform on a Poisson Distributed Quantity.
- 2004-004 Shannon Entropy, Renyi Entropy, and Information.
- 2002-002 Validating MRI Field Homogeneity Correction Using Image Information Measures.
- 2004-001 Empirical Validation of Covariance Estimates for Mutual Information Coregistration.
- 2004-005 The Equal Variance Domain: Issues Surrounding the Use of Probability Densities in Algorithm Design.
- 2009-008 Avoiding Zero and Infinity in Sample Based Algorithms.
- 2001-008 Derivation of the Renormalisation Formula for the Product of Uniform Probability Distributions and Extension to Non-Integer Dimensionality.
- 2001-005 Model Selection and Convergence of the EM Algorithm.
- 2003-007 Noise Filtering and Testing for MR Using a Multi-Dimensional Partial Volume Model.
- 2002-004 A Novel Method for Non-Parametric Image Subtraction:
Identification of Enhancing Lesions in Multiple Sclerosis from MR Images.
- 2001-014 Bayesian and Non-Bayesian Probabilistic Models for Image Analysis.
- 1997-001 The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data.
- 1999-001 The Bhattacharyya Measure requires no Bias Correction.
- 1999-004 B-Fitting: An Estimation Technique With Automatic Parameter Selection.
- 2005-008 Tutorial: Beyond Likelihood.



Imaging Science and Biomedical Engineering Division,
School of Cancer and Imaging Sciences, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Empirical Validation of Covariance Estimates for Mutual Information Registration

Paul A. Bromiley and Neil A. Thacker
Division of Imaging Science and Biomedical Engineering,
School of Cancer and Imaging Sciences,
Stopford Building, University of Manchester,
Oxford Road, Manchester M13 9PT, U.K.
paul.bromiley@manchester.ac.uk

Abstract

The problem of registration for inter-modality clinical volumes is often solved by maximising the mutual information (MI) measure. However, a full characterisation of the registration result requires not only the optimised transformation model parameters, but also an estimate of their covariance matrix. Without this information no quantitative use can be made of the registration result, as the errors on correspondences determined from the result are unknown and their influence on any subsequent image analysis procedures using the result are similarly unknown.

We describe a method for estimating the covariances of the transformation model parameters in MI registration, based on comparisons with standard maximum likelihood. Calculation of the minimum variance bound then yields the minimum achievable error. Such estimates can also be used to validate the implementation of the registration algorithm, through comparison with the covariances observed in practice, measured using Monte-Carlo simulations. The accuracy of the covariance estimation technique was confirmed using affine registrations between MR image volumes of the normal brain, with both simulated and clinical data. We conclude with some observations on the origins of the MI measure and its use in deformable registration.

1 Introduction

The aim in medical image registration is to spatially align two images, or image volumes, to achieve correspondence of anatomy and/or function and thus allow direct regional comparisons to be made. One image or volume, the source, is manipulated with a geometrical transformation model in order to align it to a second, fixed image or volume, the target. The transformation model may either be a global (e.g. rigid or affine) transformation, where the goal is to produce a general alignment of the coordinate system in which the images are embedded, or a more complex deformable transformation model (or “warp field”), where the goal is to produce an alignment at the voxel level¹. Any registration algorithm thus requires three components: the transformation model, some measure of image similarity (the similarity measure) and some optimisation routine to optimise the similarity measure with respect to the transformation model parameters. Although a wide range of similarity measures have been proposed (e.g. [21]), most recent research effort has been focused on information-theoretic measures, and in particular on the MI measure proposed independently by [35] and [9].

The MI measure $\mathcal{I}(I; J)$ for a pair of images or volumes I and J is the Kullback-Leibler divergence between their joint probability distribution $p(i, j)$ and the product of their marginal distributions $p(i)$ and $p(j)$ [10]

$$\mathcal{I}(I; J) = \sum_{i,j} p(i, j) \ln \frac{p(i, j)}{p(i) \cdot p(j)} \quad (1)$$

In principle $p(i)$ and $p(j)$ could be multi-dimensional distributions of any set of independent, transformation-invariant image descriptors: image intensities are most common in practice. In the case of complete independence of the images, the joint distribution would be identical to the product of the marginal distributions, and so maximising the MI measure is equivalent to maximising the dependence of the images. The MI measure can also

¹For reasons of clarity we adopt the terms “global” to specify registration using transformations of the first type, which consist of global translations, scalings etc., and “deformable” for registration using transformation models of the second type, where the transformation is determined locally.

be expressed in terms of the Shannon entropy [31]

$$H(P) = -\sum_{i=1}^N p_i \ln p_i \quad \Rightarrow \quad \mathcal{I}(I; J) = H(I) + H(J) - H(I, J) \quad (2)$$

where $H(P)$ represents the Shannon entropy of a sample $P = p_{i=1\dots N}$. The MI measure as expressed in Eq. 2 can be implemented simply by producing a joint histogram for a pair of images or volumes, normalising to produce a (discrete) joint probability distribution, and summing the Shannon entropy across the bins and marginals. Registration can then be performed by iteratively optimising this measure. The assumption inherent in the use of MI is that, for the allowed class of transformations, manipulation of the transformation model parameters to maximise the MI achieves the best correspondence.

It has long been recognised that, in addition to the optimised transformation model parameters, an estimate of the error on the registration result is required (e.g. [12]). A wide variety of approaches to error estimation have been described in the literature, the majority of which can be divided into four broad classes. First, a known transformation can be applied to one of a pair of well-aligned images or volumes, registration applied to bring them back into alignment, and the result compared to the applied transformation (e.g. [29, 37]). The drawback of this approach is that the applied transformation may not realistically reflect the true transformations encountered in clinical practice e.g. the application of a smooth warp field will not reflect the problems introduced by the appearance of new tissues, such as tumours. Furthermore, if simulated medical data sets are used in order to ensure that the images or volumes are well aligned to begin with, a further departure from clinical reality is introduced; such data typically do not include the whole range of imaging artefacts encountered in practice. Second, fiducial markers can be implanted in order to provide fixed landmarks, and the accuracy of their alignment measured (e.g. [14] and references therein); however, this process is highly invasive and cannot be applied retrospectively. Third, manually or automatically identified image structures, such as anatomically distinct points (e.g. [15]), edges (e.g. [11]) or regions (e.g. [17, 29]) can be used in place of fiducial markers. However, registration errors may then be overestimated due to the additional contribution from errors in structure delineation. Furthermore, in the case of inter-subject registration, exact homology may not exist. Fourth, measures of image correspondence e.g. residual intensity differences following registration, can be applied (e.g. [29]). However, there is no guarantee that image correspondence and structure correspondence are equivalent [29], and the task of defining image correspondence in multi-modality data may be challenging.

The registration errors can be divided into two components, the inability of the transformation model to describe the true transformation that has occurred between the images, and the error on the estimated transformation model parameters given the specific model adopted. The first component can in theory be eliminated if the transformation model is sufficient, and this may be achievable in practice in some situations; for example, in registration of CT and MR brain images of a single subject acquired on the same day the transformation is dominated by the positioning of the subject within the scanners, and so can be accurately modelled using an affine transformation². The same is becoming true in situations where deformable transformation models are required, as research into biomechanical modelling results in transformation models that can accurately model the true transformation (e.g. [8]). The ultimate limit on registration accuracy is therefore dictated by the extent to which the image noise destabilises the cost function i.e. the variation between the transformation model parameters that would be estimated from ideal, noise-free versions of the images, and those actually obtained from the noise-contaminated data.

The aim of the work described here, and in previous conference publications [6, 7], was to produce an estimate of the lower bound on the registration errors, i.e. that imposed by the destabilisation of the registration cost function by image noise, for MI-based registration. The approach was based on expressing the MI measure as a maximum likelihood estimator of the transformation model parameters, allowing the application of a standard technique, namely the calculation of the minimum variance bound (MVB), to find a lower bound on the errors. This bound is estimated directly from the cost function, and so does not require the use of simulated transformations, fiducial markers, or manually or automatically delineated image features. The approach thus forms a novel, statistical class of error estimation techniques and does not suffer the shortcomings listed above. [2] have described the application of a similar approach to a least-squares estimate of intensity similarity, with the intention of applying the method retrospectively to the results of registrations performed using arbitrary similarity metrics; this work is discussed further in Section 4.

