Spatial registration and normalization of images

Friston KJ, Ashburner J, Frith CD, Poline J-B, Heather JD and Frackowiak RSJ

The Wellcome Dept. of Cognitive Neurology, The Institute of Neurology, Queen Square, WC1N, UK and the MRC Cyclotron Unit, DuCane Rd., London W12 OHS, UK

Address for correspondence

Karl J Friston

The MRC Cyclotron Unit,

Hammersmith Hospital

DuCane Road, London W12 OHS

Tel (0181) 740 3162, Fax (0181) 743 3987, email karl@cu.rpms.ac.uk

Key words: realignment, registration, anatomy, imaging, stereotaxy, morphometrics, basis functions, spatial normalization. PET, MRI, functional mapping.

Abstract

This paper concerns the spatial and intensity transformations that map one image onto another. We present a general technique that facilitates nonlinear spatial (stereotactic) normalization and image realignment. This technique minimizes the sum of squares between two images following nonlinear spatial deformations and transformations of the voxel (intensity) values. The spatial and intensity transformations are obtained simultaneously, and explicitly, using a least squares solution and a series of linearising devices. The approach is completely noninteractive (automatic), nonlinear and noniterative. It can be applied in any number of dimensions.

Various applications are considered, including the realignment of functional MRI time-series, linear (affine) and nonlinear spatial normalization of PET and structural MRI images, the coregistration of PET to structural MRI and, implicitly, the conjoining of PET and MRI to obtain high resolution functional images.

Introduction

This paper is about the spatial transformation of image processes. Spatial transformations are both ubiquitous and important in many aspects of image analysis: For example in neuroimaging the realignment of a time-series of scans from the same subject (correcting for movement) is necessary for voxel based analyses of time-dependent changes. This issue is of current interest in the analysis of functional MRI (fMRI) time-series. Inter-subject averaging, with change distribution analysis or statistical parametric mapping, requires the images to be transformed into some standard stereotactic space (e.g. Fox et al 1988; Friston et al 1991a). Anatomical variability and structural changes due to pathology can be framed in terms of the transformations required to map the abnormal onto the normal. The interpretation of functional mapping studies often refers to some notion of normal anatomical variability (e.g. Steinmetz and Seitz 1991). This variability embodies neuroanatomical information of a probabilistic nature that is the focus of some important new brain mapping initiatives (e.g Mazziotta et al submitted)

Spatial transformations can be broadly classified into *label based* and *non-label based*. Label based techniques identify homologous spatial structures, features or landmarks in two images and finds the transformation that best superposes the labelled points. These transformations can be linear (e.g. Pelizzari et al 1988) or nonlinear [e.g. thin plate splines (Bookstein 1988)]. Non-label based approaches identify a spatial transformation that minimizes some index of the difference between an object and a reference image, where both are treated as unlabelled continuous processes. Again these can be linear e.g. principal axes (Alpert et al 1990); image realignment (Woods et al 1992; Collins et al 1994, and see Lange 1994) or nonlinear e.g. plastic transformation (Friston et al 1991b) with some interesting developments using neural nets (Kosugi et al 1993).

In the absence of any constraints it is of course possible to transform any image such that it matches another exactly. The issue is therefore less about the nature of the transformation but more one of defining the constraints under which the transformation is effected. The validity of the transformation can usually be reduced to the validity of these constraints. The first tenet, of the general approach described here, is that the constraints are explicit, reasonable and operationally specified. The reliability of label-based approaches is limited by the reproducibility of the labelling. Non-label based techniques are generally noninteractive and are therefore completely reliable. The second key aspect of our approach is therefore that it is non label-based and automatic. Many approaches to spatial transformation use some from of gradient descent or nonlinear minimisation to find the global minimum of an error surface. These iterative techniques are flexible but may find local solutions and are computationally very expensive [e.g. 30 minutes for a simple linear MRI to MRI matching (Collins et al 1994)]. To avoid these problems we have used an approach with explicit and unique solutions that can be obtained by solving just one equation.

In short the aim of this work was to develop an efficient, automatic and general multidimensional nonlinear spatial transformation technique (rigid body and affine transformations were considered as special linear cases of this more general approach). The only automatic nonlinear, and noniterative technique used widely at present (Friston et al 1991b) is fundamentally limited in the sense it is only valid for one dimensional images (current implementations apply this one-dimensional transformation to the radii of transverse slices in a polar space).

The paper begins by describing the idea upon which the transformations are based and provides the general operational equations (a full exposition is provided in an appendix). This is followed by a brief description of the fMRI, structural MRI and PET data used in subsequent sections. The remaining four sections deal with a range of specific applications and extensions. The first section (*within modality, within subject*) illustrates the simplest application, namely orthogonal (rigid body) transformations of single slice fMRI data that would be applied to reduce head movement artifact in fMRI time-series. The algorithm was tested using simulated head movement and using a

real time-series of 64 images. Performance was assessed by comparing the estimated and known movement parameters. The estimates were also compared with those obtained with a widely used image realignment routine (Woods et al 1992). Using a real time-series the validity of the realignment was further addressed by looking at its effect on time-dependent functional changes. The second section (*within modality, between subject*) deals with general nonlinear three dimensional transformations of the sort required to map an arbitrary PET image into a standard PET image conforming to the space described by the atlas of Talairach and Tournoux (1988). This spatial normalization transformation is used to (i) introduce a recursive application of the least squares analysis and (ii) compare it with equivalent nonlinear minimisation. The section concludes by introducing a simultaneous solution for the transformation using a series of feature-enhanced images derived from the primary image pair. This extension is demonstrated in the context of matching structural MRI images from different subjects. The third section (*between modality, within subject*) looks at the registration of structural MRI and PET from the same individual. The focus here is on the intensity transformation implicit in the overall solution. This intensity transformation can be used to create a structural MRI image whose voxel values are functional (e.g. rCBF-like) in nature. The final section deals with the most general problem of *between modality, between subject* normalization using structural MRI and functional PET images from different subjects.

This paper is methodological, however its main aim is to communicate the basic idea, mathematical formalism and the variants or extensions that ensue. The focus is therefore more conceptual than operational. Subsequent papers will present more detailed numerical information on performance, robustness and its application to some simple questions about normal and abnormal anatomical variability.

Theory

The basic idea is to formulate the most general problem of spatial transformation so that a unique least squares solution exists. This involves linearising a highly nonlinear multidimensional problem using reasonable constraints. This linearisation starts by acknowledging that the differences between two images [say $\Omega(x)$ and $\dagger(x)$] have two components: The first component is due to voxel value or intensity differences when two images are in perfect anatomical register. These differences may be artifactual (e.g. different resolutions, low spatial frequency intensity variations in MRI images or different methods of rCBF parameter estimation used for PET) or real (e.g. hypofrontality in PET scans of psychomotor poverty syndromes, differences in global activity, or experimentally induced physiological activations). We assume that, at point x, there is some operator $f_x\{.\}$ that transforms voxel values from one image to those of another (assuming perfect anatomical congruence).

The second component of the differences between the two images will be due to misalignment or indeed differences in shape and size of the objects scanned. This spatial discrepancy between the two scans is assumed to be characterized by a function of position q(x). This partitioning of the difference between one image and another can be expressed as:

$$f_x\{\dagger(x)\} \quad = \quad \Omega(q(x)) + e(x) \qquad\qquad 1$$

Where e(x) is some error. For simplicity this error term will be omitted in subsequent expressions. Eqn(1) expresses the conjecture that two images can be approximated by applying an *intensity transformation* $f_x\{.\}$ to one and a *spatial transformation* q(x) to the other. The next step is to linearise eqn(1) so that both $f_x\{.\}$ and q(x) have explicit least squares solutions. This can be effected by low order approximations and by imposing some constraints on the forms of $f_x\{.\}$ and q(x). The constraints are basically (i) that the operator $f_x\{.\}$ can be expressed in terms of a convolution and a (nonstationary) nonlinear function and (ii) that both $f_x\{.\}$ and q(x) change

slowly with location. For $f_x\{.\}$ this local-stationariness means that the 'operation' that converts voxel values from one image into the other (the intensity transformation) can change with position but is similar within a given locale. In other words the transformation $f_x\{.\}$ is locally stationary but globally unspecified. Stationariness means that something does not change with position in the image. The slowly changing nature of q(x) means that the spatial transformation is smooth and that local contiguity relationships are preserved. More formally the smooth spatial transformation ensures a positive Jacobian which preserves important features, including critical points such as peaks and saddle points (see Amit et al 1991). Stronger constraints can be imposed on q(x), for example in realigning scans from the same individual q(x) would represent an orthogonal rigid-body transformation with six parameters, but would still be smooth.

