ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

NeuroImage

NeuroImage 20 (2003) 1425–1437

www.elsevier.com/locate/ynimg

Comments and Controversies

Zen and the art of medical image registration: correspondence, homology, and quality

W.R. Crum,* L.D. Griffin, D.L.G. Hill, and D.J. Hawkes

Division of Imaging Sciences, The Guy's King's and St. Thomas' School of Medicine, London SE1 9RT, UK

Received 27 December 2002; revised 17 June 2003; accepted 11 July 2003

Abstract

Nonrigid registration (NRR) is routinely used in the study of neuroanatomy and function and is a standard component of analysis packages such as SPM. There remain many unresolved correspondence problems that arise from attempts to associate functional areas with specific neuroanatomy and to compare both function and anatomy across patient groups. Problems can result from ignorance of the underlying neurology which is then compounded by unjustified inferences drawn from the results of NRR. Usually the magnitude, distribution, and significance of errors in NRR are unknown so the errors in correspondences determined by NRR are also unknown and their effect on experimental results cannot easily be quantified. In this paper we review the principles by which the presumed correspondence and homology of structures is used to drive registration and identify the conceptual and algorithmic areas where current techniques are lacking. We suggest that for applications using NRR to be robust and achieve their potential, context-specific definitions of correspondence must be developed which properly characterise error. Prior knowledge of image content must be utilised to monitor and guide registration and gauge the degree of success. The use of NRR in voxel-based morphometry is examined from this context and found wanting. We conclude that a move away from increasingly sophisticated but context-free registration technology is required and that the veracity of studies that rely on NRR should be keenly questioned when the error distribution is unknown and the results are unsupported by other contextual information.

And what is good, Phædrus,
And what is not good . . .
Need we ask anyone to tell us these things ?
(Zen and the Art of Motorcycle Maintenance by Robert Pirsig)

Introduction

Medical image registration is a mature research field with theoretical support and two decades of practical experience. Brain applications have benefited the most from registration to the extent that automated nonrigid registration (NRR) is now a standard component of widely used analysis packages such as SPM (http://www.fil.ion.ucl.ac.uk/spm). The problem of establishing correspondence between images of biological structures is central to the image registration task, yet in nearly all problems of interest, the true correspondences are unknown and are inferred by matching imaged surrogates such as anatomical landmarks and structural boundaries. With the wider availability of different variants of NRR come an increasing number of published studies that rely on the validity of correspondences suggested by these techniques. However, the most widely used methods are essentially dumb in that, for a particular registration task, they report only a measure of image similarity which does not allow a judgement of "success" or "failure" to be made. Worse, the magnitude and spatial distribution of errors in NRR are unknown and an understanding of exactly how image-similarity measures and arbitrary transformation models combine in the matching of complex sets of imaged features remains out of reach.

In this paper we consider the definition and meaning of correspondence in the context of brain image registration to

* Corresponding author. Division of Imaging Sciences, Thomas Guy House (5th Floor), Guy's Hospital, London SE1 9RT, UK. Fax: +44-0-20 7955-4532.

E-mail address: bill.crum@kcl.ac.uk (W.R. Crum).

1053-8119/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2003.07.014

understand the shortcomings of current registration techniques and the consequences for clinical studies that rely on registration. Our perspective is predominantly from the world of medical image analysis where registration problems are nearly always couched in terms of solving the correspondence problem and registration errors are expressed as positional variances. Achieving correspondence is the primary goal with evaluation of reliability and reproducibility forming the backbone of validation strategies. By contrast, some of the approaches used in the computational anatomy and brain mapping community have deliberately sacrificed a degree of face validity (the ability to independently gauge the correctness of the approach) to ensure construct validity (the validity of the approach by comparison with other constructs) (Friston et al., 1995). With the increasing sophistication of studies relying on registration and particularly the large-scale application of techniques originally applied in studies of PET or fMRI time series to structural data, interpretation of experimental results is relying more and more on a proper understanding of the correspondence problem and the limitations of methods which have been employed to solve it.

