ELSEVIER

Medical Image Analysis 5 (2001) 143–156

MEDICAL IMAGE ANALYSIS

www.elsevier.com/locate/media

A global optimisation method for robust affine registration of brain images

Mark Jenkinson*, Stephen Smith

University of Oxford, John Radcliffe Hospital, FMRIB Centre, Oxford OX3 9DU, UK

Received 9 May 2000; received in revised form 19 September 2000; accepted 8 January 2001

Abstract

Registration is an important component of medical image analysis and for analysing large amounts of data it is desirable to have fully automatic registration methods. Many different automatic registration methods have been proposed to date, and almost all share a common mathematical framework — one of optimising a cost function. To date little attention has been focused on the optimisation method itself, even though the success of most registration methods hinges on the quality of this optimisation. This paper examines the assumptions underlying the problem of registration for brain images using inter-modal voxel similarity measures. It is demonstrated that the use of local optimisation methods together with the standard multi-resolution approach is *not* sufficient to reliably find the global minimum. To address this problem, a global optimisation method is proposed that is specifically tailored to this form of registration. A full discussion of all the necessary implementation details is included as this is an important part of any practical method. Furthermore, results are presented for inter-modal, inter-subject registration experiments that show that the proposed method is more reliable at finding the global minimum than several of the currently available registration packages in common usage.

Keywords: Affine transformation; Global optimisation; Multimodal registration; Multi-resolution search; Robustness

1. Introduction

Registration is an important component in many medical image analysis applications. It is used in motion correction, multi-modal fusion, mapping to Talairach space and many other tasks. When analysing large quantities of data, such as in a clinical study or within a busy imaging unit, it is desirable to have fully automatic registration methods. Such methods aim to offer reliability and repeatability as well as minimising user interaction.

A standard method of solving the registration problem is to treat it as a mathematical optimisation, using a cost (or similarity) function to quantify the quality of the alignment of the two images for any given transformation. In practice, this formulation relies on the use of a global optimisation method. This optimisation method is crucial for obtaining good registrations. However, to date most attention has been focused on other aspects of the problem, such as defining cost functions, rather than on the optimisation method.

Most of the mathematical optimisation methods that exist are only suitable for local optimisation and therefore will not find the global solution in general. These methods include gradient descent, Powell's method, simplex methods and so on (Press et al., 1995). Furthermore, although some global optimisation methods exist (such as Genetic Algorithms and Simulated Annealing), many of the methods require a very large number of iterations to satisfy statistical convergence criteria (Geman and Geman, 1984; Ingber, 1989). It is therefore the speed of the global optimisation methods that is the limiting factor for their successful application to this problem, since evaluation of

*Corresponding author. Tel.: +44-1865-222-739; fax: +44-1865-222-717.

E-mail address: mark@fmrib.ox.ac.uk (M. Jenkinson).

1361-8415/01/$ – see front matter
PII: S1361-8415(01)00036-6

the cost function is particularly time consuming for volumetric registration using voxel similarity measures. Consequently, although some existing global optimisation techniques may be viable (such as Genetic Algorithms) the majority of existing registration methods opt for local optimisation in order to increase the speed.

In an attempt to solve the global optimisation problem in a reasonable amount of time, many registration methods rely on a multi-resolution approach in the hope that local optimisation will then find the global minimum. The assumption is that it is easier to find the global minimum at a coarse resolution (when using large sub-sampling), since larger transformation steps can be taken, which should reduce the chance of there being a local minimum between the initial starting position and the global minimum. This global minimum is then refined by applying successive local optimisations at a series of different resolutions.

This paper examines affine registration of brain images using inter-modal voxel similarity measures such as the Correlation Ratio (Roche et al., 1998) and Mutual Information (Viola and Wells, 1997; Maes et al., 1997). It initially considers the assumptions underlying the registration problem (Sections 2 and 3) and then proposes a fast global optimisation method (Section 4) that is specifically tailored to this registration problem. A full discussion of all the necessary implementation details (Section 5) is given next, and then results are presented (Section 6) that compare the method with several currently available registration packages in common usage. These results show that the proposed method is more reliable at finding the global minimum than the other packages. Furthermore, using the above results, together with different tests which compare various optimisation schemes and cost functions, it is demonstrated that the use of local optimisation methods together with the standard multi-resolution approach is *not* sufficient to reliably find the global minimum.

2. Mathematical formulation

The standard registration problem is to find a transformation that best aligns a reference image I^{r} to another (floating) image I^{f}. This is formulated as a mathematical problem by taking a cost function, $C(I_1,I_2)$, that quantifies the quality of the registration and then finding the transformation T^* which gives the minimum cost: $T^* = \mathrm{argmin}_{T\in S_T} C(I^{\mathrm{r}},I^{\mathrm{f}}\circ T)$ where S_T is the space of allowable transformations, and $I^{\mathrm{f}}\circ T$ represents the image I^{f} after it has been transformed by the transformation T.

It is necessary when beginning to formulate an approach to the registration problem to decide what will be the space of allowable transformations, S_T. One basic class of transformations are the affine transformations which, in three dimensions, have 12 degrees of freedom (DOF). These transformations include rigid body (6 DOF) and similarity (7 DOF) transformations as particular cases. However, even for these simple cases, with a maximum of 12 DOF for the entire volume, the registration problem is still difficult. Therefore this paper examines these affine transformations because they are the simplest, most common transformations used and, furthermore, many non-linear methods rely on having an initial affine fit as a preprocessing step.

2.1. Interpolation

For discrete data, the intensity is normally only defined on a grid, of discrete locations or lattice sites. To evaluate the intensity at intermediate locations requires interpolation. The interpolation can be viewed as reconstructing a full continuous image from the discrete points, although to evaluate the cost function it is usually only necessary to know the intensity at the corresponding lattice sites. Typically, interpolation methods are based on a convolution of the discrete data with some continuous kernel such as trilinear, spline and (windowed) sinc kernels.

One major effect that the choice of interpolation has is in determining the degree that the cost function becomes continuous or discontinuous with respect to changes in the transformation parameters. This is also affected by the boundary conditions used, such as: padding with zeros; using an arbitrary intensity class that is later discarded; or by only using voxels in the overlapping volume. Studying the precise effects of interpolation is an active research area (Hajnal et al., 1995; Pluim et al., 2000; Thacker et al., 1999) but is beyond the scope of this paper. Therefore, in this paper trilinear interpolation is used on the overlapping volume. These choices require no additional parameters to be set and were motivated largely by experience.

2.2. Cost functions and optimisation

For this paper it is the intensity-based inter-modal cost functions that are investigated: for example, Mutual Information, Correlation Ratio, etc. These are difficult to optimise, since there are many non-linear, potentially discontinuous terms involved that result in functions that are non-smooth and irregular. In general, though, all cost functions require global optimisation.

The theoretical registration problem, as posed above, is fully specified by a transformation space, an interpolation method and a cost function. However, in practice, an optimisation method is required to find the transformation that minimises the cost function. It is necessary that the method be capable of finding the *global* minimum rather than one of the more common *local* minima.

Although there do exist general global optimisation methods (Genetic Algorithms and Simulated Annealing being notable examples), it has been shown (in the No Free Lunch theorem (Wolpert and Macready, 1996)) that there is no general method that is superior for all problems.

Therefore the method used should ideally be tuned for the particular problem at hand — in this case, volumetric registration using voxel similarity measures.