In what follows $f_x\{.\}$ is taken to be some nonlinear function $\gamma_x(.)$ of the original process convolved with a kernel c(x). This means that, ignoring misregistration effects, one image can be converted into another by applying a nonlinear transformation to the voxel values and then [de-]convolving into the same resolution as the second image. i.e. $f_x\{.\} = c(x) * \gamma_x(.)$ where * means convolution. Note that this form of $f_x\{.\}$ requires the nonlinear function to be applied before the convolution. In practice one would always want to arrange the images so that the one was convolving the higher resolution image to match the lower resolution image. The kernel or differential point spread function c(x) will be known or can be estimated *post hoc* using established and empirical methods of estimating image smoothness (see Friston et al 1991a). Without loss of generality one can take some expansion of $\gamma_x\{.\} = \sum u_i(x).f^i(\dagger(x))$ and expand the position-dependent coefficients $u_i(x)$ in terms of some 'smooth' spatial basis functions $\beta^{f_j}(x)$ [i.e. $u_i(x) = \sum u_{ij}\beta^{f_j}(x)$]. An expansion is simply expressing a function as the sum of (usually) simpler functions. q(x) can be similarly expanded in terms of *x* and some smooth basis functions $\beta^{q_k}(x)$. Smooth basis functions can be thought of as smooth functions of, or profiles in, space that are (usually) chosen to be independent or orthogonal. Using these expansions eqn(1) becomes:

$$c(x) * \sum \{ f^i(\dagger(x)) \sum u_{ij}\beta^{f_j}(x) \} \quad \approx \quad \Omega(x + \sum q_k\beta^{q_k}(x)) \qquad\qquad 2$$

Now if $\Omega(x)$ is smooth the effects of small distortions $q_k \text{ß}^{q_k}(x)$ will not interact to a significant degree. In other words if $\Omega(x)$ is smooth we can expand the right hand side of eqn(2) using Taylor's theorem where, ignoring high order terms:

$$\Omega(x + \textstyle\sum q_k \text{ß}^{q_k}(x)) \quad \approx \quad \Omega(x) \; + \; \textstyle\sum q_k \partial\Omega(x)/\partial q_k \qquad\qquad 3$$

(see appendix). Combining eqn(2) and eqn(3) we get:

$$\textstyle\sum\sum u_{ij} [c(x) * f^i(\dagger(x)) . \text{ß}^{f}_j(x)] \qquad \approx \qquad \Omega(x) \; + \; \textstyle\sum q_k \partial\Omega(x)/\partial q_k \qquad\qquad 4$$

This approximate equality says that, give or take some residual error, one image can be approximated to another by (i) applying an intensity transformation to the first image (where the coefficients of the transformation's expansion can change slowly with position) convolving and (ii) approximating the distortion of the second image by simply adding the effects of each component of the distortion (assuming the components are small relative to the image's resolution). The components of the distortion are defined by a set of smooth basis functions in space.

The importance of eqn(4) is that it is linear in the unknown coefficients (u_{ij} and q_k) and that these coefficients have a unique least squares solution. This may be seen more clearly by considering the matrix equivalent of eqn(4) which is of the form $\mathbf{A}.[\ \mathbf{u}_0\ \mathbf{u}_1....\ \ \mathbf{q}]^\mathrm{T}\ \ \approx \Omega$:

$$[\mathbf{c}.\mathrm{diag}(f^0(\dagger)).\text{ß}^{f}\ \ \mathbf{c}.\mathrm{diag}(f^1(\dagger)).\text{ß}^{f}...\ \ \text{-}\mathrm{diag}(\partial\Omega/\partial\mathbf{x}).\text{ß}^{q}].[\ \mathbf{u}_0\ \mathbf{u}_1....\ \ \mathbf{q}]^\mathrm{T}\ \ \approx \Omega\ \ 5$$

(see the appendix for a fuller explanation and notational details). The vector $[\mathbf{u}_0\ \mathbf{u}_1\ \mathbf{u}_2....\ \ \mathbf{q}]$ has a unique least squares solution. This solution for q_k ($\mathbf{q}$) can then be used to implement the spatial transformation that maps $\Omega(x)$ onto $\dagger(x)$ or *vice versa.* given that $q(x) = x + \sum q_k \text{ß}^{q_k}(x)$.

The Data

This section describes briefly how the data were acquired and stored. The data were obtained from [different] male subjects in accord with local and national ethical requirements.

functional MRI

The data were a time-series of 64 gradient-echo EPI coronal slices (5mm thick, with 64 x 64 voxels) through the calcarine sulcus and extrastriate areas. Images were obtained every 3 seconds from a normal male subject using a 4.0T whole body system, fitted with a small (27cm diameter) z-gradient coil (TE 25ms, acquisition time 41ms). Photic stimulation (at 16 Hz) was provided by goggles fitted with 16 light emitting diodes. The stimulation was off for the first 10 scans (30 seconds), on for the second 10, off for the third, and so on. The data were interpolated to 128 x 128 voxels. The image dimensions were 128x128x1. The voxel size was 1.25x1.25x5mm. Data were stored at 8 bit precision.

Structural MRI data

These data were high resolution structural MRI scans obtained at 1.5T using a standard T1 weighted sequence. The image dimensions (following reorientation of the saggital acquisition and resampling) were 256z256x118. The voxel size was 0.976 x1.3x1.952mm. Data were stored at 8 bit precision.

PET data

The PET images were obtained with a CTI PET camera (model 953B CTI Knoxville, TN USA). Reconstructed images had a resolution of about 8mm (Townsend et al 1992, Spinks et al 1992). [15]O was administered intravenously as radiolabelled water infused over two minutes. Total counts per voxel during the buildup phase of radioactivity served as an estimate of rCBF (Fox and Mintun

1989). The image dimensions (following reconstruction and bilinear interpolation) were 128x128x43. The voxel size was 2.09x2.09x2.45 mm.   Data were stored at 8 bit precision.

All the image manipulations and matrix calculations were performed in Matlab (MathWorks Inc. Sherborn MA, USA) on a SPARC Workstation. The PET and MRI images were scalp edited (pre-processing of this sort is not a prerequisite for anything that follows)

APPLICATIONS

Within subject, within modality

In this section we deal with one of the simpler transformations, namely the registration of homologous images acquired sequentially from the same subject. This application is important because it removes variance from time-series that would otherwise be attributed to error (i.e. decreasing sensitivity) or to treatment effects (i.e. movement artifacts). The most important source of this variance is usually due to movement during the scanning session. Although important in PET studies, these effects can seriously confound the analysis of fMRI studies, where the slightest subvoxel movement may profoundly effect the voxel value. This section is divided into two parts. In the first we compare the least squares approach to known (simulated) movement and to that estimated with a widely used image realignment program (automated image realignment - AIR, Woods et al 1992). In the second part the least squares approach is applied to the real fMRI time series in order to demonstrate that the estimated movement is real and to validate the realignment in terms of detecting functional changes. This is achieved by analyzing the fMRI time-series using singular value decomposition (SVD). SVD identifies important spatial modes and their time-dependent activity.

The reason realignment can be considered a simple case of the more general spatial transformation problem is because the form of the spatial transformation q(x) is exactly specified and the intensity transformation is known $f_x\{\dagger(x)\} \approx \dagger(x)$. The second equality follows from the fact the same object is scanned with the same imaging device (we are obviously assuming that time-dependent physiological changes of interest are small relative to the anatomic profile and that these effects can be relegated to the error term). Substituting these simple forms for q(x) and $f_x\{.\}$ into eqn(4) one has (in matrix notation):

$$[-\text{diag}(\partial\Omega/\partial\mathbf{x}).\mathbf{\beta}^q].\mathbf{q}^T \qquad \approx \qquad [\Omega - \dagger]$$

or $\qquad -\partial\Omega/\partial\mathbf{q}.\mathbf{q}^T \qquad \approx \qquad [\Omega - \dagger] \qquad\qquad\qquad 6$

In this instance the basis functions cover the space of all allowable translations and rotations i.e. $\mathbf{\beta}^q = [\beta^o_1 \; \beta^o_2 \; ....\beta^o_6 \;]$ where the columns β^o_k correspond to translations in three dimensions and the three orthogonal rotations. In practice it is easier to compute the six columns of $\partial\Omega/\partial\mathbf{q}$ directly [as implied in the second form of eqn(6)] by simply applying small translations and rotations to $\Omega(x)$ and measuring the changes in voxel values. The six elements of the row vector $\mathbf{q}$ correspond to the estimated translations and rotations that constitute the movement to be corrected.

Simulated movement

The first fMRI slice was translated over 64 logarithmically increasing distances (1 μm to 3mm) and the displacement was estimated for each simulated movement by comparing the moved image and the first or reference image. These estimates were computed with least squares according to eqn(6) and the appropriate AIR algorithm. Moved images were constructed by adding uncorrelated Gaussian noise (at 5% of the image mean), to emulate the effects of thermal noise in the fMRI scans, and translating using bilinear interpolation. Clearly this is not the same as actually moving the real object but is a reasonable approximation for small movements. The data were saved as 8 bit. This reduction to 8 bit precision introduces nonlinear noise due to round off errors

and to some extent simulates digitizing noise during acquisition. The actual and estimated translations were compared graphically.

Figure 1 presents the results of this analysis. These results highlight two points. First both the AIR and the least squares estimates are generally very good. This sort of analysis helps establish the construct validity of both techniques in terms of the other. Secondly at very large and small displacements there is a dissociation in the performance of the two techniques. AIR is better than the least squares approach when the spatial misalignment approaches the resolution of the images (~ 2.5mm). The upper line (in Figure 1 left) represents the actual movement and the AIR estimates. The lower line corresponds to the least squares estimates. This divergence at large translations is due to a failure of the first order approximation of the Taylor series implicit in eqn(3); more simply if the displacements are bigger than the image's smoothness then the linearisation in the least squares approach becomes suspect. In practice this is not a problem because (i) the images can be smoothed or (ii) the least squares analysis can be applied recursively (see below). Conversely the least squares analysis is better for small displacements. Figure 1 (right) shows the same data but on a semilog scale. The AIR algorithm starts to fail noticeably at about 100 μm whereas the least squares analysis remains reasonably robust until about 10 μm.