After a brief review of the registration task, we examine definitions and uses of correspondence in image registration and consider the relationship between registration, correspondence, and homology. The usefulness of current validation strategies is considered and implications for the use of voxel-based morphometry (Bookstein, 2001; Ashburner and Friston, 2001) in clinical studies are discussed. We conclude that prior knowledge of the application context should be incorporated into similarity measures, transformation models, and registration techniques to allow the results of clinical studies to be properly evaluated. Furthermore, we suggest that in the present climate, clinical studies whose results rely heavily on registration techniques of questionable validity should be treated with suspicion.

Background

In many ways the methodologies underlying current registration techniques have not changed significantly over the last 10 years. In all cases it is necessary to:

- specify a similarity or error measure describing the "goodness" of the match,
- specify a transformation model (e.g., rigid body, affine, elastic, fluid, B-spline),
- specify an interpolation strategy (e.g., nearest neighbour, trilinear, sinc),
- find the transformation parameters to maximise the similarity measure.

This minimal description does little justice to the sheer inventiveness that has been directed toward solving registration problems (Hill et al., 2001). There are four broad categories of medical images which feature in registration and registration tasks can be defined either within these categories or across them:

1. Structural, e.g., CT, MR, X-ray, US;
2. Functional, e.g., PET, SPECT, fMRI, contrast-enhanced CT/MR/US
3. Histological, e.g., digital images of postmortem brain slices;
4. Physical, e.g., the view through an operating microscope.

Registration outside the brain is often fraught with additional difficulties due to differences in patient positioning, deformation of soft-tissue structure, and scanning protocol (e.g., whether cardiac gating or breath holding is used). Registration of brain images is an exemplar that encapsulates most of the current debate on correspondence. Unlike most other organs, the question of structural and functional localisation in the brain is the subject of continuing study that has wide-reaching implications for future definitions of correspondence (Bjaalie, 2002; Brett et al., 2002).

The tasks of feature detection and segmentation are closely related to registration. The more corresponding features which can be identified in an image pair, the more tightly constrained the registration problem even in the absence of prior knowledge about the "true" transformation model. Labels present on both images represent powerful constraints on the possible correspondences.[1] A comprehensively labeled pair of images solves the correspondence problem but inaccurate or incomplete labeling allied with a set of trial correspondences is more common. There is a hierarchy of features which can be used to drive registration, the most tightly constraining of which are point anatomical landmarks which can be located to voxel (or subvoxel) accuracy. Edges delineate structures and features within structures. Segmentations of structures provide a priori volume correspondence which does not provide information (necessarily) about what happens within that volume but which forces the labeled set of voxels in one image to map to the labeled set in the second image. In practice these labelings are generally used to define closed boundaries and a diffeomorphic transformation model ensures that the labeled regions remain intact after registration.

Some sophisticated approaches have matched structures using labeled surface geometry. In one early approach (Thompson and Toga, 1996), selected sulci, chosen because they naturally decompose the brain into architectural compartments, were meshed and used to generate probabilistic models of anatomical variation. These models are effective in capturing structural variability and can help identify subtle anatomical trends such as those found in childhood development (Thompson et al., 2000b). These techniques have found wide application in the construction and fitting of anatomical atlases to model both normal and disease-specific brain variability (Toga et al., 2001). Related tech-

[1] Errors in the labeling process must be taken into account.

niques extract geometrically significant points on the boundaries of structures of interest and then match boundaries together, assuming elastic properties away from strong geometrical features (Davaztikos, 2001). All these techniques are distinguished by focusing strongly on the identifiable geometry of brain structures, thereby introducing a high-degree of anatomical context. However at present they are only applicable to specific delineable structures in the brain and may be unsuitable for studies of large populations due to the level of skilled manual intervention required. For that reason this paper focuses on approaches which match anatomy on the basis of voxel-intensities as these are the methods to which the neuroimaging community has established access.