The remainder of the paper is structured as follows. Section 2.1 describes the development of the error estimation technique; Section 2.2 describes its empirical validation through comparison to the errors achieved in practice in a representative image registration task, measured using Monte-Carlo simulations. The registration task chosen for the experiments involved MR images of the brain, and both clinical and simulated data sets were used. For

²Fluctuations in scanner gradients between scans can result in changes in voxel dimensions at the percentage level [19], preventing the use of a rigid transformation.

simplicity, the experiments were limited to registering images of the same subject, in which the transformation present between the images could be accurately modelled using a nine-parameter affine transform, thus avoiding the situation in which the transformation model is incapable of modelling the true transformation. The Monte-Carlo studies were designed to explore the dependence of the transformation model covariances on image noise, and so were performed by adding independent noise fields to the original MR images. In order to obtain sufficient accuracy, 1000 simulations were performed at each of 9 levels of added noise for both the simulated and clinical data. Section 3 presents a detailed comparison of the estimated and measured covariances, both in terms of their absolute values and their functional dependence on image noise. The results demonstrate that, in circumstances where the transformation model is capable of modelling the true transformation between the images and the registration algorithm is carefully implemented (with regard to numerical stability issues) it is possible to achieve the lower bound and thus use the proposed technique as a direct estimate of registration error. We conclude with some observations on the general applicability of the proposed technique, and the implications of the link between MI and maximum likelihood.

2 Method

2.1 The Covariance Matrix of MI Registration

Since MI registration involves an optimisation, there is in general no closed-form solution for the transformation model parameters to which the usual equations of error propagation [3] can be applied. Therefore, the conventional approach to covariance estimation would be to perform a Monte-Carlo analysis [27]. However, if the MI measure is expressed in the form of a likelihood, then the minimum achievable errors on the transformation model parameters can be calculated from the MVB [3]. This procedure has two advantages. First, the covariances achieved in practice, measured using Monte-Carlo techniques, can be compared to the MVB in order to validate the implementation of the algorithm: failure to reach the MVB can highlight numerical instability in the implementation. Second, expressing the MI measure in terms of conventional statistical measures can provide new interpretations of the theoretical origins of the measure.

The first stage in the derivation of MVB error estimates for MI involves identifying a maximum likelihood estimator of the transformation model parameters that is equivalent to MI. Starting from Eq. 1 and following [28], the MI measure for a pair of images or volumes I and J can be rewritten as

$$\mathcal{I}(I; J) = \sum_i p(i) \ln \frac{1}{p(i)} + \sum_{i,j} p(i, j) \ln \frac{p(i, j)}{p(j)} \quad (3)$$

Recognising that the first term on the R.H.S. is the entropy $H(I)$ of image I and that $p(i, j) = N_{ij}/N$, where N_{ij} is the number of entries in histogram bin (i, j) and N is the total number of entries in the histogram, gives

$$\ln P(I|J) = \sum_v \ln \frac{p(i, j)}{p(j)} = N[\mathcal{I}(I; J) - H(I)] \quad (4)$$

where v represents a sum over voxels rather than histogram bins. Without loss of generality let I be the target image and J the source image. In addition, following the suggestion in [28], we can insist that the region of data sampled from the target image does not change. This can be enforced in practice by excluding an appropriately sized border around the target image, so that the whole target image is contained within the source image for any values of the transformation model parameters explored during the optimisation. Under this condition $H(I)$ will be a constant, giving

$$\ln P(I|J) = \sum_v \ln \frac{p(i, j)}{p(j)} = N(\mathcal{I}(I; J)) + \text{const.} \quad (5)$$

Therefore, MI is a monotonic function of the log-probability of the target image I given the transformed source image J , and so maximisation of the MI is equivalent to a maximum-likelihood estimator of the transformation model parameters, where I plays the role of the data and the transformed source image J plays the role of the model being fitted to the data.

The MVB for a maximum likelihood estimator is given by the expectation value of the second derivative of the likelihood function L at the optimum [3]

$$C_{\theta_{m,n}}^{-1} \geq \left\langle \frac{\partial \ln L}{\partial \theta_n} \frac{\partial \ln L}{\partial \theta_m} \right\rangle = - \left\langle \frac{\partial^2 \ln L}{\partial \theta_m \partial \theta_n} \right\rangle = - \left. \frac{\partial^2 \ln L}{\partial \theta_m \partial \theta_n} \right|_{\theta_0} \quad (6)$$

where $C_{\theta_{m,n}}$ represents a term in the covariance matrix, $\boldsymbol{\theta}$ represents the vector of transformation model parameters, and $\boldsymbol{\theta}_O$ represents the parameters at the optimum. An important distinction between the processes of parameter estimation (i.e. optimisation) and error estimation in maximum likelihood techniques must be made at this point. Any normalisation of the likelihood function that does not affect the position of the optimum can be ignored during optimisation, since it will not affect the result. However, the MVB depends on the shape of the curve about the optimum, and therefore such normalisation terms must be taken into account. This situation arises in connection with MI registration. The simplest way to confirm this is to substitute the $P(I|J)$ term directly into the expression for the MVB; the resulting error estimates are dependent on the resolution of the histogram even in the limit of infinite statistics, as shown in Eqs. 18 and 19, and thus cannot be correct.

Having identified the normalisation issue, the next stage of the derivation is to identify the correct normalisation term for the $P(I|J)$ term in Eq. 5 such that it can be substituted into Eq. 6 to obtain the MVB errors. In order for a likelihood L' to be properly normalised, such that it can be used in this way, $-2L'$ must asymptotically obey a χ^2 distribution³ i.e.

$$\ln L' = -\frac{\chi^2}{2} \quad (7)$$

(the χ^2 in this equation is sometimes referred to as the ‘‘likelihood chi-square’’ [1]). This is a consequence of a standard statistical result known as the likelihood ratio test theory [30]. The correct normalisation term can then be identified by noting that the condition for achieving the MVB is that the likelihood function is Gaussian or, equivalently, the log-likelihood is quadratic [32]. Therefore, a Gaussian likelihood function

$$L(\boldsymbol{\theta}) = Ae^{-\frac{1}{2}[(\boldsymbol{\theta}-\boldsymbol{\theta}_O)C_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_O)]} \quad (8)$$

can be substituted into any expression for the MVB without loss of generality (i.e. this is not an assumption that the likelihood function for any specific problem actually is Gaussian). The MVB is calculated from the log-likelihood,

$$\ln L = -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_O)C_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_O) + \ln A \quad (9)$$

where A is the normalisation term. The first term on the R.H.S. is familiar as one-half of the χ^2 [13]. Therefore, there are two possible routes to obtain the correct normalisation. The first is to ensure that A is constant w.r.t. the parameters: it then disappears upon differentiation i.e. does not affect the shape of the likelihood function around the minimum (this is the case in simple maximum-likelihood techniques such as linear least-squares fitting). The second is to normalise the probability distribution of the data to its peak, rather than its area. A then becomes 1 and disappears upon taking logs, and so this ‘‘peak-normalised’’ likelihood obeys Eq. 7. The MVB (Eq. 6) can then be re-written as [13]

$$C_{\theta_{m,n}}^{-1} \geq \frac{1}{4} \frac{\partial \chi^2}{\partial \theta_m} \frac{\partial \chi^2}{\partial \theta_n} \bigg|_{\boldsymbol{\theta}_O} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial \theta_m \partial \theta_n} \bigg|_{\boldsymbol{\theta}_O} \quad (10)$$

The χ^2 can be expressed as a sum over individual data terms χ_v , where v represents each data point

$$\chi^2 = \sum_v (\chi_v)^2 \quad (11)$$

The quantity χ_v is sometimes called the ‘‘ χ of the χ^2 ’’. Then, using the first term on the R.H.S. of Eq. 10 and expanding the derivative of χ_v^2 w.r.t. χ_v using the chain rule gives, in the full matrix form,

$$C_{\boldsymbol{\theta}}^{-1} \geq \sum_v 2(\nabla_{\boldsymbol{\theta}} \chi_v)^T \otimes (\nabla_{\boldsymbol{\theta}} \chi_v) \bigg|_{\boldsymbol{\theta}_O} \quad (12)$$