The differences between the least squares and AIR estimates should not be over interpreted in the sense that there are many differences in the two algorithms that were not controlled for. In general results obtained from the two analyzes are consistent. It should be noted however that the least squares analysis is about an order of magnitude faster that the AIR. This is because the AIR approach is iterative.

A real time-series

It may seem to some that a precision of 100μm or less is rather irrelevant for realignment routines because this scale of movement is practically undetectable and harmless anyway. They would be wrong. The purpose of this section is to show that (i) movements measured in tens of μm can be

detected in fMRI time series and that (ii) removal of these effects by realignment can substantially effect the variance-covariance structure of the data.

The 64 scan time-series was subject to the least squares analysis using the first scan as a reference. The movement parameters reflecting estimated movement suggested translations (solid lines in the upper panel of Figure 2) in the order of $100\mu m$ or less. These translations will be referred to as Δx and Δy. The results of this analysis compared well with the equivalent AIR results (see the lower panel of Figure 2).

To demonstrate that these estimated movements were due to real movement the time-dependent hemodynamic variance was partitioned into a series of orthogonal spatial modes (or eigenimages) using SVD (the first 4 scans were ignored to avoid magnetic saturation effects). See Friston et al (1994) for a discussion of spatial modes and their identification. In brief these spatially distributed patterns reflect independent systems that share a common source of variance. Figure 3 (top right) shows the first spatial mode (that accounting for the most variance). This mode has been elicited by photic stimulation and includes striate and extrastriate regions. The time-dependent expression of this mode (the solid line in the top left of Figure 3) matches that predicted by photic stimulation (the dotted line). The dynamics predicted by photic stimulation were obtained by convolving the square wave input with a hemodynamic response function of seven seconds delay and dispersion (see Friston et al 1994).

Consider now the third mode - that accounting for the third largest source of variance (the second mode is not shown in this paper). Inspection of the third mode immediately suggests that it can be attributed to movement along a positively inclined diagonal. Figure 3 (bottom right) shows that the time-dependent expression of the third mode (solid line) is closely matched by least squares estimates (Δx and Δy) of diagonal motion $1.3.\Delta x + 1.7.\Delta y$ (dotted line). This concordance demonstrates that movements estimated by the leasts squares analysis are probably real.

The relative contribution of the first and third spatial modes can be expressed in terms of the appropriate 2-norm (a standard measure of how much a pattern or vector of voxel values contributes to the variance-covariance structure). For the original data these 2-norms are the

eigenvalues associated with the spatial modes and are presented in Figure 4 (left). The remarkable observation is that variance due to subject movement (of usually less than 100μm) was about one seventh of that introduced by the sensory stimulation. It is pleasing to see that this, and only this, source of variance (the third mode) was reduced following realignment of the images (Figure 4 right).

It may be asked at this point 'why not simply remove this mode from the time-series?'. Or more generally 'why not remove any component of all the modes that corresponds to a partial spatial derivative of the image?'. In fact this is equivalent to the least squares approach proposed.

Within modality, between subject

In this section we consider the more general problem of spatial transformation where the same imaging device has been used but the objects imaged (e.g. subjects' heads) are not exactly the same. Spatial or stereotactic normalization belongs to this class of problems when it is solved by matching an arbitrary image to some ideal image, model or template. For example in the intersubject averaging of PET activation studies, change distribution analysis (Fox et al 1988) and statistical parametric mapping (Friston et al 1991a) both require mapping into some standard space [the generally accepted international standard is the space described in the atlas of Talairach and Tournoux as proposed by Fox et al (1988)]. Friston et al (1991b) suggested that this mapping can be effected by 'matching' an individual's image with an ideal reference image, model or template, where the template conforms to the standard space in question. Clearly there are many other examples that could have been chosen for this section, but the one presented here is of immediate relevance for those engaged in functional mapping. This section also introduces two important extensions of the least squares approach which involve (i) a repeated or iterative application [where the spatially transformed object image $\Omega(q(x))$ is used recursively as a new object image $\Omega(x)$] and (ii) making two sets of images from the original image pair and simultaneously solving for the spatial transformation that conjointly maps one set of images onto the other set.

PET to PET spatial normalization

The problem in this section is more general than in the previous section because the allowable transformations are not precisely specified and $\gamma_x(.)$ - the functional component of the operator $f_x\{.\}$ is not known. This case is however not as general as situations considered later: We know that the resolution of the image process to be matched are the same (i.e. $\mathbf{c} = \mathbf{I}$ - the identity matrix) and consequently eqn(5) reduces to:

$$[\text{diag}(\dagger).\text{\ss}^f \ -\text{diag}(\partial\Omega/\partial\mathbf{x}).\text{\ss}^q].[\mathbf{u}_1 \ \ \mathbf{q}]^T \ \approx \ \ \Omega \qquad\qquad 7$$

assuming a first order approximation for $\gamma_x(.)$. This first order approximation can be justified given that the images are of the same modality. Notice also that the zeroth order term (the constant) has been omitted as we expect the zero counts in one image should correspond to zero counts in the other.. The basis functions ($\text{\ss}^f$ and $\text{\ss}^q$) can be arbitrary as long as they are smooth. In the present (and subsequent) example the basis functions for both $\mathbf{u}_1$ and $\mathbf{q}$ were the same and are shown in figure 5. These basis functions correspond to Fourier modes, obtained by incrementing the spatial frequencies by $\pi/2$ and orthogonalising the resulting basis set. These basis functions can be quite arbitrary as long as they are smooth and cover the 'space' of distortions in a reasonably comprehensive way. The exact nature and number of basis functions would probably vary from application to application. For simplicity we use the same set here throughout.

Spatial normalization

The object image $\Omega(x)$ was taken from the SPM (MRC Cyclotron Unit, London and see Friston et al 1991b) library of template images and corresponds to a transverse section lying 12mm above the intercommissural plane. The reference image $\dagger(x)$ was the corresponding slice from a normal subject acquired as described above. Although we present a spatial normalization of a single slice it should be noted that this normalization is generally applied in three dimmensions.

In this example we have used the 'template' image as the object image (the image subject to spatial transformation). The designation of which image is the object and which the reference is usually arbitrary: The spatial transformations from object to reference is given by:

$$q(x) \quad \equiv \quad \mathbf{x} \ + \ \mathbf{\beta}^q.\mathbf{q}$$

where $q(x)^{-1}$ is the complementary or inverse transformation from reference to object:

$$q(x)^{-1} \quad \equiv \quad \mathbf{x} \ - \ diag(\mathbf{I} + \partial\mathbf{\beta}/\partial\mathbf{x}.\mathbf{q})^{-1} \ \mathbf{\beta}^q.\mathbf{q} \qquad\qquad 8$$

This approximate equality is easy to derive using a first order Taylor expansion. The spatial and voxel value transformations were computed according to eqn(7) and the spatial transformations applied to $\Omega(x)$. The result [i.e. $\Omega(q(x)$] was substituted for $\Omega(x)$ in eqn(7) and the second spatial transformation added to the first. This process was repeated 3 times (see below). The results of this analysis are presented in Figure 6. Figure 6 shows the individual's transverse image (reference - top right) and the template PET image (object - top left). Figure 6 lower left shows the object image following spatial transformation. The spatial congruency between the resampled image and the reference image is seen and is largely a result of a counter-clockwise rotation (particularly of subcortical structures). The image on the lower right is also a spatially transformed object image, but the solution for the resampling was computed in a different way, as described next.

Comparison with nonlinear minimisation

The iterative solution of eqn(7) can be evaluated in relation to nonlinear minimisation [e.g. the Levenberg-Marquardt technique (More 1977)] by examining the trajectory of the paths taken in the recursive least squares approach and that taken by a nonlinear search in the space of the spatial transformation coefficients **q**. The nonlinear minimisation solution was obtained by minimizing the square of the 2-norm (over voxel locations x):

$$\| \dagger(x) \sum u_{1j} \beta f_j(x) \quad - \quad \Omega(x + \sum q_k \beta q_k(x)) \|^2 \qquad\qquad 9$$

as a function of $[u_{11} \ldots u_{1j} \ q_1 \ldots q_j] \equiv [\mathbf{u}_1 \ \mathbf{q}]$ using standard nonlinear minimisation (as implemented in Matlab, MathWorks Inc. Sherborn MA, USA). This solution is equivalent to solving eqn(7) but without the first order approximation implicit in eqn(3). The solutions for $\mathbf{q}$ were recorded at each iteration. The trajectory of these estimates was then compared with the trajectory traced out by the 4 recursive least squares solutions of eqn(7). These two trajectories were plotted in the 2-dimensional subspace of the search space associated with the largest singular values. This is simply a device to view the trajectories from a direction that reveals the greatest excursions.