Registration using voxel intensities usually maximises an image similarity measure. The most widely used measures which assume only a statistical relationship between voxel intensities in registered images are probably Mutual Information (Viola, 1995; Collignon, 1998; Maes et al., 1997), and Normalised Mutual Information (Studholme et al., 1999) together with the Correlation Ratio (Roche et al., 1998). These are low-level measures that are functions of the voxel-intensity distribution and are assumed to reach an extremum when the images are optimally matched. When stronger assumptions about the relationship between voxel intensities in the images exist, similarity measures such as the ratio image uniformity and the squared voxel intensity difference (Woods et al., 1998a, 1998b) can be used that place tighter constraints on the registration; the latter measure in particular has been found useful in conjunction with elastic and viscous fluid models of tissue deformation (Christensen et al., 1996). All these measures are also indirect functions of the parameters of the transformation model as the model specifies the set of trial voxel correspondences over which the similarity measures is evaluated.[2] Thus registration is in general a multiparameter (constrained) optimisation problem. Registration using these similarity measures has had remarkable success and, in the few gold-standard validation studies, has performed very well at establishing correspondences in restricted circumstances (e.g., West et al., 1997). There are many outstanding issues, however. There is an implicit assumption that image features, especially step changes in intensity, are corresponding features. However, similarity measures cannot generally distinguish voxel intensities arising from tissue properties from those arising through image artifacts and subjective judgment usually decides whether images are of sufficiently high quality for registration to succeed and be worthwhile. The most common reason that images may be of poor quality is patient motion or noncompliance and some limited attempts have been made to quantify this artifact as an acceptance criterion for registration (Freeborough and Fox, 1997). In practice it is not necessary for there to be a one-to-one correspondence between all features in a pair of images for registration to succeed if the degrees of freedom of the transformation model are low (e.g., rigid registration between CT and MR). However there is currently no way to assess the a priori likelihood of achieving correspondence in individual cases and it is possible that only some portions of images will be well registered, i.e., a local rather than a global extremum of the similarity measure is found.

The range and ambition of registration problems in medical imaging is increasing but fundamental problems remain because of

1. naïve association of image similarity with biological correspondence,
2. context-free similarity measures,
3. arbitrary transformation models.

A proper understanding of correspondence and errors in correspondence, which will necessarily be application specific, would help to address many of these issues. An understanding of correspondence would lead to more principled similarity measures and a more informed choice of transformation models.

What is correspondence?

Imagine two images A and B composed of the same large number of uniquely labeled volume elements. Each image contains the same set of labels and so any single element in A has a unique partner in B that can in principle be searched for systematically. This search can be repeated for all elements to generate a dense vector field linking every point A to its unique partner B. Correspondence has been established and the vector field can be parameterised as an instance of a transformation model. In practice this one-to-one labeling assumption breaks down because: (1) voxels comprising digital images do not each possess a meaningfully unique intensity; connected voxels representing biological structures often have the same or very similar intensities; (2) partial-volume effects during the imaging process can blur the signal from biologically distinct tissues; (3) tissue motion, deformation, or fluid flow during image acquisition may mean that the intensity of a single voxel is composed of signal originating from a wider locale; (4) imaging noise corrupts the true signal; (5) imaged features may be inconsistent.[3] If there is not a one-to-one correspondence of labels then it follows that there is not a unique spatial correspondence between the images. Therefore, a unique voxel-level transformation cannot be defined purely from the information present in the images. Although there are many possible transformations that could match a set of identifiable image features, the particular transformation chosen will determine what happens in homogeneous re-

[2] Except for any correspondences specified from prior knowledge.

[3] That is, biological features that are not present in both images either because of physical change or because of differences in imaging modality or both.

gions between those features where voxels lack a unique context. Good registration can be achieved by a combination of explicitly identified corresponding features and an appropriate transformation model. A dense set of verifiable feature correspondences could provide the same set of correspondences as a sparse set which is used to instantiate a well-defined transformation. If the transformation model is known and the set of feature correspondences is sufficiently dense to globally define an instance of the model then a perfect registration is achieved as correspondence can be inferred everywhere. However, if either the transformation model is unknown or only approximately known, or the set of feature correspondences is not large enough to uniquely instantiate the model, then a lower quality registration results. The particular transformation model can be of significant practical importance in applications such as image-guided surgery where the location of a structure seen only in the preoperative images is inferred using a transformation model instantiated using other features.