The penultimate stage of the derivation is to apply the peak-normalisation technique to MI in order to identify the equivalent, properly normalised likelihood. The MI measure can be split into two terms

$$NI(I; J) = \sum_v \ln \frac{p(i, j)}{p(i)p(j)} = \sum_v \ln \frac{p(i, j)}{p(i_{max}, j)} - \sum_v \ln \frac{p(i)p(j)}{p(i_{max}, j)} \quad (13)$$

where $p(i_{max}, j)$ is the peak value of the joint distribution along a row or column in the joint histogram specified by a particular j . Recognising that

$$\sum_v \ln \frac{p(i, j)}{p(i_{max}, j)} = \sum_v \ln \frac{p(i|j)}{p(i_{max}|j)} \quad (14)$$

³This observation has led some authors, e.g. [4], to argue that χ^2 is a more fundamental statistic than the likelihood, although this view is not widely accepted.

it becomes clear that the first term on the R.H.S. of Eq. 13 is then a likelihood normalised as described above, and the second term is the peak-normalised likelihood that would be obtained in the case of complete independence of the images i.e. in the case where the joint distribution of the images was identical to the product of the marginal distributions. The second differential of the second term w.r.t. the parameters is small compared to that of the first term, which therefore dominates the shape of the optimisation function around the minimum. MI registration is therefore, to a good approximation, a ML estimator of the transformation model parameters. This implies in turn that the application of the MVB to the first term of Eq. 13 will provide a good estimate of the covariances of the transformation model parameters (the validity of this assumption for data-rich, global registration tasks will be demonstrated in Section 3).

The final stage of the derivation is to substitute the likelihood term identified above into the expression for the MVB. Eq. 12, can be expanded using the chain rule in terms of derivatives of the χ of the χ^2 w.r.t. L , L w.r.t. J , and J w.r.t. θ

$$C_{\theta}^{-1} \geq 2 \sum_v \left(\frac{\partial \chi_v}{\partial L'_v} \right)^2 \left(\frac{\partial L'_v}{\partial J_v} \right)^2 (\nabla_{\theta} J_v)^T \otimes (\nabla_{\theta} J_v) \Big|_{\theta_0} \quad (15)$$

Collecting terms, we have

$$L'_v = \frac{p(i, j)}{p(i_{max}, j)} \quad \frac{\partial \chi_v}{\partial L'_v} = \frac{-1}{L'_v \sqrt{-2 \ln L'_v}} \quad (16)$$

$$\frac{\partial L'_v}{\partial J_v} = \frac{1}{p(i_{max}, j)} \left[\frac{\partial p(i, j)}{\partial J_v} - \frac{p(i, j)}{p(i_{max}, j)} \frac{\partial p(i_{max}, j)}{\partial J_v} \right] \quad (17)$$

giving

$$C_{\theta}^{-1} \geq - \sum_v \frac{\left(\frac{\partial p(i, j)}{\partial J_v} - \frac{p(i, j)}{p(i_{max}, j)} \frac{\partial p(i_{max}, j)}{\partial J_v} \right)^2}{2p(i, j)^2 \ln \frac{p(i, j)}{p(i_{max}, j)}} (\nabla_{\theta} J_v)^T \otimes (\nabla_{\theta} J_v) \Big|_{\theta_0} \quad (18)$$

It is apparent at this point that ignoring the normalisation issue described above would result in an invalid expression: the term in the denominator

$$\ln \frac{p(i, j)}{p(i_{max}, j)} \quad \text{would be replaced with} \quad \ln \frac{p(i, j)}{p(j)} \quad (19)$$

In the limit of infinite statistics, the bin size of the joint histogram will not affect the former, since it is a ratio of bin contents. However, the latter allows the estimated covariances to become arbitrarily small, and thus cannot be valid.

Eq. 18 gives the minimum achievable errors for a registration problem. The condition for achieving this minimum is that the likelihood function being optimised is Gaussian. In practical global registration problems the distributions of individual data terms generated during optimisation will not necessarily be Gaussian. However, as the likelihood function is calculated from $\sim 100,000$ voxels, the Central Limit Theorem implies that the total likelihood function should be Gaussian to a good approximation, regardless of the distribution of the individual data terms, and so the MVB should be achievable. It should also be noted that the Gaussian distribution maximises MI for a given mean and covariance matrix [10]; the process of optimising the MI measure will therefore tend to drive the distributions in the joint histogram towards Gaussian shapes.

2.2 Empirical Validation of the Covariance Estimate

Two sets of experiments were performed. In each, the estimated covariances of the transformation model parameters, as given by Eq. 18, were compared to the covariances achieved in practice, measured using Monte-Carlo simulations. Both sets of experiments involved registrations between 3D MR image volumes of the brain. The first used simulated T1- and T2-weighted images obtained from Brainweb [18]. Normal brain image volumes with 0% noise and 0% intensity non-uniformity, consisting of 1 by 1 by 3mm voxels with 56 slices in each volume, were used. A second, identical set of experiments was conducted using T1-weighted and inversion-recovery turbo spin echo (IRTSE) MR image volumes obtained from a normal volunteer, with informed consent and subject to local ethics committee approval. These images consisted of 0.89 by 0.89 by 3.3mm voxels, with 29 slices in each volume. Examples of the data are shown in Fig. 1; all image volumes consisted of axial slices. The T1-weighted image volume was used as the target in both cases. In references to the coordinate system of the images, x is the lateral direction, y the anterior-posterior direction, and z is the inferior-superior direction. Gaussian noise with a standard deviation of 1% of the dynamic range was added to the simulated data to represent the intrinsic noise fields that

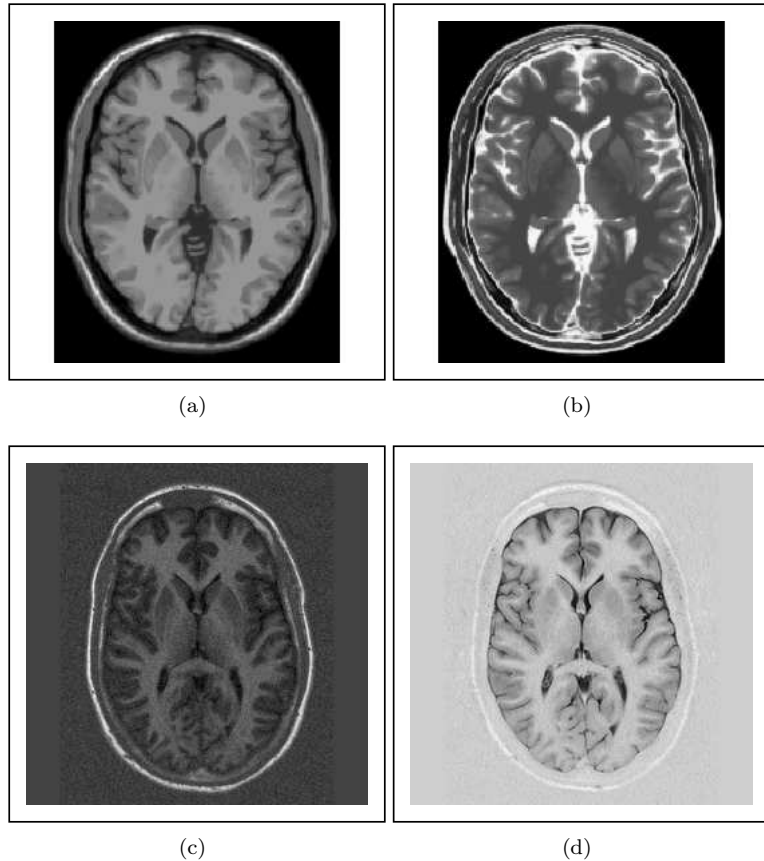


Figure 1: Example slices from the T1- (a) and T2- (b) weighted simulated (Brainweb) MR images, and T1-weighted (c) and IRTSE (d) clinical MR images.

would be present in clinical MR data⁴, so that the simulated and clinical data were as similar as possible. The experiments were performed using a global MI-based registration routine implemented in the open-source TINA software (www.tina-vision.net), using an affine transform consisting of nine parameters i.e. 3D translation \mathbf{T} , rotation \mathbf{R} (represented using Euler angles) and scaling \mathbf{S}

$$\mathbf{P}' = \mathbf{S}[\mathbf{R}.\mathbf{P}] + \mathbf{T} \quad (20)$$

where \mathbf{P} are coordinates in the original source image space, relative to the centre of the image, and \mathbf{P}' are the transformed coordinates. Both data sets were chosen such that the transformation between them could be modelled accurately using a nine-parameter affine transform, so that the ultimate, noise-based limit on the registration accuracy could be achieved. The simulated data were generated from the same anatomical model and so no initial transformation was present; the clinical data were acquired from a single subject/scanning session, and so the only transformation present should be due to subject motion.