Despite the fact that the spatially transformed images in Figure 6 (lower panels) look very similar the solutions obtained by least squares and nonlinear minimisation were different. The results of this trajectory analysis are presented below and illustrate two key points; (i) the least squares approach found the global minimum whereas the nonlinear minimisation did not and (ii) the search strategies differ fundamentally in their nature.

Figure 7 (top left and right) shows the singular distortions [c.f. principal warps (Bookstein 1988)] or spatial transformations associated with the largest two singular values following SVD of the sequence of solutions for $\mathbf{q}$ (using both the least squares and nonlinear minimisation). These distortions are simply a linear combination of the basis functions. This linear sum is determined by the singular vectors associated with the largest singular values. The two distortional modes (Figure 7 top left and right) correspond to the vertical and horizontal dimensions of the search subspace shown in Figure 7 (lower left panel). The corresponding error surface is depicted in the lower right panel. This surface reflects how 'far off' the match is, in terms of the two singular distortions. The error surface was computed by varying the first two singular distortions, whilst using the final nonlinear estimate for the other parameters (remaining singular distortions and $\mathbf{u}_1$) using eqn(9). It is immediately obvious that the two trajectories (least squares - solid and nonlinear minimisation -

broken) converge to the different solutions. The nonlinear minimisation found a local minimum. Furthermore the recursive least squares approach used a handful of steps whereas the nonlinear minimisation took over twenty iterations to converge. The number of floating point operations (flops) per iteration taken by the nonlinear minimisation was 2.7 times the number of flops required by one least squares solution. In short the least squares approach will usually be an order of magnitude more efficient than an iterative nonlinear equivalent.

It is instructive to contrast the explicit least squares and nonlinear minimisation. In anthropomorphic terms both approaches want to find the global minimum in a landscape defined by eqn(9) (the error surface in Figure 7). The nonlinear minimisation starts at an arbitrary location and, with acknowledged ignorance of all but the local landscape, sets of one step at a time always trying to chose the 'best' path based on local features. The least squares approach is more pretentious and makes some (possibly) unwarranted assumptions about the global shape of the landscape (e.g. it assumes the landscape is parabolic) and, based on local features, goes straight to the (anticipated) lowest point. However because the landscape is not quite as 'well behaved' as the least squares thinks, it may need to 'home in' on the global minimum; hence the iterative approach adopted in this section.

MRI to MRI

The last part of this section deals with a simple extension to the least squares approach. Here we deal with MRI to MRI matching and assume, as in the previous section, that $\mathbf{c} = \mathbf{I}$ and $\gamma_x\{.\}$ has only a first order term. In fact in this application we assume $f_x\{y\} = y$. Now consider some arbitrary function $\gamma_i\{.\}$ applied to both $\dagger(x)$ and $\Omega(x)$ where $\gamma_i\{\dagger(x)\} = \dagger_i(x) \equiv \dagger_i$ and $\gamma_i\{\Omega(x)\} = \Omega_i(x) \equiv \Omega_i$. Eqn(7) can be generalized to:

$$[\text{diag}(\dagger_i).\text{ß}f \quad -\text{diag}(\partial\Omega_i/\partial\mathbf{x}).\text{ß}q].[\mathbf{u}_1 \quad \mathbf{q}]^\top \quad \approx \quad \Omega_i$$

Because we assume $f\text{x}\{y\} = y$. this simplifies to:

$$\dagger_i \quad -\text{diag}(\partial\Omega_i/\partial\mathbf{x}).\text{ß}^q.\mathbf{q}^T \qquad \approx \qquad \Omega_i$$

Because the spatial transformations are the same irrespective of the intensity transformations $\gamma_i\{.\}$ we can solve for $\mathbf{q}$ simultaneously using several $\gamma_i\{.\}$ functions:

$$\dagger_1 \quad -\text{diag}(\partial\Omega_1/\partial\mathbf{x}).\text{ß}^q . \quad [\mathbf{u}_1 \quad \mathbf{q}]^T \quad \Omega_1$$
$$\dagger_2 \quad -\text{diag}(\partial\Omega_2/\partial\mathbf{x}).\text{ß}^q \qquad \approx \qquad \Omega_2$$
$$\dagger_3 \quad -\text{diag}(\partial\Omega_3/\partial\mathbf{x}).\text{ß}^q \qquad \qquad \Omega_3 \qquad\qquad\qquad 10$$

and so on. In simple terms this extension means that we can extract any feature from both the reference and object image and match the resulting images. Because the spatial transformation is the same for all possible features we can combine several feature-specific image pairs to estimate the best spatial transformation. This extension depends on the fact that the $f_x\{.\}$ has a simple form. A more general extension of this sort is presented in the discussion.

The following example is meant to illustrate this idea more clearly. In matching two homologous structural (T1 weighted) MRI images from different subjects one might simply use eqn(7) and proceed in an analogous way to the PET - PET normalization in the previous section. However in so doing one would be implicitly matching white matter to white matter (because in T1 weighted images white/non-white boundaries dominate). A useful extension would be to segment the images into gray matter and white matter and match gray matter to gray matter and white to white. This can be effected simultaneously with eqn(10) and is simply implemented by chosing $\gamma_i\{.\}$ to perform a (crude) segmentation. In the present example we used:

$$\gamma_i\{\Omega(x)\} \qquad = \qquad c(x) * \exp(-(\Omega(x) - v_i)^2/2\sigma^2) \qquad\qquad 11$$

where v_1 corresponds to the average gray matter intensity levels (0.52 of the image maximum) and v_2 corresponds to white matter levels (0.86 of the image maximum). σ controls the width of the 'segmentation window' (we used 0.3). $c(x)$ was a convolution kernel of FWHM 4 voxels to ensure that smoothness constraints were met. Note that this 'segmentation' is not very rigourous but is sufficiently good for the current purposes.

The results of this analysis are shown in figure 8. The object image (that to be spatially transformed) is seen in the upper left panel and the reference image is seen in the upper right panel. The lower panels show the transformed object images (lower left) and the same superimposed on the reference image (lower right panel). The reshaping of the object image is particularly pronounced in the bottom right (occipital) region (Figure 8 lower left). The two segmentation functions were applyed to both the spatially normalized object image and the reference MRI images (omitting the convolution) to reveal the conjoint spatial matching of gray matter (Figure 9 upper panels - transformed object and reference images) and white matter (lower panels). Generalization to three dimensions is only a question of computer resources, however the resources required to manipulate volumetric high resolution MRI data should not be underestimated.

Between modality, within subjects

This section is concerned with the problem of cross-modality registration. The importance of this problem lies in relating anatomy to functional organization in the brain. Structure-function relationships are being elucidated with increasing precision by mapping data from functional time-series onto an individual's high resolution MRI image. The problem considered here is to match a PET image to a transverse high resolution MRI scan obtained from the same subject. This problem, once solved (in three dimensions) uniquely specifies the orthogonal transformation which brings the MRI and PET data into alignment.

In this section we concentrate more on the intensity transformation (the expansion of $f_x\{.\}$). In previous sections this expansion was trivial because the images were acquired with the same modality and a simple relationship could be assumed. In this section there is no simple relationship

between MRI and PET voxel values even if the images were congruent. Indeed not only is this relationship highly nonlinear it is also likely to be highly non-monotonic and nonstationary. There are many forms that the expansion of $f_x\{.\}$ could take. We present here an expansion designed specifically for the current application and which uses the same linearising devise (Taylor's theorem and a first order approximation) as for the spatial transformation.

  If one accepts the conjecture that the source of the PET signal is predominantly from gray matter then the intensity transformation that maps an MRI image $\dagger(x)$ onto a PET image $\Omega(q(x))$ would involve gray matter segmentation $s_x(\dagger(x),v(x))$, nonstationary scaling and convolution into the resolution of the PET image. Mathematically one could express this as:

$$f_x\{\dagger(x)\} \quad = \quad c(x) * u_0(x).s_x(\dagger(x),v(x))$$

$$\text{where } s_x(\dagger(x),v(x)) \quad = \quad \exp(\,-(\dagger(x) - \{v_g + v(x)\}\,)^2/2\sigma^2\,) \qquad\qquad 12$$

where $c(x)$ is a convolution kernel, $u_0(x)$ is a nonstationary but smoothly varying scaling coefficient, σ is the 'standard deviation' of the segmentation kernel, v_g is the fixed estimate of gray matter intensity and $v(x)$ is the deviation from this estimate at a particular point in the brain x. Eqn(12) is not in a form that permits a linear least squares solution for the unknowns $u_0(x)$ and $v(x)$. In order to make eqn(12) amenable to a least squares analysis one can adopt the same approach used for the spatial transformation in eqn(3); namely using a first order approximation of the Taylor series. In this case the derivatives are not with respect to spatial distortions but the derivative of the segmentation function with respect to small deviations in the grey matter intensity $[\ \partial s(\dagger(x),0)/\partial v(x)\ ]$.

$$f_x(\dagger(x)) \quad \approx \quad c(x) * u_0(x).\{\ s(\dagger(x),0) \ + \ v(x).\partial s(\dagger(x),0)/\partial v(x)\ \}$$

$$\approx \quad c(x) * \{\ u_0(x).s(\dagger(x),0) \ + \ u_1(x).\partial s(\dagger(x),0)/\partial v(x)\ \}$$

As usual we can expand the nonstationary coefficients u_0 and u_1 in terms of basis functions $u_i(x) = \sum u_{ij} \beta f_j(x)$ giving, in matrix notation:

$$f_x(\dagger(x)) \quad \equiv \quad [\mathbf{c}.\mathrm{diag}(\dagger).\boldsymbol{\beta} f \quad \mathbf{c}.\mathrm{diag}(\partial \mathbf{s}/\partial \mathbf{v}).\boldsymbol{\beta} f].[\ \mathbf{u}_0 \ \mathbf{u}_1\ ]^T \qquad 13$$

The important thing to note here is that the *form* of the 'segmentation function' $f_x(\dagger(x))$ is the same everywhere but the actual function can vary from place to place. This is important because one cannot guarantee exactly the same relationship between MRI and PET voxel values in every part of the image (e.g. lesions, focal activations and field heterogenieties will introduce nonstationariness into the relationship).