The spatial scale of measured correspondence is set by the spatial resolution and voxel dimensions of the imaging modalities and is typically in the range [0.1, 10.0] mm for brain applications. Therefore registration is matching features significantly above the cellular level ($\sim$10 μm). Registration works well because larger biological structures persist both in individuals with a remarkably constant morphology and across individuals with a remarkably similar morphology even though the constituent cells are different. For intra-individual registration of *normal* anatomy the implicit assumption is often that corresponding points in the registered images mapped back to the in vivo organ will coincide. With suitable mathematical models of anatomy and imaging physics, this assumption could form the basis of a validation strategy (Schnabel et al., 2003). A similar model is used by Davatzikos (2001) where curvature functions summarising the significant geometric structure of edges are assumed to define correspondences. The method assumes an explicit link between geometrical features and underlying biology — *"geometry has a biological substrate"*; equally it could be suggested that registration is exploiting the fact that *imaged biology has a geometrical substrate.* Between individuals, the variability of the cerebral cortex precludes establishing correspondence at the level of individual gyri. It has been suggested that in these cases the fundamental definitions of correspondence need to be well below the scale of imaged morphology at the level of cytoarchitectural[4] borders (Roland et al., 1997). This is inconsistent with the level of correspondence currently achieved by registering structural images.

An intuitive definition of structural correspondence is that by visual inspection, identifiable structures are mapped to one another, e.g., Warfield et al. (2001), where "We define alignment as every voxel of an acquisition being in correspondence with precisely the same anatomy in each scan." This implies correspondence in the location of structure boundaries or of boundaries of features within structures. The required correspondences are often defined to suit the application and are limited by the available image data, experimental design, and prior knowledge in the form of labels or models. For instance, when a mapping is sought between an atlas and an unlabeled brain image for the purpose of segmentation, a correspondence sufficient to map labels from the atlas to the brain image is sufficient. These correspondences need not correctly describe detailed structural differences within labeled regions or correctly map functional areas. Most approaches assume that one-to-one correspondence exists but there are two common classes of registration task where it is known a priori that one-to-one correspondence does not exist: (1) the presence of pathology in one image but not the other (or equally, different stages of pathology, or the surgical removal of tissue) and (2) intersubject registration where the anatomy may vary. Additionally in intermodal registration the same features may not be visible in images acquired of the same anatomy by different scanning modalities.

A simple example of registration with pathology is the case of intrasubject serial MR brain-image registration where a new lesion has developed in the interval between scans. It is relatively straightforward to segment the brains in both images and register them with a rigid-body transformation. Naively applying the resultant transformation to the lesion infers correspondence between the lesion and a volume of normal tissue in the earlier scan. This is a correspondence between different tissues at the same geometrical location and is valid for a lesion-detection task where we are looking for change. The lesion may have replaced normal tissue in which case the rigid-body transformation is appropriate even though there is not true biological correspondence. Alternatively the lesion may be space occupying and has displaced surrounding normal tissue. If the application is surgical resection then NRR that includes a model of tissue growth must be applied if tissue correspondence is to be established. This would be vital to help guide the surgeon away from displaced healthy tissue. Therefore the selection criterion for the transformation model can depend both on the data and on the application context.

Where a specific physical transformation model is not available, generic assumptions of "smoothness," "localisation," "invertibility," or "compactness of description" can be used to select transformation types which do not use explicit knowledge of the particular registration task but have performed well on a variety of image matching problems. Alternatively physical analogies can be used to select transformation types that approximate the known behaviour of tissue, e.g., elastic, viscous fluid. For intrasubject registration perhaps the most principled approaches use biomechanical models of the brain tissue and surrounding structures, often for applications in image-guided neurosurgery. Provided the mechanics can be modeled well enough, these

[4] Discernible groups or patterns of arrangement of cells.

should select only those transformations that are physically feasible (Hagemann et al., 2003). For intersubject registration it is unprincipled to use a biomechanical model since morphological differences between corresponding anatomy are not the result of a biomechanical process. In these cases statistical models of the population variation in brain structure can be used to select the transformation (Ashburner et al., 1997; Cootes et al., 1999; Nikou et al., 2001). However suitable correspondences must be found to construct the models in the first place and these correspondences must be found efficiently and robustly (Davies et al., 2001; Rueckert et al., 2001). There is a danger of circularity here.