An accurate registration algorithm was required in order to provide a realistic comparison between the estimated and measured covariances, and so careful attention was paid to numerical stability issues in the implementation. Of particular significance is the interpolation required to resample the source image on the voxel grid of the target image, such that pairs of intensities entered into the joint histogram are generated from the same spatial positions in both. Interpolation may change the amount of dispersion in the joint histogram e.g. trilinear interpolation applies local averaging and thus reduces the dispersion, increasing the MI. This effect will be absent at points during the optimisation where the voxel grids align over all or part of the images. As a result, oscillatory artefacts, first identified by [25] and demonstrated in Fig. 2 for a MI-based registration algorithm using trilinear interpolation, can be observed in the MI as the voxel grids shift in and out of phase. The amplitude of the oscillation is inversely related to the statistical power of the data [25], and so in this case is greatest in the z direction i.e. the interslice direction, where the voxel dimension is approximately three times larger than in x, y plane. The presence of such oscillations will preclude registration to sub-voxel accuracy and may also introduce problems with local

⁴The noise model in MR magnitude images in practice is Rician [16]; however, the difference between Gaussian and Rician noise is insignificant at signal-to-noise ratios greater than 5. This condition was met for all but 3% of the voxels in these images, so the effects of using Gaussian rather than Rician noise were insignificant.

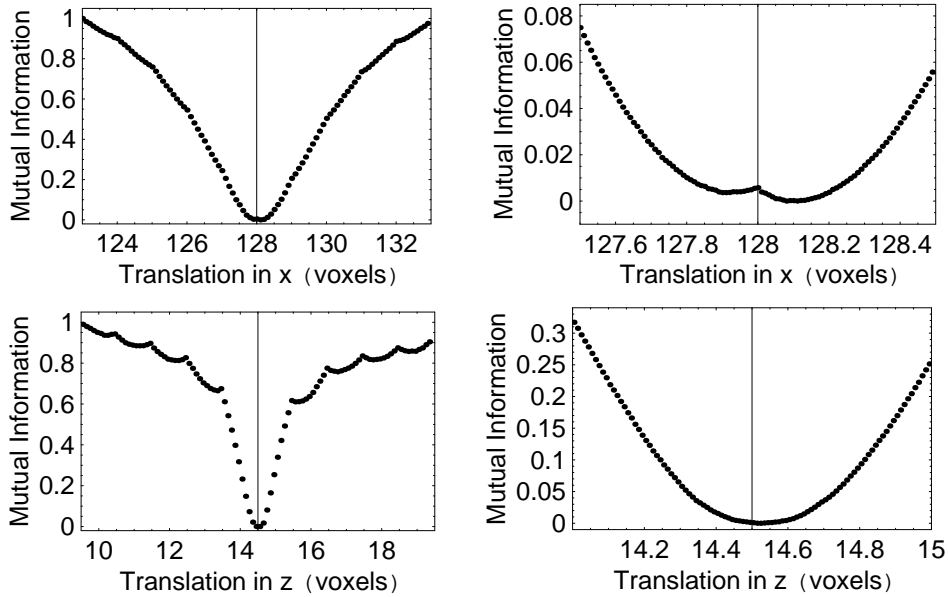


Figure 2: Interpolation artefacts in a naive implementation of MI-based affine registration applied to T1-weighted and IRTSE MR image volumes of the brain of a normal volunteer. The right-hand images show expanded views of the minimum. The MI is given in arbitrary units and has been inverted.

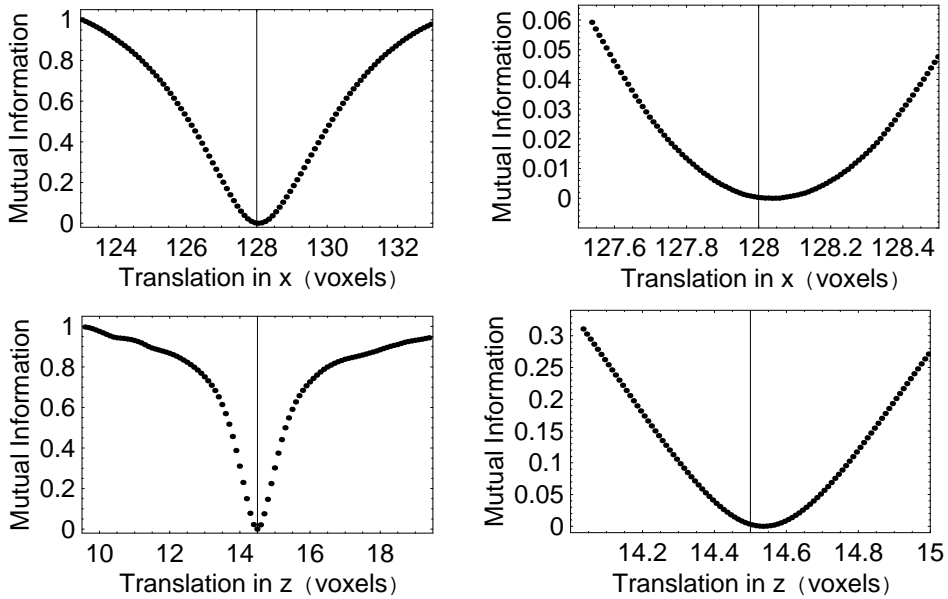


Figure 3: The MI measure in the stabilised implementation of MI-based affine registration applied to T1-weighted and IRTSE MR image volumes of the brain of a normal volunteer. The right-hand images show expanded views of the minimum. The MI is given in arbitrary units and has been inverted. Compared to Fig. 2, the effects of interpolation artefacts have been greatly reduced without changing the minimum or width of the MI function.

minima if a strictly local optimiser, such as simplex [23], is used. In addition, the presence of the local minima would increase the scatter in the results of the Monte-Carlo experiments, increasing the measured variances of the transformation model parameters, whilst also increasing the average gradient of the MI function, reducing the estimated covariance.

In order to avoid such effects, implementation details suggested by [25] were adopted. Both data sets used here were well-aligned to begin with, and all registrations were started from the position of correct alignment in order to prevent contamination of the results with registration failures. A five degree rotation was applied to the source image volume (prior to noise addition and registration), and the images were re-sliced using renormalised sinc interpolation with a $9 \times 9 \times 9$ kernel [33]. This avoided exact alignment of the voxel grids, and thus the introduction of a large interpolation artefact, close to the optimum of the registration cost function: it is not anticipated that this

will be necessary in general medical registration tasks, where the images are not well aligned prior to registration. During the registration itself, renormalised sinc interpolation with a $5 \times 5 \times 5$ kernel was used. Sampling theory implies that this is the optimal interpolation routine [26]. In addition, the bin width of the histogram was set to 1σ in each direction⁵, where σ represents the standard deviation of the image noise, measured using the techniques described by [24]. It was also necessary to ensure that the sample of voxels drawn from the target image remained constant. Therefore, the procedure described by [28] was adopted; this involved defining a border around the target image of sufficient size (5 voxels in this case) to ensure that the source image always overlapped the whole of the target image for any values of the transformation model parameters explored during the optimisation. Gaussian smoothing was also applied to the joint histogram in order to stabilise it. The effects of these implementation details can be seen in Fig. 3, which shows the MI measure for the same registration problem as Fig. 2 using the stabilised algorithm. The effects of interpolation artefacts have largely been eliminated, without either biasing the position of the minimum or increasing its width.