The performance of the optimisation method is important since evaluating the cost function requires a large amount of computation. This, together with the general combinatorial explosion which occurs for higher dimensional spaces (for example, $\mathscr{R}^{12}$) makes any simple search algorithm impractical. For example, a search which uses N values per degree of freedom D, with a cubic voxel size of n mm (n-mm resolution) and a fixed FOV, requires a number of calculations that is proportional to N^D/n^3. In our implementation a single cost function at 1-mm resolution takes approximately 1 second to calculate, so that with only $N = 10$ values per parameter, the total search time would be approximately 11 days with six DOF ($D = 6$) and over 10000 years for the full 12 DOF!

To overcome these problems, one common tactic is to adopt a multi-resolution approach in conjunction with a local optimisation method (Woods et al., 1993; Studholme et al., 1996), that is, to start at a coarse resolution of n mm (where the volume is sub-sampled such that the voxel side-length becomes n mm), find a solution with the local optimisation method and then refine the solution by progressively increasing the resolution (decreasing the value of n). For instance, a method may start with $n = 9$ then progress to $n = 3$ and finally to $n = 1$.

An advantage of this multi-resolution approach is that the initial optimisation, at large n, has a dramatically reduced computational load, since the number of sample points is substantially less. In addition, for large sub-samplings it should be the gross features of the image which dominate, and so the overall alignment should be easier to find. This belief is equivalent to saying that there are less local minima for the optimisation method to get caught in, although, as has been shown in (Pluim et al., 2000) that this is not necessarily true.

2.3. Difficulties

The standard formulation for cost function optimisation described above, is based on the assumption:

1. The location of the global minimum of the cost function corresponds to the desired solution.

To use (multi-resolution) local optimisation techniques to solve the problem requires the following additional assumptions.

2. At the maximum sub-sampling the global minimum is the nearest minimum to the starting position and can be found by local optimisation,
3. The location of the global minimum found using one sub-sampling is inside the basin of attraction (as defined by the optimisation method) of the global minimum for the next sub-sampling.

Here the basin of attraction for some minimum is defined as the set of initial transformations that, after successive applications of the local optimisation algorithm, converge to that minimum. This basin depends on the precise optimisation algorithm used and can have very complicated boundaries, often with a fractal nature.

However, these assumptions do not always hold. The following describes some cases where they are not valid.

- If the cost function for some extreme transformation gives a low value then the global minimum will be (at least degenerately) given by this limiting case. For example, large scalings can create low cost values even though the registration is poor. Furthermore, limiting the domain is not a general solution to this as it would be necessary to guarantee that the cost at the edge of the domain was higher than the global minimum value which is unknown.
- Sub-sampling to lower resolutions may not reduce the number of local minima sufficiently. In fact, some work on interpolation (Pluim et al., 2000) has shown how it can actually create additional local minima.
- Minima can move in scale-space, so that the location of a minimum for some sub-sampling may fall inside the basin of attraction of a different minimum for another sub-sampling.

To develop a reliable, automatic registration method it is necessary to examine these assumptions and tailor the method to this problem. In order to do this the characteristics of typical cost functions need to be understood, and these are examined in the next section.

3. Characterisation of cost functions

In this section, the behaviour of five different intermodal cost functions will be examined. The cost functions that will be compared here are: the Woods function (Woods et al., 1993), Correlation Ratio (Roche et al., 1998), Joint Entropy (Studholme et al., 1995; Collignon et al., 1995), Mutual Information (Viola and Wells, 1997; Maes et al., 1997), and Normalised Mutual Information (Studholme et al., 1999; Maes, 1998). Definitions of these functions are given in Table 1. Note that to form the *cost* functions it is necessary that low values represent good registrations, so many definitions are reversed such as $C^{\mathrm{MI}} = -\mathrm{Mutual\ Information}$.

All of these definitions require the intensity values in at least one image to be 'binned'. That is each intensity, I, has a bin number k assigned to it if $I_{k-1} < I \leqslant I_k$ and $\{I_0, I_1, \ldots, I_M\}$ is a partition of the intensities. Then, given this bin number, iso-sets or the joint histogram are easily determined.

3.1. Asymptotic behaviour

In order that the global minimum correspond to the desired registration it is necessary that the cost functions have the correct asymptotic behaviour. That is, for extreme

Table 1
Mathematical definitions of several inter-modal intensity-based cost functions[a]

Cost function	Definition	Minimum	Maximum
C^{W}	$\Sigma_k \frac{n_k}{N} \frac{\sqrt{\mathrm{Var}(Y_k)}}{\mu(Y_k)}$	0	∞
C^{CR}	$\frac{1}{\mathrm{Var}(Y)} \Sigma_k \frac{n_k}{N} \mathrm{Var}(Y_k)$	0	1
C^{JE}	$H(X,Y)$	0	∞
C^{MI}	$H(X,Y)-H(X)-H(Y)$	$-\infty$	0
C^{NMI}	$\frac{H(X,Y)}{H(X)+H(Y)}$	0	1

[a] The notation used is defined by: quantities X and Y are images, each represented as a set of intensities; $\mu(A)$ is the mean of set A; $\mathrm{Var}(A)$ is the variance of the set A; Y_k is the kth iso-set defined as the set of intensities in image Y at positions where the intensity in X is in the kth intensity bin; n_k is the number of elements in the set Y_k such that $N = \Sigma_k n_k$; $H(X,Y) = -\Sigma_{ij} p_{ij} \log p_{ij}$ is the standard entropy definition where p_{ij} represents the probability estimated using the (i,j) joint histogram bin, and similarly for the marginals, $H(X)$ and $H(Y)$.

(and hence erroneous) registrations they should return high cost values, otherwise the first assumption outlined in Section 2.3 is invalid. In this section two asymptotic registrations are investigated: large translations where the volume overlap is minimal; and large scale disparity where a small portion of one volume is stretched to cover the entire other volume.

When examining the asymptotic behaviour it is important to define what the boundary conditions are. That is, what is done for points in space which do not lie in one or other of the volume domains. One option is to (conceptually) pad the volumes with zeros to give them infinite domain. However, this creates artificial intensity boundaries when the FOV only includes part of the head (for example, when the top few mm of the head/brain are not included the intensity suddenly drops from that of brain matter to zero) – which is relatively common. These artificial boundaries would then bias the registration, which is undesirable. Therefore, better alternatives are to either use an arbitrary intensity class that is later discarded or to calculate the cost function only for the region where the volumes overlap. The only difference between these options is whether the total number of voxels used to normalise the calculations remains constant, as in the former option, or varies with the size of the overlapping volume, as in the latter option. It is apparent that these alternatives should provide less biased registrations; however, a proof of unbiasedness remains a topic for future work.

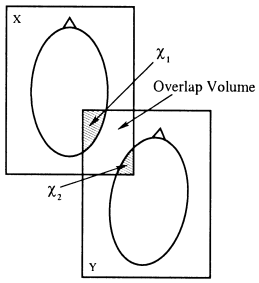


Fig. 1. Example of an asymptotically large transformation between images X and Y where the overlap volume becomes small. The proportion of non-background voxels in the overlap volume (shaded) is denoted by χ_1 for X and χ_2 for Y.