Registration, correspondence, and homology

The idea of homology originated with Richard Owen (1843) who famously suggested that a homologue is "The same organ in different animals under every variety of form and function." Homology is a vital concept for those seeking to understand evolutionary processes across species. With the advent of genomics and proteomics, structural and genetic homology will assume greater importance as the relationships between animal models and humans are probed more deeply than was previously possible. The original definition of Owen has given way to at least three more detailed definitions (Wagner, 1994) (1) phylogenetic, where the existence of synapomorphy[5] is the main criterion; (2) morphological, where spatial configuration and structural similarity indicate homology; and (3) developmental, where adult genetic pathways ("morphostatic") are distinguished from those of early development ("morphogenetic"). The homologies of interest in medical image registration are largely morphological and functional; however, synapomorphy and developmental homology are important in registration of the cerebral cortex. An accessible overview of homology in this context is given in Maudgil et al. (1998). A rather more philosophical account of the relationship between homology and its various definitions and correspondence is given by Brigandt (2002).

Recently, there has been a temptation to interpret patterns in regulatory gene expression repeated across species as being enough to assert that the structures governed by such genes are homologous. This has necessitated a reevaluation of the meaning of homology among geneticists (Abouheif et al., 1997) and a quite general definition of homology has been suggested by Fitch (2000) as "the relationship between two characters that have descended, usually with divergence from a common ancestral character." Here "character" is defined as "any genic, structural or behavioural feature of an organism." One of the 15 problems described in this reference is the "structure/function problem" where it is pointed out that across many organisms, the same structure can have different functions or different structures can have the same function. This is a generalisation of the problem of structural versus functional correspondence in human brain studies. At present, there are immense problems reconciling these correspondence issues in humans and it is likely to be even more difficult to draw inferences from animal studies by invoking arguments which rely on homology. Novel analytical methods may help to solve these problems. One recent approach (Van Essen et al., 1998) used constrained surface-based warpings of the cortical sheet to preserve topological and spatial relationships between functionally distinct areas in the cerebral cortex rather than striving inappropriately for exact correspondence.

For exploratory studies of the human genome to be carried out using animal models, e.g., with so-called knockout mice (Kooy et al., 1999, 2001), it will be necessary to resolve correspondence problems both at a genetic and a morphological level, not least, as many investigations will be image-based (Balaban and Hampshire 2001; Jacobs and Cherry, 2001; Johnson et al., 2002). Under strictly controlled imaging conditions it may be possible to bypass the structure/function correspondence problem in individual animals (or humans) by acquiring images simultaneously from more than one modality (Marsden et al., 2002; Garlick, 2002), although this still will not address the wider problems of intersubject correspondence. The influence of genes on human brain structure has already been probed directly by imaging matched monozygotic and dizygotic twins to correlate regional grey-matter density with genetic affinity (Thompson et al., 2001). In this study, monozygotic twins had virtually identical grey-matter structure, indicating that their development had been relatively unaffected by environmental effects. Dizygotic twins had significantly more similar sensorimotor and occipital-parietal structure than random pairs that the authors conclude is indicative of a continuum of structural differences caused by genetic differences. Another study of mono and dizygotic twins used statistical shape analysis to encode central sulcal variation (Le Goualher et al., 2000) and showed that sulcal shape varied much less between monozygotic twins than dizygotic and that genetic effects were stronger in the left hemisphere than in the right hemisphere. A third approach was taken by Good et al. (2001a) where voxel-based morphometry was used to correlate the degree of X chromosome deletion with differences in regional grey- and white-matter density in women suffering Turner's syndrome. Significant regional differences in grey-matter density were observed in 15 subjects with the full syndrome and a spectrum of regional differences correlating with cognitive-behavioural assessments were observed in 11 subjects suffering partial deletions. While impressive, the interpretation of studies relying on interindividual spatial comparisons requires care. The sophistication and prevalence of such studies is likely to increase as gene-mapping becomes more routine. The chal-

[5] Inheritance of shared derived characteristics. If members of a group have an evolutionary history in common then they share unique features which are not present in distant ancestors.

lenges of linking anatomy to genetics should not be underestimated, as simple experimental paradigms, where large number of subjects can be split into well-defined groups by a small number of genetic and structural factors, are unlikely to be common.