The Monte-Carlo simulations were performed by repeatedly registering the source and target images after addition of Gaussian noise fields with standard deviations of 1σ to the target images and $n\sigma$ to the source images, where σ represents the standard deviation of the intrinsic image noise and n was varied from 0.25 to 2.25 in steps of 0.25. One thousand registrations were performed at each of these 9 noise levels, and covariance matrices were calculated for the transformation model parameters. Distinct local minima were observed in the results at higher noise levels, and these were removed so that the covariances were measured only from the global minima; this typically amounted to removal of $\approx 10\%$ of the results in the experiments for $n \geq 1.5$. Eq. 18 was then used to estimate the covariances of the parameters at each noise level (and, in addition, at zero added noise i.e. with only the intrinsic image noise present). Since this relies on normalisation to the peaks of the distributions in the joint histogram, rather than their areas, it is an inherently less stable process than the calculation of the MI. In addition, processes such as Gaussian smoothing of the joint histogram could not be applied as Eq. 18 implicitly estimates the noise from the width of the distributions; Gaussian smoothing would change this without changing the shape of the MI cost function, which is dictated by the step-to-step change in the histogram during optimisation rather than the shapes of the distributions in the histogram at any given step⁶. Therefore, to increase the stability of the covariance estimation process the bin size of the joint histogram was reduced to 0.5σ and iterative tangential smoothing [34], which smooths along a tangent to the maximum direction of local slope and thus does not significantly broaden the distributions, was applied. All other implementation details remained identical to those used in the Monte-Carlo experiments. Furthermore, the $\partial\theta$ corresponding to three times the range of the results of the Monte-Carlo experiments was found, and covariance estimates were calculated at 100 points across this range; median filtering was then applied to calculate the final covariance estimates.

3 Results

For a given system of model parameters with n components, the covariance matrix will contain n^2 components. However, since it has only n degrees of freedom, a full characterisation of the extent to which the observed and estimated covariances match can be obtained through comparing only n components of the matrix. Clearly, the n components comprising the diagonal elements of the matrix, the variances, are the most informative and so analysis of the results is limited to the variances alone. Figures 4 and 5 show the estimated and measured (via. the Monte-Carlo experiments) errors on the parameters of the transformation model for nine-parameter affine registration of the simulated and clinical MR data, plotted against the level of added noise. It should be noted that these are not errors on the alignment of individual voxels: error propagation would have to be applied in order to calculate such errors from the measurements given here. It should also be noted that the original images, in both sets of experiments, contained a fixed, intrinsic noise field i.e. the acquisition noise, as well as the independent noise fields added prior to each registration. This intrinsic noise does not affect the Monte-Carlo results: repeated registration of the same images, with fixed noise, would always result in the same transformation model parameters to within machine precision. In contrast, the error estimation technique implicitly estimates the image noise from the dispersion in the joint histogram, and so takes account of both the intrinsic and added noise. A correction for this was applied by adding 1 in quadrature to the x-axis coordinates of the estimated errors in Figs. 4 and 5.

A linear dependence between the standard deviations of the added noise and the errors on the parameters would be expected from simple error propagation, and the estimated errors follow this relationship well; for clarity, linear least-squares fits to the measured errors are shown in place of the raw data. Since the registration problem is

⁵This high resolution was possible because the experiments were performed on global registration problems, and so the entire image pair was represented in the histogram for all parameters; it ensured that the distributions in the joint histogram were represented accurately.

⁶This is directly equivalent to the error scale invariance of least-squares techniques i.e. as long as the data are i.i.d, error weighting can be omitted from least-squares measures, as it does not affect the results.

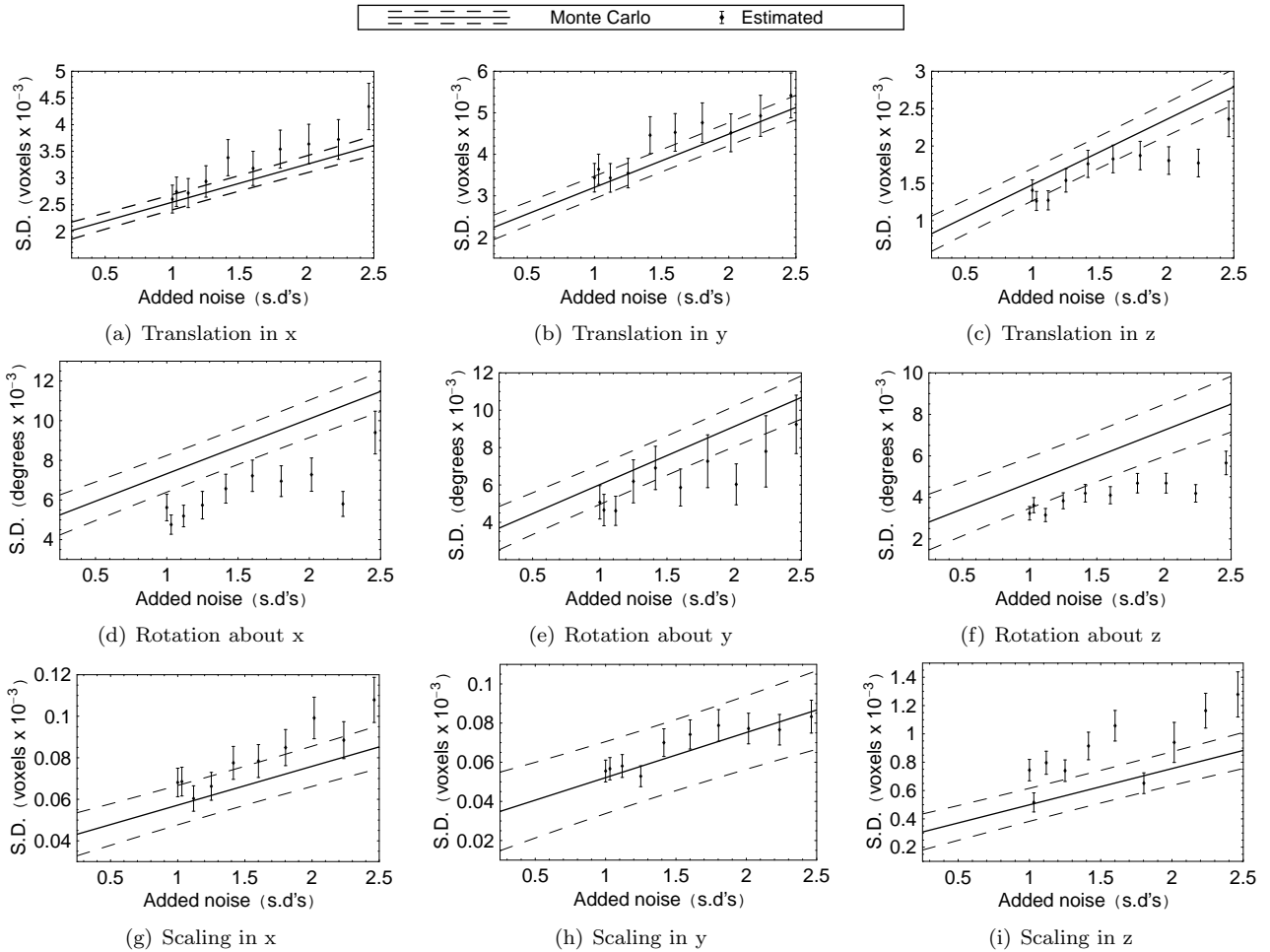


Figure 4: The standard deviations of the registration parameters for the Brainweb data, plotted against the level of added noise as multiples of the intrinsic image noise. The solid lines show linear least-squares fits to the Monte-Carlo results and the dashed lines show the upper and lower 1σ error bounds.

approximately equivalent between the two experiments, but the simulated data contained approximately twice as many voxels as the clinical MR data, the absolute values of the variances should be approximately twice as large in the clinical MR experiments, and this is indeed seen in the results. The estimated error on the scaling parameter in the z direction of the clinical MR data shows a distinct instability. This is due to the low information content along this axis of the data, which represented the inter-slice direction of the MR volume. As stated above, effects like interpolation artefacts become more significant as the information content of the data falls, destabilising the registration cost function and in turn destabilising the covariance estimation process, which operates by estimating the gradient of the cost function. All other parameters show good agreement between the estimated and measured errors, to within the noise on the data.