Because this section deals with within subject transformations the spatial component is a rigid body transformation with six parameters. The simultaneous solution for both intensity and spatial transformations is given by the solution of:

$$[\mathbf{c}.\mathrm{diag}(\dagger).\boldsymbol{\beta} f \quad \mathbf{c}.\mathrm{diag}(\partial \mathbf{s}/\partial \mathbf{v}).\boldsymbol{\beta} f \quad -\partial\Omega/\partial\mathbf{q}].[\ \mathbf{u}_0 \ \mathbf{u}_1 \ \mathbf{q}]^T \quad \approx \ \Omega \qquad\qquad 14$$

for $\mathbf{u}_0 \ \mathbf{u}_1$ and $\mathbf{q}$. $\partial\Omega/\partial\mathbf{q}$ was estimated directly as in the section on realignment of fMRI time series. Because the effective resolution of the MRI image is approximated by its voxel dimensions and that of the PET image was about 8mm we chose $\mathbf{c}$ to correspond to a two dimensional convolution with a Gaussian filter with FWHM 8mm. v_g was set at 0.52 of the MRI image maximum and σ was 0.3.

The results of this analysis, for roughly homologous slices, are presented in Figure 10. Figure 10 shows the object image (best guess transverse slice from the PET volume - top left) and the reference image (an arbitrary slice from a volume MRI image of the same subject - top right). The lower image is the PET slice following spatial transformation. The congruence is demonstrated on the lower right were the spatially transformed PET image and the reference MRI image are superposed. In fact the registration here was already quite good and the spatial normalization

involved only a small translation and rotation. Of greater interest is the solution for the MRI intensity transformation:

The intensity transformation

Here we consider the intensity transformation that is implicitly applied to the MRI image in eqn(14). Hitherto this component of the solution has not been very interesting but in between-modality applications the transformation can be considered of primary importance and the spatial component of secondary interest (see applications section in discussion). The intensity transformed MRI image that best matches (in a least squares sense) the spatially transformed PET image is given by eqn(13). This estimate is shown in Figure 11 (bottom left) and can be thought of as an MRI image emulating a PET image. The correspondence with the real PET image (above) is apparent and remarkable good considering we made the somewhat unreasonable assumption that all the PET signal came from gray matter.

The intensity transformed MRI image obtained before convolution [setting $\mathbf{c} = \mathbf{I}$ in eqn(13)] is seen in Figure 11 on the right. This is the 'best' distribution of functional activity that could explain the observed PET scan before convolution with the point spread function, where the distribution is constrained by structural information in the MRI.image. Equivalently this is a least squares solution for the distribution of functional activity based on the underlying anatomy. We expand on the potential importance of this 'virtual modality' in the discussion.

The actual solution for $f_x\{.\}$ in eqn(13) clearly depends on position in the image. Two examples are given in Figure 12 according to the estimates obtained. The examples come from cortical and subcortical gray matter voxels indicated by black dots in the left lateral secondary sensory cortex and the left thalamus. The corresponding cortical (solid line) and subcortical (broken line) functionals $f_{\text{cortical}}\{.\}$ and $f_{\text{subcortical}}\{.\}$ are shown on the right of Figure 12. It is immediately obvious that the cortical gray matter activity is greater than the subcortical estimate. The fact that this relative difference in reversed in the PET image is due to partial voluming effects on signals from the cortical sheet. More importantly the gray matter intensity values for the cortical MRI

voxels were lower than for the subcortical voxels. I.e. the peak of $f_{\text{cortical}}\{.\}$ corresponds to a lower MRI voxel intensity than the $f_{\text{subcortical}}\{.\}$ peak.

In this section we have presented a simple (but still effective) example of PET to MRI matching. One could easily generalize the approach to include white matter contributions or indeed make the form of the segmentation kernel more general using a greater number of parameters.

Between modality, between subjects

This final section is included for completeness. Here we match (†) a transverse MRI slice from the individual used in the previous sections and (Ω) the corresponding transverse section from the the MRC library of PET templates used in the section on PET to PET normalization. This is the most general problem considered in this paper and requires a solution for the transformation coefficients with no simplifications. In this example we have deliberately degraded the MRI data by resampling to render voxel size the same as the PET data. Here the spatial transformation is not constrained to be affine and eqn(14) has to be generalized to

$$[\mathbf{c}.\text{diag}(\dagger).\text{\ss}^{f} \quad \mathbf{c}.\text{diag}(\partial\mathbf{s}/\partial\mathbf{v}).\text{\ss}^{f} \quad -\text{diag}(\partial\Omega/\partial\mathbf{x}).\text{\ss}^{q}].[\ \mathbf{u}_{0}\ \mathbf{u}_{1}\ \mathbf{q}]^{\mathrm{T}} \quad \approx\ \Omega \qquad 15$$

The basis functions ($\text{\ss}^{f}$ and $\text{\ss}^{q}$) for intensity and spatial transformations were the same as in previous sections (Figure 4). The expansion of $f_{x}(x)$ was in terms of a Gaussian segmentation kernel as in the previous section. $\mathbf{c}$ corresponded to a two dimensional convolution with a Gaussian kernel of FWHM of 8mm.

The results of this analysis are presented in Figures 13 and 14 using a similar format as in the previous section (Figure 10 and 11). The spatial congruence of the MRI image and transformed PET image and delimitation of gray matter in the MRI image are evident. The technique appears to be fairly resistant to the degradation of the MRI images.

Discussion

This paper has presented a generic approach to the spatial transformation of image processes. The approach is automatic (is noninteractive), finds explicit solutions (is noniterative) and solves general problems (is nonlinear). The technique depends on partitioning the differences between an object image and a reference image into two components. The first component is due to differences that remain after discounting the effects of image mis-registration or non-congruence. The second source of differences results from spatial translations, rotations and distortions. Both types of difference are modeled with expansions in terms of basis functions, polynomials, and Taylor series using low order approximations in such a way as to render the problem linear. The requisite spatial and image transformations are then solved for in a least squares sense.

In this paper we have focussed on the realignment of neuroimaging time-series, spatial normalization required for voxel-based analysis of activation studies, cross-modality registration and spatial normalization of images from different modalities. The nature of the transformation is defined by the constraints under which the transformation is applied. These constraints are embodied in the operational equations described. The minimal constraints adopted in this work pertain to the preservation of local contiguities and the local stationariness of operators which map one modality into another. These constraints can be framed in terms of smoothness; namely the spatial transformation represents a smooth mapping and the coefficients of any intensity transformation (functionals) change slowly with position. Smoothness is imposed by using smooth basis functions.

We have presented a series of ideas and techniques some of which are fundamental, some of which are not. We consider the following points to be important aspects of the approach

♦   Reliability takes precedence over validity. This requires a non-label based approach, that, in this instance, minimizes the sum of squares between two image processes.

♦   The differences between two images can be attributed to (i) non-congruence and (ii) differences extant when the images are in perfect register.

♦   The transformations which minimize these differences (in a least square sense), namely the spatial transformation and intensity transformation, are solved for simultaneously.

♦   Both transformations are defined in the space of some basis functions. The nature of the basis functions (e.g. smoothness) embody the constraints under which the transformations are effected.

♦   The minimal constraints on the basis functions relate to preservation of local contiguity relationships and the local stationariness of the intensity transformation. This is equivalent to using smooth basis functions.

♦   The solution for the spatial and voxel value transformations can, if appropriate, be formulated in linear terms, permitting an explicit least squares solution. The use of Taylor series and other expansions are particularly useful in this regard. If not appropriate standard nonlinear minimisation or gradient descent techniques can be used.

Although non-optimal Fourier basis functions are efficient in that they do not require empirical characterization of the spatial autocovariance functions (c.f. the Karhunen Loeve expansion or SVD).

Applications

The applications of spatial and intensity transformations are numerous. We have demonstrated applications to the realignment problem and spatial normalization. Here we consider other applications and extensions. Applications can be classified according to the goal of the

transformation. The objectives of most transformations fall under three general headings: (i) Reducing differences of a specific sort to facilitate comparison among images. (ii) Characterization of differences in spatial topography. (iii) Image restoration and segmentation.