3.1.1. Large translation

Consider the case where there is a large translational misalignment between one image and another. In these conditions the intersection of the volume domains becomes small and generally any tissue in one volume corresponds

Table 2
Asymptotic forms of the cost functions for large translational misalignment

Cost function	Approximation	Limit as $\chi_1 \to 0$	Limit as $\chi_2 \to 0$
C^{W}	$(1-\chi_1)\left(\frac{\chi_2}{1-\chi_1}B^2-\left(\frac{\chi_2}{1-\chi_1}B\right)^2\right)^{1/2}\left(\frac{\chi_2}{1-\chi_1}B\right)^{-1/2}$	$\left(\frac{1}{\chi_2}-1\right)^{1/2}$	∞
C^{CR}	$(1-\chi_1)\left(\frac{\chi_2}{1-\chi_1}B^2-\left(\frac{\chi_2}{1-\chi_1}B\right)^2\right)(\chi_2 B^2-\chi_2^2 B^2)^{-1}$	1	1
C^{JE}	$-\chi_1 \log(\chi_1)-\chi_2 \log(\chi_2)-(1-\chi_1-\chi_2)\log(1-\chi_1-\chi_2)$	0	0
C^{MI}	$-\chi_1 \log(\chi_1)-\chi_2 \log(\chi_2)-(1-\chi_1-\chi_2)\log(1-\chi_1-\chi_2)+\chi_1 \log(\chi_1)$ $+(1-\chi_1)\log(1-\chi_1)+\chi_2 \log(\chi_2)+(1-\chi_2)\log(1-\chi_2)$	0	0
C^{NMI}	$\frac{\chi_1 \log(\chi_1)+\chi_2 \log(\chi_2)+(1-\chi_1-\chi_2)\log(1-\chi_1-\chi_2)}{\chi_1 \log(\chi_1)+(1-\chi_1)\log(1-\chi_1)+\chi_2 \log(\chi_2)+(1-\chi_2)\log(1-\chi_2)}$	1	1

Table 3
Asymptotic forms of the cost functions for large scale disparity

Cost function	Approximation	Limit
C^{W}	$\Sigma_i \frac{n_i}{N} \frac{\sqrt{\mathrm{Var}(Y_i)}}{Y}$	0
C^{CR}	$\frac{1}{\mathrm{Var}(Y)} \Sigma_i \frac{n_i}{N} \mathrm{Var}(Y_i)$	1
C^{JE}	$H(X,Y) \geqslant H(X)$	$H(X)$
C^{MI}	$H(X,Y)-H(X)$	0
C^{NMI}	$\frac{H(X,Y)}{H(X)}$	1

only to background in the other volume. More specifically, consider the case where there is a single, small amount of tissue contained in either volume, with the proportion of voxels containing tissue to total number of voxels in the intersection volume being χ_1 for X and χ_2 for Y (see Fig. 1). Furthermore, without loss of generality, take the background intensity as zero and the tissue intensity as B. The cost functions then tend to the asymptotic values shown in Table 2.

From the results in Table 2 is can be seen that most cost functions approach their maximum values, indicating poor registrations, as desired. However, Joint Entropy actually approaches its minimum value indicating, erroneously, that the registration is good.

3.1.2. Large scale disparity

Consider now the case where a small portion of the floating volume is stretched to cover the reference volume.[1] The floating volume (Y) is therefore approximately constant over the volume of overlap. This implies that $H(Y) \approx 0$, leading to the results shown in Table 3. Note that it is also assumed that the spatial variation of Y is uncorrelated with the iso-sets of X, such that $\Sigma_i (n_i/N) \mathrm{Var}(Y_i) \approx \mathrm{Var}(Y)$.

In this instance both the Woods function and Joint Entropy give low cost values indicating a good registration, and so are likely to violate the assumption (stated in Section 2.3) that the global minimum is the desired solution. Consequently, these functions should not be used. Note that although the Joint Entropy does not approach zero, the value is $H(X) = H(X,X)$ which is the same as it would be for a *perfect* registration of the image X with itself.

Overall, therefore, only the Correlation Ratio, Mutual Information and Normalised Mutual Information display the correct asymptotic behaviour and so these are the only suitable cost functions (from this selection) to use.

3.2. Extent of minima

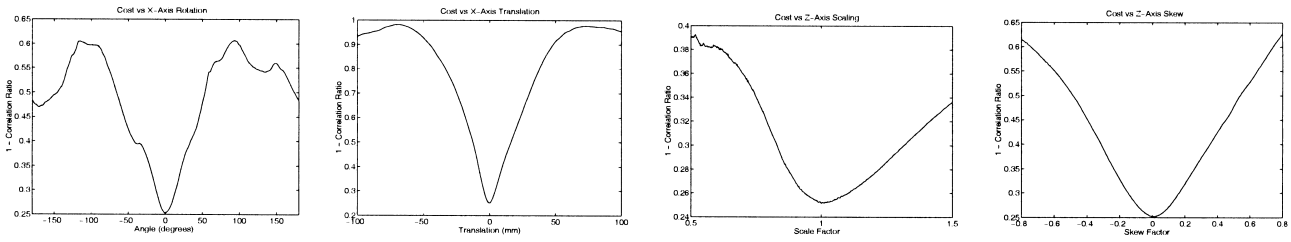
The number and location of local minima for a given cost function can only be determined empirically, as otherwise the registration problem could be solved analytically. However, the typical extent of local minima, the size of their basins of attraction and their inherent smoothness are important for designing effective optimisation strategies.

One approach for examining the cost function is to look at how it changes as individual parameter values vary about some fixed point, that is, to plot the function values versus one parameter value, with all other parameters held constant.

As it is the global minimum that is of most interest, the fixed point is chosen to be the global minimum itself. Fig. 2 shows the cost function plots for a selection of parameters (rotation angle about x, translations in x, scaling in z and skew in z) using the images shown in Fig. 3. This image pair was chosen as it had been found to be a difficult pair to register successfully. However, despite this fact, the global minimum is still quite broad and well defined.

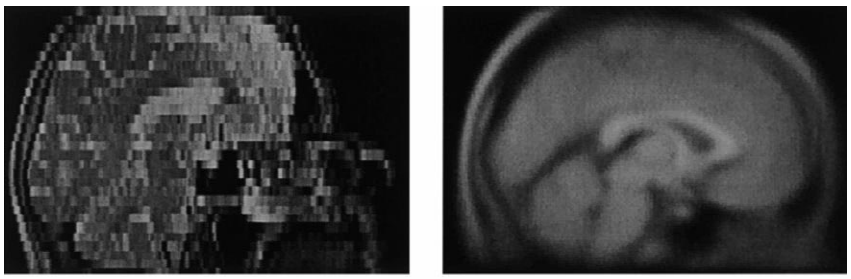
The plots shown in Fig. 2 give an idea of the gross

[1] Similar results are obtained when the floating volume is compressed, although the analysis is slightly different as the overlap volume varies with scale.



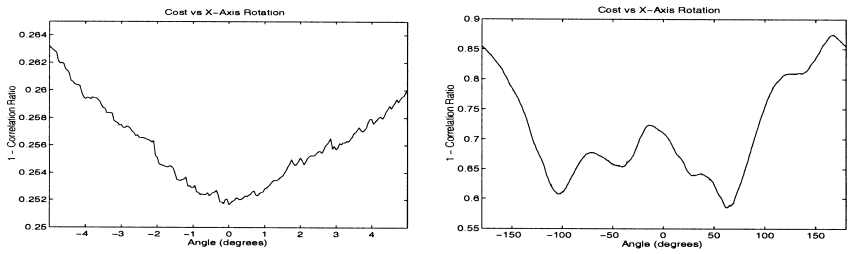
(a) (b) (c) (d)

Fig. 2. Plots of the Correlation Ratio cost function versus some of the individual parameter values. In each plot, a single parameter is varied over a large range while the others are kept fixed. The central point about which the parameters varied was the global minimum. All cost function calculations were done using the image pair shown in Fig. 3 with a resampling (including appropriate pre-blurring) to cubic voxels with 8 mm side-length.



(a) (b)

Fig. 3. Example slices from the volumes used for plotting the cost function. (a) A T2-weighted image of a subject; (b) the T1-weighted MNI 305 image (Collins et al., 1994). The voxel dimensions for these images were $0.93 \times 0.93 \times 5$ mm for (a), and $1 \times 1 \times 1$ mm for (b).