A more quantitative comparison between the measured and estimated errors was then performed, in two stages. First, the absolute values were compared, testing for any correlated offset between the two sets of data. The presence of a fixed noise field across all experiments, as described above, might be expected to introduce some correlated offset into the estimated errors, and there is some indication of such correlated differences between the estimated and measured errors in Figs. 4 and 5. Fortunately, since these are measurements of errors, rather than parameters, a definitive meaning can be placed on such differences. They relate directly to the confidence limits on the results of any hypothetical image analysis method using registration as a component. Two definitions of statistically significant difference are in common usage. In comparisons of data to theory, or to data from an alternative measurement technique, via a standard method such as a paired T-test, the probability of identity between the two sets of measurements must be lower than either 5% (i.e. 1.96σ difference) or 1% (i.e. 2.576σ difference). Therefore, for two sets of results that are identically distributed (i.e. the expectation value of the difference is 1σ), the errors must be underestimated by at least a factor of two to produce a test statistic that erroneously indicates significant difference. This relationship has led to the general rule-of-thumb that any error estimate that is correct to within a factor of two is statistically usable. The averaged difference between the

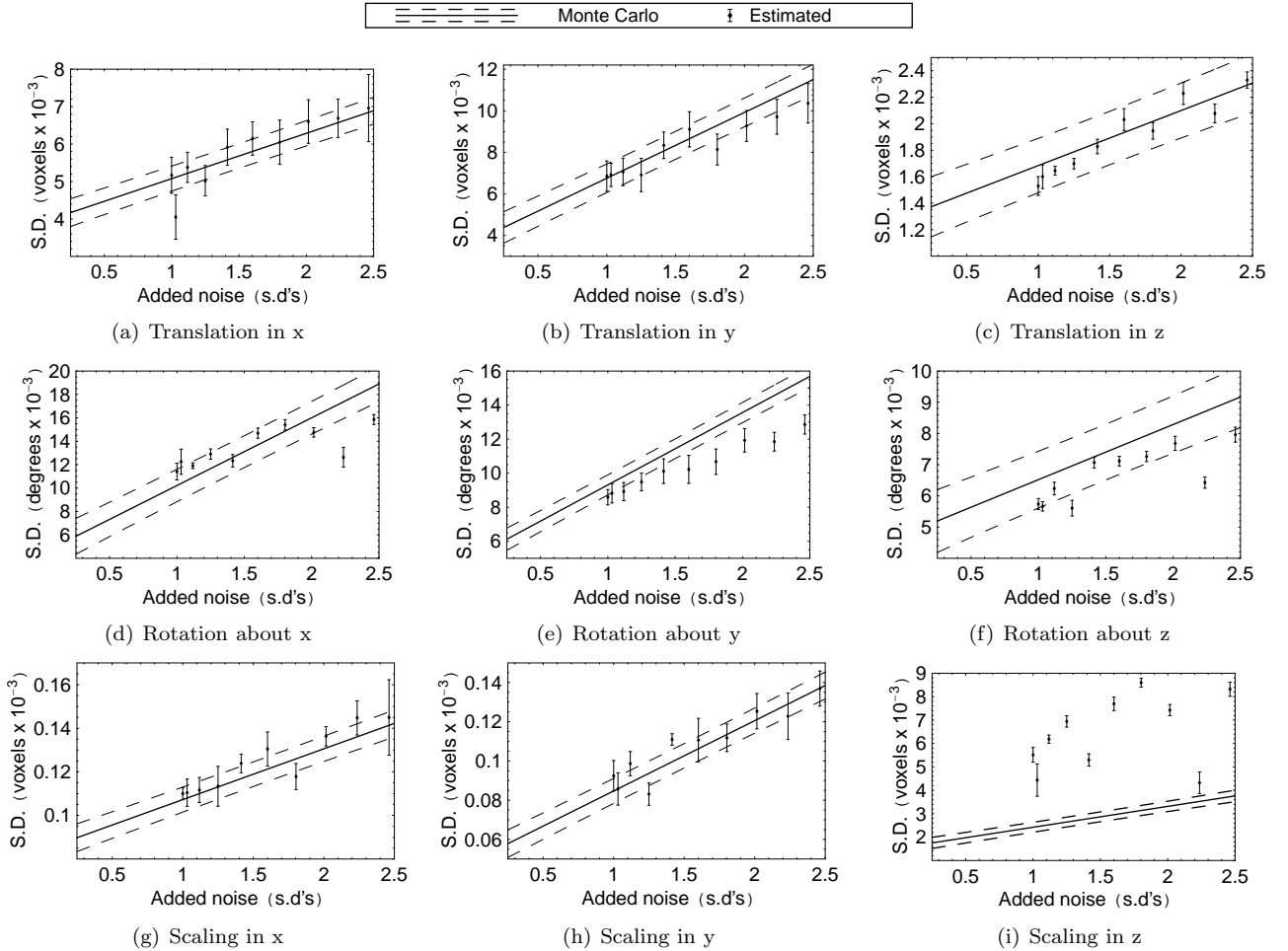


Figure 5: The standard deviations of the registration parameters for the clinical MR data, plotted against the level of added noise as multiples of the intrinsic image noise. The solid lines show linear least-squares fits to the Monte-Carlo results and the dashed lines show the upper and lower 1σ error bounds.

Parameter	Brainweb			Clinical		
	x	y	z	x	y	z
Translation	0.40	0.54	0.28	0.43	0.23	0.77
Rotation	0.36	0.52	0.28	0.13	0.11	0.52
Scaling	0.29	0.98	0.53	0.86	0.49	0.66

Table 1: Student’s T-test probabilities of agreement between the gradients of the Monte-Carlo and estimated errors as a function of image noise, as shown in Figs. 4 and 5.

observed and estimated error measurements is 0.60σ for the simulated and 0.33σ for the clinical MR data and so, despite any offset, the error estimation technique still gives results that are well within the bounds of statistical utility.

In order to compare the absolute values of the estimated and measured errors on the registration parameters in a more quantitative fashion, Bland-Altman plots [5] were produced, and are shown in Fig. 6. These show the differences between the two measurements, in units of the standard deviation of the difference: in this case the sum in quadrature of the standard deviations of the estimated and measured transformation model parameters (i.e. the errors on the error measurements themselves, computed from the residuals about linear fits to the results), plotted against the average of the two measurements. None of the groups of translation, rotation and scaling parameters fall consistently outside the 95% confidence limits for either the simulated or clinical data. Therefore, any correlated differences between the estimated and measured errors are within the noise i.e. there is no evidence of a statistically significant difference between the absolute values of the estimated and measured errors.