The examples in this paper have focussed on the first class of applications in the sense that image realignment and normalization are usually implemented as a prelude to comparison among scans. The second class of applications asks 'what are the important modes of anatomical (topographic) variation?' or 'what are the anatomical differences between one set of scans and another?'. The answers to both these questions depend on a complete specification of the spatial topography of each image. We propose that this specification could be in terms of the *spatial distortion required to map an arbitrary image onto some reference*. For example the topography of an image can be characterized in terms of the coefficients corresponding to the spatial basis functions. This simple list of coefficients, taken in conjunction with the reference image, is a complete specification of the topography of the original image (down to the resolution imposed by the basis functions). The importance of this observation is that anatomical topography can be characterized by a multivariate measure (the coefficients) and subject to conventional multivariate statistics.

Important examples of this characterization could include the normal modes of anatomical variability defined on a series of MRI scans from normal subjects. The exciting concept here is that once normal modes of anatomical variation are established *they can then be used as the basis functions in the transformations*. The charm of this 'bootstrapping' is in constraining the transformations to lie in the space of normal anatomical variability. This would increase the face validity of the transformation and probably reduce the number of basis functions considerably. This is the subject of current work. Alternatively important neurodevelopmental modes can be identified using MRI scans obtained during development (Nick Lange personal communication). These modes are simply defined by SVD of the spatial transformations (or the coefficients of the spatial basis functions).

The third set of applications was referred to in the first section on PET-MRI matching. In this section we solved simultaneously for a spatial transformation and an intensity transformation that

jointly matched a PET scan and a MRI scan from the same subject. The intensity transformation, considered in isolation, assigns functional (PET) values to structural objects in the MRI image. This transformation can be appreciated from two points of view (i) It is a functional categorisation of anatomical structures of the sort implied by image segmentation. Alternatively (ii) the intensity transformed MRI image represents a least squares solution for the underlying flow distribution seen in the PET image. In this sense the transformed MRI image emulates a 'restored' PET image, where the restoration embodies anatomical information (from the MRI). This second perspective suggests it may be possible to analyze single subject functional time-series (e.g. PET activation studies) using not the original functional data but the transformed MRI images. This multimodality application is currently being explored.

Extensions

There is a potentially important extension to the technique that we did not demonstrate. This extension involves solving for the spatial and intensity transformations simultaneously for the images and the images convolved with any number of convolution kernels. This is simply effected by stacking the matrices in eqn(5) on top of each other after premultiplying by the appropriate convolution matrix. For example if we take an arbitrary convolution matrix **§** then from eqn(5)

$$\mathbf{\S}.[\mathbf{c}.\mathrm{diag}(f^0(\dagger)).\mathbf{\beta}^f \quad \mathbf{c}.\mathrm{diag}(f^1(\dagger)).\mathbf{\beta}^f... \quad -\mathrm{diag}(\partial\Omega/\partial\mathbf{x}).\mathbf{\beta}^q].[\ \mathbf{u}_0\ \mathbf{u}_1.... \quad \mathbf{q}]^T \quad \approx \quad \mathbf{\S}.\Omega$$

so for a series of convolution matrices **§**$_i$:

$$\mathbf{\S}_0.\mathbf{c}.\mathrm{diag}(f^0(\dagger)).\mathbf{\beta}^f \quad \mathbf{\S}_0.\mathbf{c}.\mathrm{diag}(f^1(\dagger)).\mathbf{\beta}^f... \quad -\mathbf{\S}_0.\mathrm{diag}(\partial\Omega/\partial\mathbf{x}).\mathbf{\beta}^q \quad .[\ \mathbf{u}_0\ \mathbf{u}_1..\quad \mathbf{q}]^T \qquad\qquad \mathbf{\S}_0.\ \Omega$$
$$\mathbf{\S}_1.\mathbf{c}.\mathrm{diag}(f^0(\dagger)).\mathbf{\beta}^f \quad \mathbf{\S}_1.\mathbf{c}.\mathrm{diag}(f^1(\dagger)).\mathbf{\beta}^f... \quad -\mathbf{\S}_1.\mathrm{diag}(\partial\Omega/\partial\mathbf{x}).\mathbf{\beta}^q \qquad\qquad\qquad \approx \qquad \mathbf{\S}_1.\ \Omega$$
$$\mathbf{\S}_2.\mathbf{c}.\mathrm{diag}(f^0(\dagger)).\mathbf{\beta}^f \quad \mathbf{\S}_2.\mathbf{c}.\mathrm{diag}(f^1(\dagger)).\mathbf{\beta}^f... \quad -\mathbf{\S}_2.\mathrm{diag}(\partial\Omega/\partial\mathbf{x}).\mathbf{\beta}^q \qquad\qquad\qquad\qquad\qquad \mathbf{\S}_2.\ \Omega$$

and so on.   16

For example $\S_0$ could represent the identify matrix and $\S_1$ a differential operator. In this case the transformation solutions would simultaneously approximate the original images (Ω and $\dagger$) and their first derivatives ($\mathbf{d\Omega/dx}$ and $\mathbf{d\dagger/dx}$). The constraints on $\S_i$ are that it should not make the first order approximation of the effect of spatial transformations unreasonable. This means the result of convolving Ω should still be smooth.

Limitations

  The limitations of the technique described here relate to the 'reasonableness' of the first order approximation in eqn(3) and any other constraints imposed by expanding in terms of basis functions, Taylor series or polynomials. The first order approximation [eqn(3)] is only a good one when the spatial distortions are small relative to smoothness. In a sense this is not a fundamental limitation because (i) the images can always be made sufficiently smooth using the formalism of the previous section [eqn(16)] or (ii) spatial differences can be successively reduced using the linear piece-wise approximation to a nonlinear search, as implemented in the recursive application (see above). Because the initial mismatch between object and reference should be small (relative to smoothness), whenever two images are matched the image with the poorer spatial resolution can be considered as Ω. Clearly one should start with the 'best guess' registration.

  Missing data appears to be handled in a reasonably graceful fashion as long as the least squares solution is restricted to voxels at which data exists for both image processes.   This is no problem in image registration because the affine transformations specified for one part of the image exactly specify the transformations everywhere. For nonlinear transformations if one part of the image is missing no solution can be obtained for the homologous part of the other image. In this sense it is advisable to ensure the reference image is more 'complete' than the object image.

  One potentially important limitation of smooth basis functions is that they are not always necessarily appropriate.   For example at the apposition of the cortices in the inter-hemispheric

fissure there is spatial proximity but contiguity across the falx cerebri is a biological impossibility. In this instance basis functions that are smooth over the mid-saggital plane may not be appropriate. In practice we have never found this to be a problem at the resolutions we commonly work with. If it were a problem then the basis functions could be redesigned, or equivalently each hemisphere could be normalized separately.

One practical limitation is the amount of working memory available to compute the least squares solution. This can severely compromise the volume of data that can be dealt with and the number basis functions (or the order of other expansions) that can be used. It should be noted that the actual software implementation does not adhere to the matrix equations above. For example one would never actually construct diag(†) but emulate the required matrix operation with the equivalent element by element operations (unless one had a software environment that could handle sparse matrices). In dealing with nonlinear spatial normalization of large volume data sets we have found the following strategy useful: (i) three dimensional affine and nonlinear transformations based on sparse sampling (i.e. subsampling) followed by (ii) two dimensional or piece-wise (e.g. slice) nonlinear deformations with complete sampling.

Issues of validity

The criteria for 'good' spatial transformations can be framed in terms of validity, reliability and computational efficiency. The validity of a particular transformation device is not easy to define or measure and indeed varies with the application. For example an orthogonal transformation may be perfectly valid for realignment but not for spatial normalization of an arbitrary brain into a standard stereotactic space. In general the sorts of validity that are important in spatial transformations can be divided into (i) *Face validity*, established by demonstrating the transformation does what it is supposed to and (ii) *Construct validity*, assessed by comparison with other techniques or constructs. In functional mapping face validity is a complex issue. At first glance face validity might be equated with the co-registration of anatomical homologues in two images. This would be complete and appropriate if the biological question referred to structural differences or modes of

variation. In other circumstances however this definition of face validity is not appropriate. For example the purpose of spatial normalization (either within or between subjects) in functional mapping studies is to maximize the sensitivity to neurophysiological change elicited by experimental manipulation of sensorimotor or cognitive state. In this case the better definition of a valid normalization is that which maximizes condition-dependent effects with respect to error (and if relevant intersubject) effects. This will probably be effected when functional anatomy is congruent. This may or may not be the same as registering structural anatomy.

In the present work we have addressed validity at a number of levels. In the first section we established face validity with respect to known simulated movement and construct validity with respect to another approach (Woods et al 1992). Another aspect of face validity was addressed by showing the algorithm removed a component from the functional variance that could be attributed to movement artifact. In the second section we examined construct validity in terms of standard nonlinear minimisation and trajectories on an error surface.

Finally with reference to the construction of 'virtual modalities' using intensity transformations: We emphasise that the validity of this 'modality' depends on the validity of the forms of the relationship, between voxel values in both modalities, that are assumed in the least squares solution.

Conclusion

In conclusion we hope to have presented a reasonable solution to a fairly simple problem: How to match one image to another reliably, quickly, automatically and with some degree of validity.

Note

Many of the algorithms presented in this paper have been implemented in MATLAB (MathWorks Inc, Sherborn MA, USA) as part of the SPM (Statistical Parametric Mapping) package. The spatial normalization component developed for PET uses a twelve parameter affine-six paramter nonlinear transformation in three dimensions to match an individual's image to a library template. Our ASCII files (which are interpreted by MATLAB) are available from the authors.