(a) (b)

Fig. 4. Plots of the Correlation Ratio cost function versus rotation about the x-axis, showing the presence of local minima. (a) An expansion of the plot shown in Fig. 2(a), over a small range about the global minimum. It can be seen that several small, local minima exist, generated by small fluctuations in the cost function. In (b) the cost function is plotted about a different central point — one that is offset from the global minimum by a translation of 5 mm in z and a rotation of 40° about the y-axis. This produces significant changes in the cost function, showing the presence of large local minima.

behaviour of the cost function about the global minimum. However, on a small scale there exist many small, local minima, as shown in Fig. 4(a). In addition, there also exist larger local minima as shown in Fig. 4(b), where the central point is no longer the global minimum, but a transformation more typical of an initial alignment. Note that even though there is a clearly defined, large local minimum in this plot, in the full 12-dimensional transformation space this minimum may or may not lead (along a 'valley') to the global minimum. Consequently a successful optimisation method must be able to cope with both small, densely packed local minima and much larger, and widely separated, local minima. These problems are, respectively, dealt with in Section 4 by (1) an adaptive step local minimisation method (Section 4.2) and (2) an initial search stage (Section 4.3).

4. Global optimisation method

This section presents a global optimisation method that is specifically tailored for volumetric registration of brain images. It uses trilinear interpolation with either the Correlation Ratio or Mutual Information as a cost function. However, the optimisation method is quite general and can be used for most cost functions, provided that they have the correct asymptotic behaviour.

4.1. Overview

The fundamental idea of this method is to combine a fast local optimisation Powell's method (Press et al., 1995) with an initial search phase, tuned to be computationally feasible. Our implementation of the whole optimisation method executes in approximately 30 minutes on a modern PC.

There are four different resolution levels that are used by the optimisation algorithm: $n = 8, 4, 2, 1$ mm. Initially, 8-mm cubed voxels are used and a full search is conducted over the rotation angles. Following this, various local optimisations are performed with a variety of starting points in the local neighbourhood of the best points identified in the search. These local optimisations are done using 4-, 2- and finally 1-mm cubed voxels. An exception to this occurs when the initial images are inherently low-resolution in which case the resolution n is not allowed to decrease below a certain limit. For instance, if the original images had voxel sizes of 4-mm cubed then the final stages are performed with $n = 4$ rather than $n = 2$ or $n = 1$. Each

of the optimisation stages is described below in more detail.

4.2. Local optimisation method

A major component in the overall optimisation process is the local optimisation method. This local optimisation is called repeatedly and so should be efficient. Here Powell's method was chosen since it is a commonly used and efficient local optimisation method (Press et al., 1995).

Since the objective at coarse resolutions is only to find an approximate transformation, suitable for refinement at finer resolutions, it follows that the changes in transformation used by the optimisation method should also be adapted to the resolution of the volumes. For example, computing very small changes in transformation when the voxel size is large corresponds to very small sub-voxel changes and is a waste of computation. Consequently, a lower limit for the transformation step size is enforced. This step size is easily calculated by imposing the condition that the maximum voxel position shift should be less than half a voxel. More precisely, at a resolution n (with cubic voxels of side-length n mm), a brain of radius R, and an origin located at the centre of mass for the brain, the maximum shifts are $x_{\max} \leqslant n/2$ which gives $\Delta t \leqslant n/2$ for translation, $\Delta\theta \leqslant n/2R$ for rotation, $\Delta s \leqslant n/2R$ for scale and $\Delta k \leqslant n/2R$ for skew.

Given these lower limits on the parameter steps, Powell's algorithm is modified by changing the termination conditions so that when the 1D optimisation (Brent's method) has bounded the minimum within an interval that is less than one parameter step, then it returns the midpoint of the current interval. In this way, significant savings in computation can be had without any sacrifice of accuracy. Moreover, the final accuracy is largely determined by the final pass, which can have more conservative bounds if better accuracy is required.

4.3. Initial search

As shown in Section 2.2, the amount of time required to perform all but the most perfunctory search is prohibitive, especially when the voxels are smaller than 8-mm cubed. However, in order to find the global minima reliably, some search must be performed.

To allow a useful search to be conducted in under 20 minutes, a multi-stage search was devised. This search is based on the empirical observation that finding the correct orientation, or rotation, is the most difficult task, since the three rotation parameters are highly coupled and most of the erroneous registrations that we have examined have been primarily due to an incorrect orientation. Therefore, the search concentrates on the rotational part of the transformation space. The outline is:

1. Initially form a coarse grid over the three rotation parameters (Euler angles), with M angles per dimension (total of M^3 grid points).
2. At each point in the grid, perform a four-DOF *local optimisation* to find the translation and (global) scale — using initial values of translation = 0 and scale = 1.
3. Calculate the median scaling from the optimised results.
4. Form a finer grid over the rotation angles, with N angles per dimension.
5. Provide initial parameter estimates for translation at each fine grid point by interpolating the optimised translation values found at the coarse grid points.
6. *Evaluate* the cost function (no optimisation) at all points on the fine grid with scale set to the previously calculated median scale.
7. Find all points that have lower cost than their neighbours (local minima) and perform a seven-DOF optimisation for each of these points, storing the results of the transformations (and costs) both before and after optimisation.

Ideally, by choosing a large value for M, the initial stage (steps 1 and 2) of the search would be sufficient. However, as it typically takes about 2 seconds to perform each 4-DOF optimisation in our implementation, only a small value for M can be chosen if the search is to be carried out in a limited amount of time. Therefore another two stages are introduced into the search to allow a finer grid ($N > M$) to be used. In practice, it has been found that values as low as $M = 6$ and $N = 20$ give good results.

To find the cost at each point on a finer grid (where $N > M$) only a single evaluation, rather than an optimisation, is done. However, in order that the cost value is close to the optimal value it is necessary to ensure that they are not significantly influenced by translation and scale differences. In order to achieve this, good initial estimates for these parameters are required. The initial estimates used are obtained from the parameters stored after optimisation on the coarse grid. Since the translation is strongly coupled to the rotation, the initial translation parameters are obtained by interpolating the values from the coarse grid. However, the scaling parameter is not strongly coupled with rotation, and so a single, robust estimate (the median) is taken from all the coarse grid values.

This method relies on certain assumptions, namely:

1. At least one point in the fine grid will occur in the global minimum's basin of attraction, and be lower than its neighbours.
2. The initial estimates for translation and scale, obtained by a four-DOF optimisation at each coarse grid point, are reasonable.
3. The typical anisotropic scaling and skews are sub-voxel with 8-mm cubed voxels, and can be ignored.

The first assumption can be partly justified by the results shown in Section 3.2. It can be seen that the typical size of the global minimum's basin of attraction is large enough that a suitably fine grid should manage to have at least one

point within the basin of attraction, even though the initial orientation may lie outside.

4.4. Re-optimising with perturbations

Following the search at 8-mm resolution, a progressive refinement is required, starting with the 4-mm resolution.

The 8-mm stage identifies many candidate local minima. Since the relative costs of each candidate solution may change at higher resolutions, where structures become more sharply defined, several candidate solutions are considered in the 4-mm stage, rather than just the single best solution. This effectively makes the algorithm a multi-start algorithm, similar to the approach used in Genetic Algorithms. In practice, the best three solutions from the 8-mm stage are used. Furthermore, perturbations about each of these solutions are applied, as described below.