A second, and in some ways more informative, comparison between the measured and estimated errors was per-

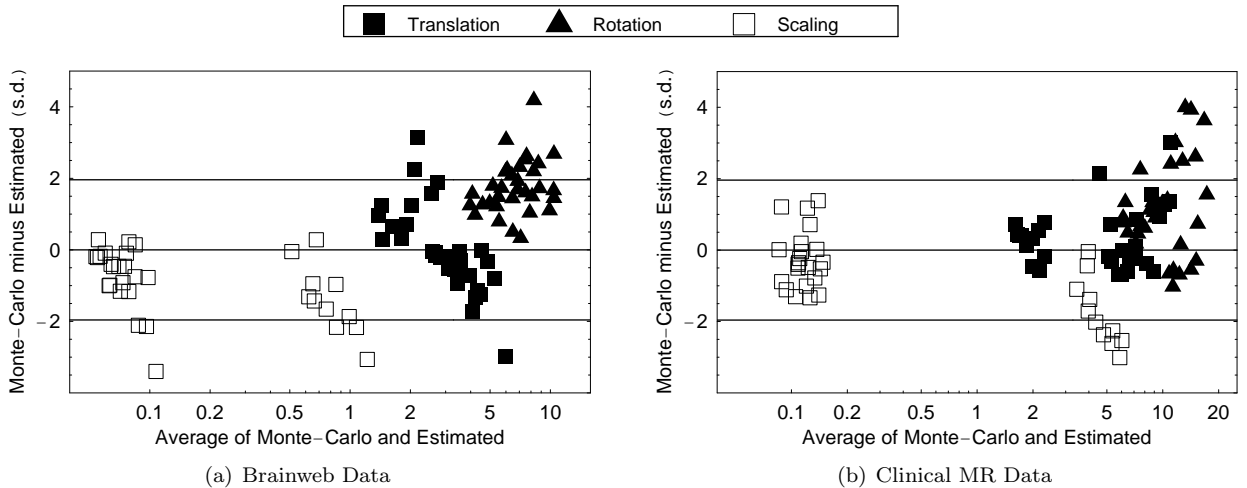


Figure 6: Bland-Altman plots of the estimated and measured (Monte-Carlo) errors on the parameters of the nine-parameter affine transformation model for registration of the Brainweb (a) and clinical (b) MR images. The upper and lower 95% confidence bounds are shown. The units on the abscissa are voxels $\times 10^{-3}$ for the translation and scaling parameters, and degrees $\times 10^{-3}$ for the rotation parameters.

formed by comparing their dependence on the added noise i.e. the gradients of linear fits to the errors as a function of the added noise. Any consistent difference in the gradient would indicate a functional difference between the error estimation technique and the Monte-Carlo results in terms of their dependence on the image noise. Standard errors on the linear fit parameters were provided by the fitting package used, allowing the construction of a T-statistic [3] for the difference between the gradients of the estimated and measured errors for each transformation model parameter. Finally, Student’s T-test probabilities were computed for these T-statistics, and are given in Table 1. These probabilities would have to fall below the 5% limit in order to demonstrate a significant difference between the gradients: none do, indicating that there is no significant functional difference between the error estimation technique presented here and the actual errors on the transformation model parameters achieved in practice, as measured by the Monte-Carlo experiments, to within the accuracy of these experiments.

4 Discussion

MI has become the most popular similarity measure used for inter-modality registration of clinical images or image volumes in recent years. The measure originated in the field of information theory, and is ultimately derived from Shannon’s concept of entropy. However, the work presented here has demonstrated an alternative interpretation. Through comparison of the measure to standard maximum likelihood, it is possible to express MI as a difference of two terms: the conventional χ^2 of the target image in terms of the source image, and a second term representing the χ^2 in the case of complete independence of the images. Assuming the effects of this second term to be small, the MVB can be applied to the first term in order to obtain the minimum achievable errors on the transformation model parameters for MI registration. Covariances estimated in this way were compared to covariances measured from Monte-Carlo experiments for two medical registration problems, both involving the registration of MR image volumes of the brain, one using simulated data and the other clinical MR scans. Several statistical tests were applied to the resulting errors, and indicated that there were no statistically significant differences between either their absolute values or their functional dependence on image noise. This agreement between the estimated and measured errors confirms the assumption that the second term has little effect on the shape of the likelihood function in the data-rich, global registration problems studied (although it may not be more generally true). We therefore conclude that, in the registration problems studied, it is possible to achieve the MVB and, conversely, to use MVB errors directly as estimated errors on the transformation model parameters.

Achieving the MVB requires careful attention to the implementation of the registration algorithm. This demonstrates one of the most useful aspects of a quantitative statistical approach to algorithm development. The covariance estimation technique calculates the minimum achievable errors on the transformation model parameters. If the errors seen in practice, measured using Monte-Carlo techniques, fail to reach these lower bounds it may indicate that some feature of the implementation is sub-optimal. Further investigation, for instance plotting the cost function using exhaustive search techniques, can reveal these problems, allowing improvements to be made. Iterating this approach until convergence will yield an implementation of the algorithm that is provably

optimal. The converse is also true: agreement between the results of an optimal implementation of the algorithm and the theoretical MVB proves that the quantitative statistical interpretation of the algorithm used to derive the covariance estimate is correct. The prominence of such techniques has grown in recent years, as machine vision algorithms have become ever more complex and so greater care must be taken in the validation of the software. For example, [20] suggest applying algorithms to data for which the distribution of the results can be predicted, and testing for agreement using the Komolgorov-Smirnov test. We believe that, in the present case, comparison of the predicted and observed covariances achieves the same goal.

The ability to achieve the MVB depends on four main factors: the numerical stability of the implementation of MI registration; the suitability of the chosen transformation model for a given registration problem; the ratio of the amount of data to the number of transformation model parameters being estimated; and the type of image data being registered. Significant changes to any one of these factors require that the comparison of MVB error estimates to the results of Monte-Carlo experiments be performed again, in order to confirm the ability of the algorithm to reach the MVB for the particular registration problem at hand. However, once this has been done on a small number of datasets, the MVB error estimates can be used directly as error estimates on subsequent registrations of the same type. The estimated errors on the transformation model parameters, combined with error propagation on the transformation equation, then provide estimated errors on the alignment of individual voxels. The technique described here is therefore of most use in medical image analysis experiments involving large numbers of images.

For completeness it must be noted that some implementations of MI registration, notably that described by [36] use a Parzen window based density method for estimating the joint image probabilities, rather than the histogram-based approach adopted here. The advantage of this approach is that it allows direct evaluation of the derivatives of the probability densities, simplifying optimisation: however, since the absolute values of the cost function are not available, no error estimation at the minimum is possible. Nevertheless, the Parzen window and histogram-based approaches simply represent alternative ways of sampling the same underlying probability densities. Therefore, assuming that numerical instabilities have been eliminated from the calculations, both approaches should yield the same registration result, and the same accuracy. The error estimation technique presented here is therefore valid in either case, although it is clearly easier to implement if the full joint image histogram is available.

In a recent paper, [2] described an approach to registration error estimation similar to the one presented here. They assume Gaussian conditional densities in the joint image histogram and a non-linear functional relationship between the source and target image intensities. This function, applied to the transformed source image, provides a model of the target image, allowing the construction of a least-squares estimator of the transformation model parameters from the differences between this model and the actual target image intensities. They then obtain the covariance matrix of this estimator, and use this to derive errors on the alignment of individual voxels. As stated above, the assumption of Gaussian conditional densities (and therefore a Gaussian likelihood function) has a particular relevance in error estimation, since this is the condition for achieving the MVB. Therefore, the Bansal et al. approach, in the terminology used in this paper, provides a method for estimating the MVB on registration algorithms that optimise functional intensity similarity or a monotonic function of it.

This paper has focused exclusively on global registration problems. However, in many respects deformable registration is the more interesting problem, since the transformation model is determined locally and so the spatial variation of the errors may be much more complicated. In that case, knowledge of the errors is essential in order to perform further quantitative analysis of the image volumes using the transformation model. We can make two observations concerning the application of the technique described here to deformable registration problems. First, two competing effects will be observed. Since less information is used to calculate each transformation model parameter, it may be more difficult to produce stable estimates of the covariances. However, for the same reason, the random errors on each parameter (ignoring any systematic error caused by the inability of the transformation models to describe the true transformation) will typically be much higher, and so the requirement for a stable estimate of the similarity measure at sub-voxel resolutions may be significantly relaxed. Ultimately, the balance between these two effects will be determined by the number of transformation model parameters. Second, the derivation presented here has shown that, to a good approximation, MI is equivalent to a ML estimator of the transformation model parameters. This interpretation of MI is essential if quantitative use is to be made of any MI-based parameter estimation. In general, the behaviour of ML estimators is only guaranteed where the model matches the true functional dependency of the data [3]. In the experiments presented here, the actual transformation between the image pairs is expected to be well described by the nine-parameter affine transform used in the registration, and so this requirement is satisfied. However, in many implementations of deformable registration the transformation model is based on assumptions that are not expected to be valid in detail. Diffeomorphic warp fields [22], for example, impose assumptions of local smoothness and unique one-to-one matches of spatial locations in the two image volumes. Such assumptions will frequently be violated in practical medical image registration tasks where, for example, deformable registration is applied to align MR brain images across populations of sub-

jects with varying cortical topologies. Research into biomechanical modelling may eventually remedy this situation by providing transformation models that accurately describe such effects. However, the interpretation of MI as a ML estimator makes clear the problems that will be encountered when this condition is not met: applying ML estimation when the model is not a good description of the data will result in systematic errors that cannot easily be quantified.