Acknowledgements

Appendix:

This appendix contains nothing new for the general reader. It is included for those who wish to verify the derivations in detail and/or implement the matrix equations in an application.

Consider the general problem of spatially transforming one image so that it matches another. The differences between the object images $\Omega(x)$ and the reference image $\dagger(x)$ can can be attributed to (i) differences in intensity (given the images are spatially congruent) and (ii) differences due to a spatial distortion of one image relative to the other. Let the first, generally nonlinear, relationship between voxel values from the same point (x) in both images be denoted by the operator $f_x\{.\}$ and the distortion by $q(x)$ such that:

$$f_x\{\dagger(x)\} \quad = \quad \Omega(q(x)) \ + \ e(x) \qquad\qquad a.1$$

where $e(x)$ is a normally distributed error term with zero mean. For clarity we will omit the error term and deal with one dimensional images (generalizing to three dimensional images at the end of the appendix). Images are assumed to be good lattice representations of the continuous processes $\Omega(x)$ and $\dagger(x)$ in a Euclidean space x.

The problem of spatial transformation reduces to finding approximate solutions for $f_x\{.\}$ and $q(x)$ subject to reasonable constraints. The constraint on the operator $f_x\{.\}$ is one of local-stationariness, in the sense that $f_x\{.\}$ does not change very much in a local region [$f_x\{.\}$ is allowed to vary between remote regions of the images]. The minimal constraint on the spatial transformation $q(x)$ is assumed to be a preservation of local contiguity relationships (i.e. $q(x)$ is a smooth and the associated Jacobian is positive). We will deal here with the most general case of these minimal constraints.

At first glance a.1 may appear so ill posed as to make any explicit solution impossible: However if we assume the images are smooth (or that they can be rendered smooth - see main text) then a

first order approximation of a.1 can be constructed in which both the operator $f_x\{.\}$ and q(x) have a least squares solution. The remainder of this section describes how this is done.

First decompose $f_x\{.\}$ into a convolution and a functional:

$$f_x\{.\} \quad = \quad c(x) * \gamma_x(.) \qquad\qquad\qquad \text{a.2}$$

where * denotes convolution. c(x) is a kernel or differential point spread function. Without loss of generality one can take some expansion of $\gamma_x\{.\} = \Sigma\ u_i(x).f^i\{\dagger(x)\}$ where $f^i\{.\}$ could correspond to terms in a polynomial, a Taylor series, a Fourier series and so on. The local-stationariness constraint on $f_x\{.\}$ can be implemented by expanding the coefficients of $f^i\{.\}$ [i.e. $u_i(x)$] in terms of some 'smooth' spatial basis functions $\text{\ss}^{f_j}(x)$:

$$u_i(x) \quad = \quad \Sigma\ u_{ij}\text{\ss}^{f_j}(x) \qquad\qquad\qquad \text{a.3}$$

q(x) can be similarly expanded in terms of x and some smooth basis functions $\text{\ss}^{q_k}(x)$:

$$q(x) \quad = \quad x + \Sigma\ q_k\text{\ss}^{q_k}(x) \qquad\qquad \text{a.4}$$

$$\text{and:} \qquad \partial q(x)/\partial q_k \quad = \quad \text{\ss}^{q_k}(x) \qquad\qquad\qquad \text{a.5}$$

Using these expansions a.1 becomes:

$$c(x) * \Sigma\ \{\ f^i(\dagger(x))\ \Sigma\ u_{ij}\text{\ss}^{f_j}(x)\ \} \quad \approx \quad \Omega(x + \Sigma\ q_k\text{\ss}^{q_k}(x)\ ) \qquad \text{a.6}$$

If $\Omega(x)$ is smooth the effects of small distortions $q_k\text{\ss}^{q_k}(x)$ will not interact to a significant degree and we can expand the right hand side of a.6 using Taylor's theorem where, ignoring high order terms:

$$\Omega(x + \sum q_k \text{ß}^{q_k}(x)) \quad \approx \quad \Omega(x) \quad + \quad \sum q_k \text{ß}^{q_k}(x).\partial\Omega(x)/\partial x \qquad \text{a.7}$$

a.7 is asymptotically true for small q_k and reasonably true, if $\Omega(x)$ is smooth, for larger q_k. The right hand side of a.7 can be expressed directly in terms of q_k by noting that from eqn(5):

$$\text{ß}^{q_k}(x) \cdot \partial\Omega(x)/\partial x \quad = \quad \partial q(x)/\partial q_k.\partial\Omega(x)/\partial x \quad = \quad \partial\Omega(q(x))/\partial q_k \qquad \text{a.8}$$

i.e. $\quad \Omega(x + \sum q_k \text{ß}^{q_k}(x)) \quad \approx \quad \Omega(x) \quad + \quad \sum q_k \partial\Omega(q(x))/\partial q_k \qquad \text{a.9}$

this first order approximation (a.7 or a.9) is substituted into a.6:

$$c(x) * \sum\sum [f^i(\dagger(x)) \cdot u_{ij} \text{ß}^{f_j}(x)] \; - \; \sum q_k \text{ß}^{q_k}(x).\partial\Omega(x)/\partial x \quad \approx \quad \Omega(x) \qquad \text{a.10}$$

Given the 'good lattice' assumption a.10 can be expressed in matrix notation as:

$$[\mathbf{c}.\text{diag}(f^0(\dagger)).\text{ß}^f \; \mathbf{c}.\text{diag}(f^1(\dagger)).\text{ß}^f... \; -\text{diag}(\partial\Omega/\partial\mathbf{x}).\text{ß}^q].[\; \mathbf{u}_0 \; \mathbf{u}_1.... \quad \mathbf{q}]^T \quad \approx \quad \Omega$$

or $\quad [\mathbf{c}.\text{diag}(f^0(\dagger)).\text{ß}^f \; \mathbf{c}.\text{diag}(f^1(\dagger)).\text{ß}^f... \; -\partial\Omega/\partial\mathbf{q}].[\; \mathbf{u}_0 \; \mathbf{u}_1.... \quad \mathbf{q}]^T \quad \approx \quad \Omega \qquad \text{a.11}$

where $\mathbf{c}$ is a Toeplitz matrix of the convolution kernel $c(x)$. $\text{diag}(f^i(\dagger))$ represents a diagonal matrix with leading diagonal elements $f^i(\dagger(x))$ at the location of all voxels. The matrix ß^f and ß^q contain a basis function in each column and have the same number of rows as voxels analyzed. $\mathbf{u}_i$ and $\mathbf{q}$ are row vectors of the unknown coefficients. Ω and $\partial\Omega/\partial\mathbf{x}$ are column vectors with one element per voxel. Let :

$$\mathbf{A} \quad = \quad [\mathbf{c}.\text{diag}(f^0(\dagger)).\text{ß}^f \; \mathbf{c}.\text{diag}(f^1(\dagger)).\text{ß}^f... \; -\text{diag}(\partial\Omega/\partial\mathbf{x}).\text{ß}^q]$$

then from a.11 $\mathbf{A}.[\ \mathbf{u}_0\ \mathbf{u}_1....\quad \mathbf{q}]^{\mathrm{T}}\qquad \approx\qquad \Omega$

and $\qquad [\ \mathbf{u}_0\ \mathbf{u}_1....\quad \mathbf{q}]^{\mathrm{T}}\qquad \approx\qquad (\mathbf{A}^{\mathrm{T}}.\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}.\Omega\qquad\qquad$ a.12

a.12 represents the least squares solution of the unknown coefficients $[\ \mathbf{u}_0\ \mathbf{u}_1....\quad \mathbf{q}]^{\mathrm{T}}$ The elements of these vector specify the approximations of $f_x\{.\}$ and q(x) in the space defined by the (generally nonlinear) basis functions. The corresponding expressions in three dimensions (x, y and z) are:

$\mathbf{A}\qquad =$

$[\mathbf{c}.\mathrm{diag}(f^0(\dagger)).\mathbf{\text{ß}}^f\ \ \mathbf{c}.\mathrm{diag}(f^1(\dagger)).\mathbf{\text{ß}}^f...\ \ -\mathrm{diag}(\partial\Omega/\partial\mathbf{x}).\mathbf{\text{ß}}^q\ \ -\mathrm{diag}(\partial\Omega/\partial\mathbf{y}).\mathbf{\text{ß}}^q\ \ -\mathrm{diag}(\partial\Omega/\partial\mathbf{z}).\mathbf{\text{ß}}^q]$

and $\qquad [\mathbf{u}\ \ \mathbf{v}\ \ \mathbf{q}_x\ \ \mathbf{q}_y\ \ \mathbf{q}_z]^{\mathrm{T}}\qquad \approx\qquad (\mathbf{A}^{\mathrm{T}}.\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}.\Omega$

where (c.f. a.4): $\qquad q(x)\qquad \equiv\qquad \mathbf{x}\ \ +\ \ \mathbf{\text{ß}}^q.\mathbf{q}_x\qquad\qquad$ a.13

and similarly for q($\mathbf{y}$) and q($\mathbf{z}$). The spatially normalized image $\Omega_n(x,y,z)$ is simply:

$$\Omega_n(x,y,z)\qquad =\qquad \Omega(q(x),q(y),q(z))\qquad\qquad \text{a.14}$$

Table of Symbols: (Could be placed in a Box)