For each candidate solution, an initial 7-DOF optimisation is run at 4 mm in order to find the position of the local minima with this new resolution, which is likely to have moved slightly from the position found at the 8-mm resolution. Furthermore, 7-DOF optimisations are performed for several perturbations of this newly optimised candidate solution. These perturbations attempt to correct for typical registration errors, with the perturbations being $\pm\frac{1}{2}\Delta\theta_{\text{fine}}$ in each rotation angle and $\pm\Delta s$, $\pm 2\Delta s$ in scale, giving a total of 10 different perturbations, where $\Delta\theta$ and Δs are the local optimisation step sizes evaluated for the 8-mm resolution. It has been found that this simple, relatively inexpensive procedure, corrects for the majority of mis-registrations.

This perturbation stage is equivalent to using extra local searches about the candidate solutions. A similar approach is used in Simulated Annealing to escape from local minima, with the difference here being that a set number of deterministic perturbations (determined from an understanding of the previous stage) are used, rather than some number of stochastic perturbations terminated by some convergence criterion.

Overall, therefore, a total of 33 optimisations with 7 DOF are performed at the 4-mm resolution. This acts as a local search about each of the most promising local minima found by the previous (8-mm) stage. The time taken to complete this stage is typically less than 10 minutes but it is a critical stage in eliminating many common mis-registrations (see results in Section 6.1.2).

4.5. Refining the transformation

After the 4-mm stage has completed, and the best solution at this resolution found (the least cost solution from the previous stage), this single solution is refined further at the higher resolutions. No further multi-start optimisation is performed here for reasons of speed, since the computational burden now becomes much more significant.

For the 2-mm case, the optimisation is initially run with 7 then 9 then 12 DOF in order to iteratively improve the fit. Note that it is only at this resolution that anisotropic scalings and skews are included in the optimisation, as in coarser resolutions they are only expected to have a small, sub-voxel effect.

In the 1-mm case, the computational burden is high, such that a single optimisation pass with 12 DOF takes approximately 10 minutes. Therefore, only a single optimisation pass is performed, in order to make minor adjustments in the transformation. Even so, this stage typically takes 10 minutes to run which is approximately 25% of the entire execution time.

4.6. Discussion

No practical guarantee of finding the global minimum is available using this method. However, this is true for most global optimisation methods which at most provide only statistical guarantees for convergence which can never be met in practice (as they require infinite time in theory). The timing for this method, on the other hand, is more or less guaranteed to be less than an hour for our implementation.

Much is based on empirical observations; however, the performance will certainly not be worse than using the standard (no search) multi-resolution approach as this is a special case of the above method, equivalent to having $M = N = 1$ and using no perturbations. A justification of this method on theoretical grounds, although desirable, remains a topic for future research. At present the justification is instead found in the results presented in Section 6, which shows that having a fast global optimisation method is an important practical issue for *fully automatic* registration.

5. Implementation

In almost all methods there are several additional choices which need to be made at the implementation stage. These choices are important and a re-implementation of a method will be unlikely to give similar results (and may even fail completely) unless these details are treated the same way.

5.1. Parameterisation

The first implementation choice to make is the way in which the transformation is represented. The best parameterisation is one that decouples the parameters in a sensible way. One standard decomposition of an affine matrix, which decouples the parameters, is to split the transformation into three rotation (Euler angles), three translation, three scale and three skew parameters.

It is also necessary to choose an origin for the above decomposition. Our implementation uses the *Centre of*

Mass (COM) as the origin since it helps to minimise the coupling between the translation and other parameters. In addition, the COM of each volume is initially aligned at the beginning of the search phase.

5.2. Histogram bin size

Another important choice to make for most cost functions is the histogram bin size, or equivalently, number of bins. All the cost functions examined in Section 3.1 require some form of intensity binning, and the number of intensity bins determines the accuracy of the calculated statistics.

For a uni-modal distribution it has been shown (see (Izenman, 1991) for details) that the optimal histogram bin size, which provides the most efficient, unbiased estimation of the probability density function, is achieved when $W = 3.49\sigma N^{-1/3}$ or $W = 2(IQR)N^{-1/3}$ where W is the width of the histogram bin, σ is the standard deviation of the distribution, N is the number of available samples, and IQR is the interquartile range (the 75th percentile minus the 25th percentile).

In both formulations the width is proportional to $N^{-1/3}$. Assuming that this proportionality will also hold for multi-modal distributions, and using the fact that the number of voxels at resolution n mm is proportional to n^{-3}, then the optimal bin width will be proportional to n. In practice, the number of intensity bins used at resolution n is $256/n$. This sets the bin width proportional to n and fixes the number of bins for the 1-mm resolution to 256, in common with many other registration implementations and which has been found, empirically, to be a good choice for most images.

5.3. Sub-sampling and resolution

The precise method of sampling and sub-sampling the volumes to calculate the cost function is also important. In the method proposed here the reference volume is initially re-sampled (if of sufficient quality) to an isotropic grid with voxel size 1-mm cubed. This is done by interpolating the values available in the original volume which usually has anisotropic voxel dimensions. Once this isotropic 1-mm resolution reference volume has been obtained, the 2-, 4- and 8-mm sub-sampled versions are created.

Sub-sampling the reference volume is done by first blurring the intensities using a convolution with a discrete, 3D Gaussian kernel where $FWHM = n$ mm (or $\sigma = 0.425$ n mm), with n being the size of the required sub-sampling (that is, 2, 4 or 8). This blurring is done so that all points on the lattice contribute equally to the sub-sampled version. The sub-sampling then simply takes every nth point on the lattice in each direction. Therefore, the new volume contains $1/n^3$ as many points as the original.

The floating image is also blurred at each resolution using the same size of Gaussian kernel ($FWHM = n$ mm). However, it is not resampled afterwards since it is computationally more efficient to only store the initial grid values and just calculate the required interpolated intensities as required in each cost function evaluation.

6. Results

The registration method described above in Sections 4 and 5 has been implemented in C++ and is called FLIRT–FMRIB's Linear Image Registration Tool. This program has undergone extensive trials over several months, being used by various researchers including trained neurologists, psychologists and physiologists. During this time it has been used to perform many thousands of registrations in the context of FMRI analysis and structural studies (Smith et al., 2001).

Feedback from the users has been positive, with the vast majority of registrations producing acceptable results and only a few cases of failure where the misalignment of some anatomical structures was substantial (more than several millimetres). It was also found that, of the failures reported, the other registration methods available (see below) also failed to find an acceptable registration. Furthermore, these volumes were usually difficult to manually register as they often had unusual features such as particularly enlarged ventricles which, if aligned, caused large misalignments in other structures, while if the other structures were aligned the ventricles were then substantially misaligned. Consequently, these cases are ones where the main problem appears to be the transformation space. That is, a good *affine* registration does not exist. It would require the use of higher order transformations (non-linear warpings) to achieve a good solution.

6.1. Consistency test

The feedback discussed above is purely qualitative, based on the subjective assessment of many individuals. However, as is generally the case for many registration problems in practice, there was no ground truth available to test the registration against. This makes the area of quantitative assessment of methods quite difficult.

In order to test the method more quantitatively, a comparative consistency test was performed. That is, the registrations obtained using various different initial starting positions of an image are compared to see if they are consistent with one another. Consistency is a *necessary* but not *sufficient* condition that all correctly functioning registration methods must possess. This is essentially a measure of the robustness rather than the accuracy (West et al., 1997) of the registration method. Robustness is defined here as the ability to get close to the global minimum on all trials, whereas accuracy is the ability to precisely locate a (possibly local) minimum of the cost

function. Ideally a registration method should be both robust and accurate.

More specifically, the consistency test for an individual image I involved taking the image and applying several pre-determined affine transformations, A_j, to it. All these images (both transformed and un-transformed) were registered to a given reference image, I^r, giving transformations T_j. If the method was consistent the composite transformations $T_j \circ A_j$ should all be the same, as illustrated in Fig. 5.