Acknowledgement

This work was funded by the MIAS IRC (Medical Images and Signals Interdisciplinary Research Collaboration), under EPSRC GR/N14248/01, and UK MRC Grant No. D2025/31. The software used is freely available from our website www.tina-vision.net.

References

- [1] S Baker and R D Cousins. Clarification of the use of chi-square and likelihood functions in fits to histograms. *Nucl Instrum Methods*, 221:437–442, 1984.
- [2] R Bansal, L H Staib, A F Laine, D Xu, J Liu, L F Posecion, and B S Peterson. Calculation of the confidence intervals for transformation model parameters in the registration of medical images. *Med Image Anal*, 13(2):215–233, 2009.
- [3] R J Barlow. *Statistics: A Guide to the use of Statistical Methods in the Physical Sciences*. John Wiley and Sons Ltd., UK, 1989.
- [4] J Berkson. Minimum chi-square, not maximum likelihood. *Ann Stat*, 8(3):457–487, 1980.
- [5] J M Bland and D G Altman. Statistical method for assessing agreement between two methods of clinical measurement. *Lancet*, 327:307–310, 1986.
- [6] P A Bromiley, M Pokric, and N A Thacker. Computing covariances for mutual information coregistration. In *Proceedings MIUA '04*, pages 77–80, 2004.
- [7] P A Bromiley, M Pokric, and N A Thacker. Empirical evaluation of covariance estimates for mutual information coregistration. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2004*, volume 3216 of *Lecture Notes in Computer Science*, pages 607–614. Springer, 2004.
- [8] T J Carter, C Tanner, W R Crum, and D Hawkes. Prone-supine breast MR image registration for image-guided surgery. In *Proceedings MIUA '05*, pages 2–5, 2005.
- [9] A Collignon, F Maes, D Delaere, D Vandermeulan, P Suetens, and G Marchal. Automated multi-modality image registration based on information theory. In Y Bizais, C Barillot, and R Di Paola, editors, *Information Processing in Medical Imaging*, pages 263–274. Kulwer Academic, Dordrecht, 1995.
- [10] T M Cover and J A Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.
- [11] W R Crum, L D Griffin, and D J Hawkes. Automatic estimation of error in voxel-based registration. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2004*, volume 3216 of *Lecture Notes in Computer Science*, pages 821–828. Springer, 2004.
- [12] W R Crum, L D Griffin, D V G Hill, and D J Hawkes. Zen and the art of medical image registration: correspondence, homology, and quality. *NeuroImage*, 20:1425–1437, 2003.
- [13] G D'Agostini. Bayesian inference in processing experimental data: Principles and basic applications. *Rep Prog Phys*, 66:1383–1422, 2003.
- [14] J M Fitzpatrick, J B West, and C R Maurer Jr. Predicting error in rigid-body point-based registration. *IEEE Trans Med Imaging*, 17(5):694–702, 1998.
- [15] I D Grachev, D Berdichevsky, S L Rauch, S Heckers, D N Kennedy, V S Caviness, and N M Alpert. A method for assessing the accuracy of intersubject registration of the human brain using anatomical landmarks. *NeuroImage*, 9:250–268, 1999.
- [16] H Gudjbartson and S Patz. The Rician distribution of noisy MRI data. *Magn Reson Med*, 34(6):910–914, 1995.
- [17] P Hellier, C Barillot, I Corouge, B Gibaud, G Le Goualher, D L Collins, A Evans, G Malandain, N Ayache, G E Christensen, and H J Johnson. Retrospective evaluation of intersubject brain registration. *IEEE Trans Med Imaging*, 22(9):1120–1130, 2003.
- [18] R K-S Kwan, A C Evans, and G B Pike. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans Med Imaging*, 18(11):1085–1097, 1999.
- [19] L Lemieux and G J Barker. Measurement of small inter-scan fluctuations in voxel dimensions in magnetic resonance images using registration. *Med Phys*, 25(6):1049–1054, 1998.

- [20] A Liu, T Kanungo, and R M Haralick. On the use of error propagation for statistical validation of computer vision software. *IEEE Trans Pattern Anal Mach Intell*, 27:1603–1614, 2005.
- [21] J Maintz and M Viergever. A survey of medical image registration. *Med Image Anal*, 2(1):1–36, 1998.
- [22] M I Miller, S C Joshi, and G E Christensen. Large deformation fluid diffeomorphisms for landmark and image matching. In A. W. Toga, editor, *Brain Warping*, pages 115–131. Academic, New York, 1999.
- [23] J A Nelder and R Meade. A simplex method for function minimisation. *Computer Journal*, 7:308–313, 1965.
- [24] S I Olsen. Estimation of noise in images: an evaluation. *CVGIP: Graphical Models and Image Processing*, 55:319–323, 1993.
- [25] J P W Pluim, J B Antoine Maintz, and M A Viergever. Interpolation artefacts in mutual information-based image registration. *Computer Vision and Image Understanding*, 77:211–232, 2000.
- [26] J P W Pluim, J B Antoine Maintz, and M A Viergever. Mutual information based registration of medical images: a survey. *IEEE Trans Med Imaging*, 22(8):986–1004, August 2003.
- [27] W H Press, B P Flannery, S A Teukolsky, and W T Vetterling. *Numerical Recipes in C*. Cambridge University Press, New York, 2nd edition, 1992.
- [28] A Roche, G Malandain, N Ayache, and S Prima. Towards a better comprehension of similarity measures used in medical image registration. In *Proceedings MICCAI'99*, pages 555–566, 1999.
- [29] P Rogelj, S Kovacic, and J C Gee. Validation of a non-rigid registration algorithm for multi-modal data. In M Sonka and J M Fitzpatrick, editors, *Medical Imaging 2002: Image Processing*, volume 4684 of *Proceedings of SPIE*, pages 299–307, 2002.
- [30] W A Rolke, A M Lopez, and J Conrad. Limits and confidence intervals in the presence of nuisance parameters. *Nuclear Instrumentation and Methods in Physics Research A*, 551:493–503, 2005.
- [31] C E Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423 and 623–656, Jul and Oct 1948.
- [32] A Stuart and K Ord. *Kendall's Advanced Theory of Statistics Volume 2: Classical Inference and the Linear Model*. Arnold, London, 1994.
- [33] N A Thacker, A Jackson, D Moriarty, and B Vorkurka. Renormalised sinc interpolation. *J Magn Reson Imaging*, 10(4):582–588, 1999.
- [34] N A Thacker, M Pokric, and D Williamson. Noise filtering and testing using a multi-dimensional partial volume model of MR data. In *Proceedings BMVC'04*, pages 909–918, 2004.
- [35] P Viola and W M Wells. Alignment by maximisation of mutual information. In *Proceedings ICCV'95*, page 16, 1995.
- [36] P Viola and W M Wells. Alignment by maximisation of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.
- [37] J West, J M Fitzpatrick, M Y Wang, B M Dawant, C R Maurer Jr, R M Kessler, R J Macuiunas, C Barillot, A Lemoine, A Collignon, F Maes, P Suetens, D Vandermeulen, P A van der Elsen, A Napel, T S Sumanaweera, B Harkness, P F Helmer, D L G Hill, D J Hawkes, C Studholme, J B Antoine Maintz, M AA Viergever, G Malandain, X Pennec, M E Noz, G Q Maguire Jr, M Pollack, C A Pelizzari, R A Robb, D Hanson, and R P Woods. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J Comput Assist Tomogr*, 21:554–566, 1997.