$\dagger(x)$ and $\dagger$ — The image designated the template or reference image

$\Omega(x)$ and Ω — The image designated the object or observed image

$f_x\{.\}$ $=$ $c(x) * \gamma_x(.)$ The operator mapping voxel values from reference to object at x

$c(x)$ and $\mathbf{c}$ — The convolution kernel or matrix

$\gamma_x(.)$ $=$ $\sum u_i(x).f^i\{.\}$ — The nonlinear functional component of $f_x\{.\}$ at x

$f^i\{.\}$ — The ith function in the expansion of $\gamma_x(.)$

$u_i(x)$ $=$ $\sum u_{ij}\beta f_j(x)$ The ith coefficient of the expansion of $\gamma_x(.)$ at x

u_{ij} and $\mathbf{u}_i$ — The coefficients of the expansion of $u_i(x)$ in space (an unknown)

$\beta f_j(x)$ and $\boldsymbol{\beta}f$ — The spatial basis functions of the expansion of $u_i(x)$ in space

$q(x)$ $=$ $x + \sum q_k\beta q_k(x)$ — The function describing the spatial transformation or movement

$\beta q_k(x)$ and $\boldsymbol{\beta}q$ — The spatial basis functions of $q(x)$ - x

q_k and $\mathbf{q}$ — The coefficient of the expansion of $q(x)$ (an unknown)

$\partial\Omega(x)/\partial x$ and $\partial\Omega/\partial\mathbf{x}$ — The derivative of the object with respect to x

$\partial\Omega(q(x))/\partial q_k$ and $\partial\Omega/\mathbf{dq}$ — The derivative of the object with respect to q_k

References

Alpert NM, Bradshaw JF, Kennedy D and Coreia JA. (1990) The principal axis transformation - a method for image registration. J Nucl Med 31:1717-22

Amit Y, Grenander U and Piccioni M (1991) Structural image restoration through deformable templates. J. Am. Statistical Assoc. 86: 376-387

Bookstein FL (1989) Principal warps: Thin plate splines and the decomposition of deformations. IEEE Trans Pattern Anal. Mach. Intell. PAMI11:567-85

Collins DL, Neelin P, Peters TM and Evans AC (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18:192-205

Fox PT Mintun MA Reiman EM Raichle ME Enhanced detection of focal brain responses using intersubject averaging and change distribution analysis of subtracted PET images J Cereb Blood Flow Metab. 1988;8:642-653

Fox PT, Mintun MA (1989) Non-invasive functional brain mapping by change distribution analysis of averaged PET images of $H^{15}O_2$ tissue activity J.Nucl. Med. 30:141-149

Friston, K.J Frith, C.D Liddle, P.F and Frackowiak, R.S.J. (1991a) Comparing functional (PET) images: The assessment of significant change. J. Cereb. Blood Flow Metab. 11:690-699.

Friston KJ Frith CD Liddle PF & Frackowiak RSJ (1991b) Plastic transformation of PET images J. Comput. Assist. Tomogr. 15:634-639

Friston KJ, Jezzard P and Turner R (1994) Analysis of functional MRI time-series. Human Brain Mapping 2:69-78

Kosugi Y, Sase M, Kuwatani H, Kinoshita N, Momose T, Nishikawa J and Watanabe T (1993) Neural network mapping for nonlinear stereotactic normalization of brain MR images. J Comput. Assist, Tomogr. 17:455-60

Lange N (1994) Some computational and statistical tools for paired comparisons of digital images. Statistical methods in medial research 3:23-40

More JJ (1977) The Levenberg-Marquardt Algorithm: Implementation and theory, Numerical Analysis, ed Watson GA, Lecture notes in mathematics 630, Springer Verlag, p105-116

Pelizzari CA, Chen GTY, Spelbring DR, Weichselbaum RR and Chen CT (1988) Accurate three-dimensional registration of CT, PET and MR images of the brain. J Comput Assist Tomogr 13:20-6

Spinks TJ Jones T Bailey DL Townsend DW Grootnook S Bloomfield PM Gilardi MC Casey ME Sipe B and Reed J (1992) Physical performance of a positron tomograph for brain imaging with retractable septa. Phys. Med. Biol. 37:1637-1655

Steinmetz H and Seitz RJ (1991) Functional anatomy of language processing: Neuroimaging and the problem of individual variability. Neuropsychologia 29:1149-60

Talairach J. and Tournoux P. Co-planar stereotaxic atlas of the human brain Stuttgart:Thieme 1988

Woods RP, Cherry SR and Mazziotta JC (1992) Rapid automated algorithm for aligning and reslicing PET images. J Comput. Assist. Tomogr. 16:620-33

Legends for Figures

Figure 1

Comparison of the least squares technique with known (simulated) movement parameters and parameter estimates using AIR (Woods et al 1992). Left: y translation expressed in mm over the different simulated movements. The two solid lines correspond to the actual and least squares estimates of movement. The least squares estimate deviates on the extreme right (lower curve) from actual movement. The AIR estimates (broken line) are largely superimposed on the actual movement curve. Right: The same data plotted on a semi-log scale. *Straight line* - actual, *solid line* - least squares and *broken line* - AIR.

Figure 2

Estimated movement from a real time series of 64 coronal fMRI slices through visual cortex. Upper x and y translation (solid lines) and rotation (broken line) as estimated by the least squares approach. Lower: x translation estimated using AIR regressed on least squares estimates.

Figure 3

SVD analysis of the fMRI time series referred to in the previous figure. Left: first (top) and third (lower) spatial modes or eigenimages following SVD. The grayscale is arbitrary and the images have been normalized to their maximum. Right: Time dependent expression of the spatial modes. Right top: the periodic expression of the first mode due to photic stimulation (solid line) and that predicted by convolving the stimulation waveform with the hemodynamic response function, assuming a delay and dispersion of seven seconds (broken line). See Friston et al (1994) for a full exposition. Right lower: time-dependent expression of the third spatial mode due to movement artefact (solid line). The broken line corresponds to estimated diagonal movement using on the data in the previous Figure (Figure 3).

Figure 4

Singular (squared) or eigenvalue spectrums following an SVD analysis of the fMRI time series before and after spatial transformation (realignment). The key difference is a decrease in the amount of variance attributable to the third spatial mode. This can be interpreted as a removal of movement artefacts

Figure 5

The basis functions used in subsequent nonlinear spatial transformations and intensity transformation expansions. The gray scales are arbitrary and each function has been normalized to its maximum.

Figure 6

Spatial normalization of PET images effected by spatial normalization of one PET image to another. The object image (top left) is presented after spatial normalization (lower left) and shows a greater degree of correspondence with the reference image (top right). The equivalent spatial transformation of the object image estimated using nonlinear minimisation is shown on the lower right. All images have been scaled to their image maximum and are displayed on the proportional grid used by the atlas of Talairach and Tournoux (1988).

Figure 7

Comparison of the least squares approach and nonlinear minimisation in terms of searches over an error surface. Top: singular warps or distortions defined by a SVD of the search trajectories in the space of the basis functions. These two singular warps correspond to the axes in the lower panels. The two warps accounted for the greatest excursions of the search trajectories and are orthogonal. Each singular warp has been applied to the object image (Figure 6 - top left).

Lower left. Search trajectory in the space defined by the singular warps. *solid line* - least squares, *broken line* - nonlinear minimisation. Lower right: The corresponding error surface computed as a two dimensional subspace of the search space. This subspace includes the final solution of the nonlinear minimisation algorithm.

Figure 8

Matching MRI to MRI. Left: The object images before (upper) and after (lower) spatial transformation. The reference image that has been spatially approximated in shown on the upper right. The spatially transformed object and reference images have been superimposed to illustrate the coregistration (lower right). All images have been scaled to their maximum.

Figure 9

The efficacy of the spatial transformation in terms of the (crudely) segmented MRI images. Top: gray matter segmented images following spatial normalization of the object image (left) and the reference image (right). Lower: equivalent images segmented for white matter. All images have been scaled to their maximum.

Figure 10

Cross modality spatial transformations. Left: transverse PET slices before (top) and after (lower) spatial normalization to a reference MRI image (top left). The coregistration is illustrated in the lower right panel. All images have been scaled to their maximum.

Figure 11

Cross modality intensity transformations. Upper left: spatially transformed PET image (as in Figure 10 lower left). Right: The intensity transformed MRI image without convolution. After convolution (lower left) this image approximates the PET image in a least squares sense. All images have been scaled to their maximum.

Figure 12

  The transformation functions for two gray matter voxels in the MRI image. Left: the position of the two voxels in cortical and subcortical regions (black dots). Right: the corresponding intensity transformation functions *solid line* - cortical voxel, *broken line* - subcortical voxel.

Figure 13

  As for Figure 10 but in this instance the PET and MRI images are subject to nonlinear spatial transformations because they did not derive from the same subject. All images have been scaled to their maximum.

Figure 14

  As for Figure 11 but in this instance the PET and MRI images are subject to nonlinear spatial transformations because they did not derive from the same subject and the original MRI image is shown on the top left.