The transformations are quantitatively compared using the RMS deviation between the composite registration $T_j \cdot A_j$ and the registration from the un-transformed case T_0. This RMS deviation is calculated directly from the affine matrices (Jenkinson, 2000). That is, $d_{\mathrm{RMS}} = \sqrt{\frac{1}{5}R^2\mathrm{Tr}(M^\top M) + t^\top t}$ where d_{RMS} is the RMS deviation in mm, R is a radius specifying the volume of interest, and $\begin{pmatrix} M & t \\ 0 & 0 \end{pmatrix} = T_j \cdot A_j \cdot T_0^{-1} - I$ is used to calculate the 3×3 matrix M and the 3×1 vector t.

6.1.1. Comparison with existing methods

A comparison of FLIRT with several other registration packages was initially performed using the consistency test explained above. The other registration packages used were AIR (Woods et al., 1993), SPM (Friston et al., 1995), UMDS (Studholme et al., 1996) and MRITOTAL (Collins et al., 1994). These methods were chosen because the authors' implementations were available, and so this constituted a fair test as opposed to a re-implementation of a method described in a paper, where often the lack of precise implementation details makes it difficult to produce a good working method.

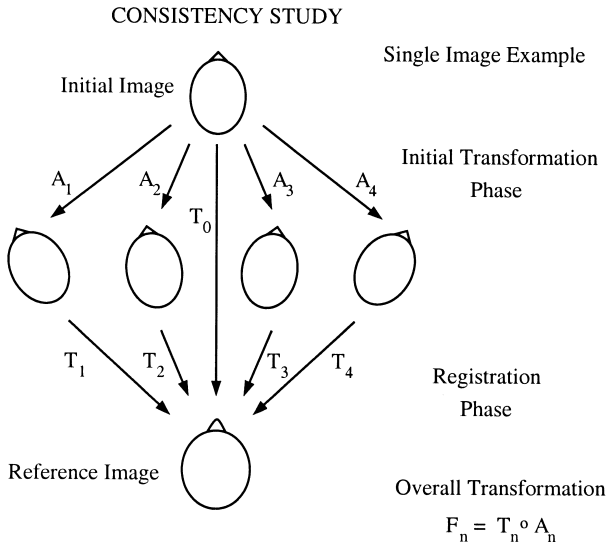


Fig. 5. Illustration of the consistency test for a single image. An image (top) has a number of initial affine transformations A_j (rotations are used in this study) applied to it. The resulting images (middle) are then registered to the reference image (bottom), giving transformations T_j. Therefore, the overall transformation from the initial image to the reference image is $F_j = T_j \cdot A_j$, and these are compared with T_0 which is the registration of the initial image directly to the reference image. For a consistent method, all the transformations, F_j, should be the same as T_0.

The particular experiment that was performed (which is also described in (Jenkinson and Smith, 1999)) was *inter-subject* and *inter-modal* using 18 different images as the floating images (like the one shown in Fig. 3(a)), all with the MNI 305 brain (Collins et al., 1994) as the reference image. The 18 images were all $256 \times 256 \times 30$, T2-weighted MR images with voxel dimensions of $0.93 \times 0.93 \times 5$ mm, while the MNI 305 template is a $172 \times 220 \times 156$, T1-weighted MR image with voxel dimensions of $1 \times 1 \times 1$ mm. Note that this is an *inter-subject*, *inter-modal* registration problem.

The results of such a test, using six different rotations about the anterior–posterior axis, are shown in Fig. 6. It can be seen that only FLIRT and MRITOTAL were consistent with this set of images. This indicates that the other methods, AIR, SPM and UMDS, get trapped in local minima more easily, and are not as robust.

A further consistency test was then performed comparing only MRITOTAL and FLIRT. This test used initial scalings rather than rotations. The reason that this is important is that MRITOTAL uses a multi-resolution local optimisation method (Gradient Descent) but relies on initial pre-processing to provide a good starting position. This pre-processing is done by finding the principle axes of both volumes and initially aligning them. However, this initial alignment does not give any information about scaling.

The results of the scaling consistency test are shown in Fig. 7. It can be seen that, although generally consistent, in three cases MRITOTAL produces registrations that deviate by more than 20 mm (RMS) from each other. In contrast, FLIRT was consistent (less than 2 mm RMS) for all images.

6.1.2. Comparison of optimisation methods and cost functions

To further investigate the effects of the proposed optimisation method two more consistency tests were performed. These tests were designed to separate out the contribution that the cost function and the various components of the optimisation method make to achieving the consistency demonstrated in the previous tests.

Firstly, the rotation consistency test was applied using the Correlation Ratio cost function but with various components of the optimisation method removed. These results are shown in Fig. 8 for two cases: (a) no initial search (that is, $N = M = 1$) and no perturbations, and (b) an initial search, but no perturbations. In neither case are the results consistent (as they are in Fig. 6(a)), demonstrating that both components of the optimisation method are necessary to achieve the consistent results.

Secondly, the same test as above was applied but using Mutual Information as a cost function. The results shown

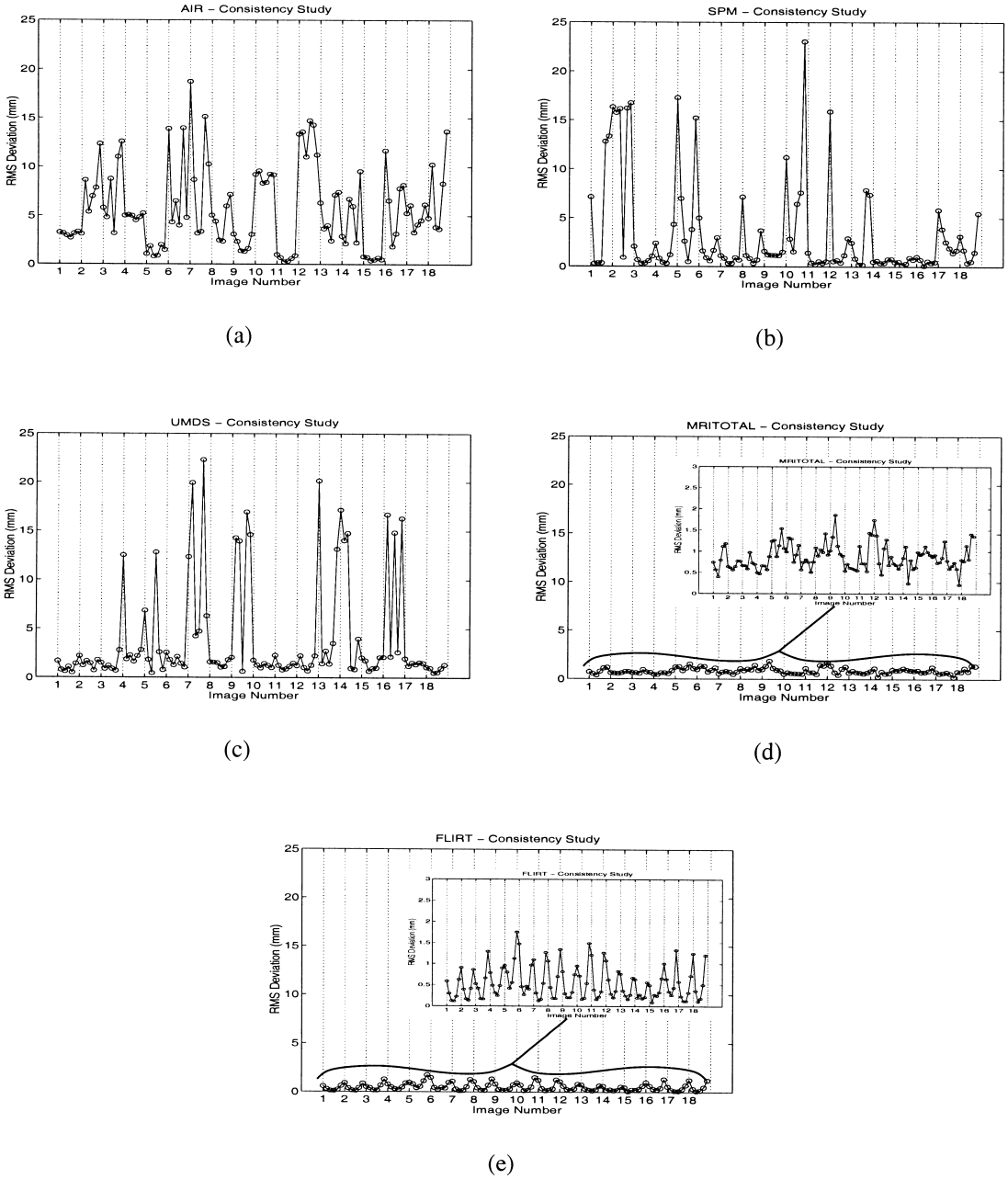


Fig. 6. Results of the consistency study, plotting RMS deviation (in mm) versus image number for (a) AIR, (b) SPM, (c) UMDS, (d) MRITOTAL and (e) FLIRT. For each of the 18 source images (T2-weighted MRI images with voxel dimensions of $0.93 \times 0.93 \times 5$ mm) there are 6 results corresponding to initial starting rotations of -10, -2, -0.5, 0.5, 2 and 10 degrees about the y-axis (anterior–posterior axis). All of the methods, except FLIRT and MRITOTAL, show large deviations and are therefore inconsistent and non-robust.

in Fig. 9 are for (a) no initial search and no perturbations and (b) full optimisation, including initial search and perturbations.

These last two tests indicate clearly that both parts of the optimisation method (the initial search and the subsequent use of perturbations) are critical for obtaining the observed consistency. That is, using a multi-resolution local optimisation method alone (Figs. 8(a) and 9(a)) is insufficient to achieve the desired robustness. Furthermore, the full optimisation method performs robustly using either the Correlation Ratio or Mutual Information as cost functions. Therefore, the improvement in performance, by comparison with the methods examined in Section 6.1.1, is due to the optimisation method and not due to the different cost function used, since Mutual Information is used as the cost function in UMDS, SPM and MRITOTAL. Note that for FLIRT, the Correlation Ratio is adopted as the default cost function since, for our implementation, it is quicker to compute and the results are, as judged qualitatively in many different cases, slightly more accurate.

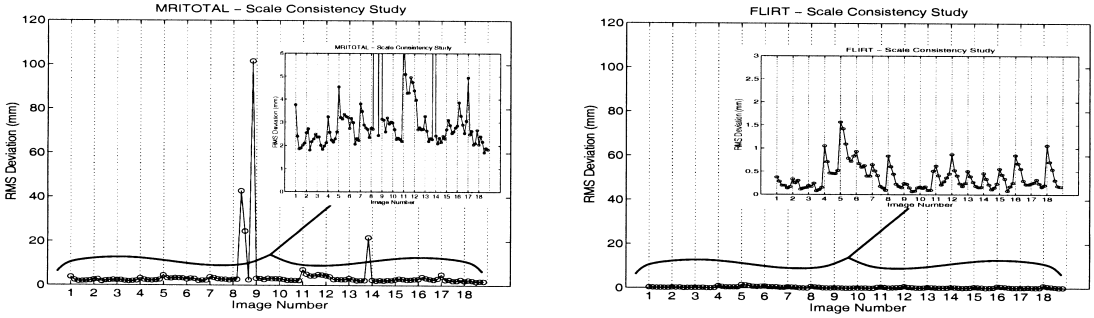


Fig. 7. Results of the scale consistency study, plotting RMS deviation (in mm) versus image number for (a) MRITOTAL and (b) FLIRT. For each of the 18 source images (T2-weighted MRI images with voxel dimensions of $0.93 \times 0.93 \times 5$ mm) there are 6 results corresponding to initial scalings of 0.7, 0.8, 0.9, 1.1, 1.2 and 1.3 about the Centre of Mass. In three cases MRITOTAL shows large deviations and so is less consistent and robust than FLIRT in this case.

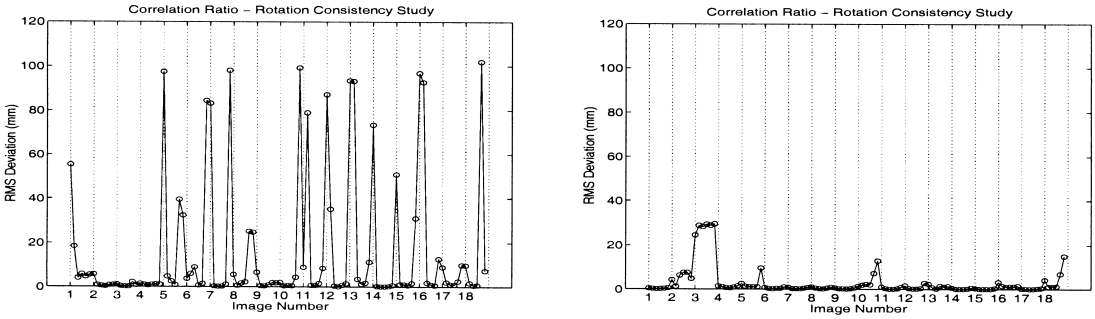


Fig. 8. Results of the Rotation Consistency Test for FLIRT using the Correlation Ratio cost function and different optimisation methods. The results shown in (a) are with no initial search (that is, $N = M = 1$) and no perturbations, while in (b) the initial search is performed, but no perturbations. Comparing these to the results obtained using the full optimisation method (shown in Fig. 6(e)) it can be seen that both components of the optimisation method are needed to obtain a robust method.

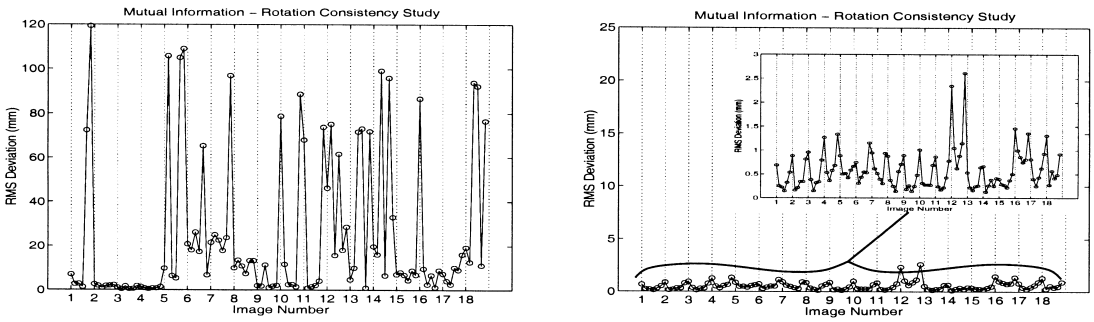


Fig. 9. Results of the Rotation Consistency Test for FLIRT using the Mutual Information cost function and different optimisation methods. The results shown in (a) are with no initial search and no perturbations, while in (b) the full optimisation method (initial search and perturbations) is used. This demonstrates that the full optimisation method is required for robustness. Furthermore, comparing (b) with Fig. 6(e) shows that the results using the full optimisation method are similar for either Mutual Information or the Correlation Ratio cost functions.

7. Discussion

In this paper the problem of global optimisation for fully automatic registration of brain images was examined. Only affine registration was examined as, although this is a much easier problem than general, non-linear transformations, finding the global minimum is still difficult. Furthermore, many non-linear methods rely on an initial affine registration to find a good starting position, and so having a good method of affine registration is important.

The standard mathematical formulation of the registration problem was detailed and its assumptions examined. This included an analysis of the asymptotic behaviour of several multi-modal cost functions and showed that only the Correlation Ratio, Mutual Information and Normalised Mutual Information correctly yielded high costs for these extreme misalignments. Consequently, only these functions (out of those considered) are suitable for the global optimisation formulation.

A fast global optimisation method, which was tailored to this registration problem, was then proposed together with a complete discussion of the implementation details required to turn the method into a practical registration tool. This method uses the Correlation Ratio or Mutual Information (although any suitable cost function could be used) and combines a common local optimisation method (Powell's method with step size proportional to the voxel size to improve the efficiency) with several search strategies. The main search is a global search through the transformation space using the most sub-sampled volumes, with smaller neighbourhood-based searches taking place as the sub-sampling decreases. In such methods, the major constraint is the amount of time that is considered reasonable. In the design of this method the search time was limited to 20 minutes using our implementation. Even with this constraint it was still possible to perform a full search and identify the global minimum more reliably.

Evaluation of this registration method was done in two ways. Firstly, the software has been used routinely in the FMRIB Centre, allowing many people to use and comment on the performance. The qualitative feedback was positive in its own right and in comparison with other available methods. This method has now been used to satisfactorily solve thousands of registration problems.

Secondly, quantitative results were found for a consistency test. This test is designed to examine the robustness of a registration method. Results showed that the method was highly consistent on a set of difficult images. Furthermore, several other available packages were tested on the same set of images and did not achieve the same level of consistency, sometimes demonstrating substantial inconsistencies. This was further supported by tests using the method with a different cost function (Mutual Information) and, additionally, removing stages from the optimisation method. These tests showed that the robustness was not due to the choice of cost function and that each component of the optimisation method was necessary to achieve the robust registrations, hence showing that multi-resolution local optimisation alone is insufficient.

The global optimisation method proposed here does not, however, guarantee finding the global minimum. This is typical though, as even methods such as Simulated Annealing and Genetic Algorithms only provide a statistical guarantee which cannot be met in practice. The results, though, are encouraging, and by using finer search grids, the likelihood of finding the global minimum can be increased. This requires that there be sufficient time at hand, or a sufficiently fast computer. However, even with modest resources this method can find the global minimum (solving the registration problem) within 1 hour more reliably than the other methods tested.

Optimisation is only one aspect of the registration problem, although it is practically a very important one. Other aspects such as interpolation, alternative cost functions and understanding the properties of existing cost functions remain important areas for further work. In addition, a theoretical justification for the current method and finding a method suitable for higher dimensional transformations are important areas for future research.

Finally, as stated before, it is important to be precise about the implementational details of such methods. This allows (1) the methods to be more easily re-implemented by others, (2) the various methods to be compared fairly (by using the author's parameters) and (3) the results to be repeated. The alternative is finding the best value for various implementation parameters by trial and error which is extremely tedious and error prone. In addition to including the implementation details in this paper the source code for the FLIRT package is available for downloading from www.fmrib.ox.ac.uk/fsl. This should avoid others needing to re-implement the method and facilitate the evaluation of the method, hopefully leading to further improvements.

Acknowledgements

The authors wish to thank the Medical Research Council and the European MICRODAB project for supporting this work.

References

Collignon, A., Vandermeulen, D., Suetens, P., Marchal, G., 1995. 3D multi-modality medical image registration using feature space clustering. In: Ayache, N. (Ed.), Lecture Notes in Computer Science, LNCS, Vol. 905. Springer. Proceedings of the 1st International Conference on Computer Vision, Virtual Reality and Robotics in Medicine, Nice, France, pp. 195–204.

Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. Journal of Computer Assisted Tomography 18 (2), 192–205.

Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.-B., Heather, J.D., Frackowiak, R.S.J., 1995. Spatial registration and normalization of images. Human Brain Mapping 2, 165–189.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans. on Pattern Analysis and Machine Intelligence 6 (6), 721–741.

Hajnal, J.V., Saeed, N., Soar, E.J., Oatridge, A., Young, I.R., Bydder, G.M., 1995. A registration and interpolation procedure for subvoxel matching of serially acquired MR images. Journal of Computer Assisted Tomography 19 (2), 289–296.

Ingber, A.L., 1989. Very fast simulated re-annealing. Jounral of Mathematical and Computational Modelling 12, 967–973.

Izenman, A.J., 1991. Recent developments in nonparametric density estimation. Journal of the American Statistical Association 86 (413), 205–224.

Jenkinson, M., 2000. Measuring transformation error by RMS deviation. Internal Technical Report TR99 MJ1, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, Oxford University, Oxford, UK.

Jenkinson, M., Smith, S.M., 1999. An investigation of the robustness of registration methods. In: Int. Workshop on Biomedical Image Registration, pp. 200–210.

Maes, F., 1998. Segmentation and registration of multimodal medical images: from theory, implementation and validation to a useful tool in clinical practice. PhD thesis, Catholic University of Leuven, Belgium.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximisation of mutual information. IEEE Trans. on Medical Imaging 16 (2), 187–198.

Pluim, J.P.W., Maintz, J.A., Viergever, M.A., 2000. Interpolation artefacts in mutual information based image registration. Computer Vision and Image Understanding 77 (2), 211–232.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1995. In: Numerical Recipies in C, 2nd Edition. Cambridge University Press, Cambridge.

Roche, A., Malandain, G., Pennec, X., Ayache, N., 1998. The correlation ratio as a new similarity measure for multimodal image registration. In: Wells, W., Colchester, A., Delp, S. (Eds.), MICCAI'98, Lecture Notes in Computer Science, LNCS, Vol. 1996. pp. 1115–1124.

Smith, S.M., De Stefano, N., Jenkinson, M., Matthews, P.M., 2001. Normalised accurate measurement of longitudinal brain change. Journal of Computer Assisted Tomography. in press.

Studholme, C., Hill, D.L.G., Hawkes, D.J., 1995. Multiresolution voxel similarity measures for MR-PET registration. In: Proceedings of Information Processing in Medical Imaging, Brest, France, pp. 287–298.

Studholme, C., Hill, D.L.G., Hawkes, D.J., 1996. Automated 3D registration of MR and CT images of the head. Medical Image Analysis 1 (2), 163–175.

Studholme, C., Hill, D.L.G., Hawkes, D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition 32, 71–86.

Thacker, N.A., Jackson, A., Moriarty, D., 1999. Improved quality of re-sliced MR images using re-normalized sinc interpolation. Journal of Magnetic Resonance Imaging 10, 582–588.

Viola, P., Wells, W.M., 1997. Alignment by maximization of mutual information. International Journal of Computer Vision 24 (2), 137–154.

West, J. et al., 1997. Comparison and evaluation of retrospective intermodality brain image registration techniques. Journal of Computer Assisted Tomography 21 (4), 554–566.

Wolpert, D.H., Macready, W.G., 1996. No free lunch theorems for search. Technical Report SFI-TR-95-02-010, The Santa Fe Insititute.

Woods, R.P., Mazziotta, J.C., Cherry, S.R., 1993. MRI-PET registration with automated algorithm. Journal of Computer Assisted Tomography 17 (4), 536–546.