Validation

Validation tests the ability of registration to establish correspondence. Early in the development cycle of a nonrigid registration technique, nonrepeatable validation methods including special or multiple image acquisition, the use of implanted markers as "gold-standard" correspondences (West et al., 1997), or biomechanical models which produce known anatomical variation or deformation (Schnabel et al., 2003) have been used. Markers are complementary to biomechanical models in that they provide a sparse set of true gold-standard in vivo correspondences whereas biomechanical models provide a dense-set of simulated correspondences.[6] These techniques provide information which characterises a technique in a particular setting but cannot usually be applied retrospectively or to estimate registration error in individual cases. During testing, or sporadically during the course of a study, postregistration validation by expert visual inspection, labeling of corresponding landmarks or regions, and deformation field analysis are used to increase confidence.

A recent comprehensive approach to comparing NRR techniques used 128 landmarks identified in each hemisphere of each subject by an experienced neuroanatomist (Grachev et al., 1999). The landmarks were well distributed and chosen to provide a good summary of the individual topography of the brain including the neocortex. These techniques have wide applicability for characterising registration error but must be applied on a case-by-case basis and require expert anatomical knowledge and significant operator time.

Registration validation has been studied in detail by Fitzpatrick et al, (1998, 2001) for the case of rigid-body point-based registration leading to expressions for the expected squared target registration error[7] (TRE) and an approximation to the distribution of the TRE. The TRE was recently estimated for a nonrigid registration application by Schnabel et al. (2003), using a deforming biomechanical model. In real registration tasks, gauging success relies on postregistration point or feature identification as it is not generally possible to identify a dense set of point correspondences from which to calculate the TRE as in simulated data. Regional correspondence can, however, be estimated by considering a set of M volumes of interest labeled independently in each of a pair of registered images. Then if V_i and V_i' are the set of voxels comprising the volume of interest in each image after registration then the fractional overlap of the i^{th} volume $O_i(\tau)$, where τ is the correspondence tolerance, is given by

$$O_i(\tau) = \frac{(V_i' \cap V_i)(\tau)}{V_i' \cup V_i}.$$

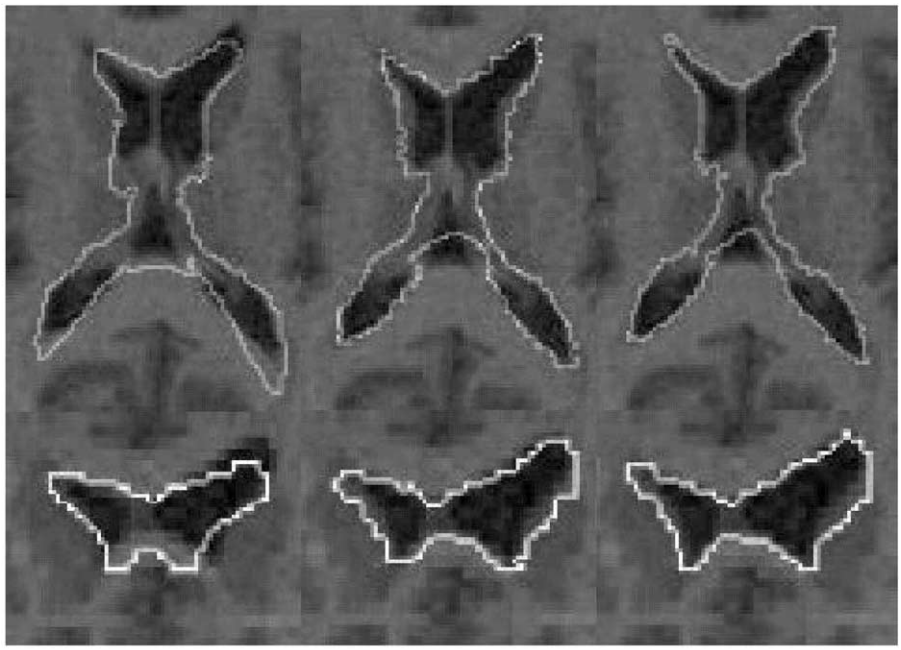
When $\tau = 0$ this measure reduces to the commonly used region of interest overlap (Andreasen et al., 1996; Crum et al., 2001). For $\tau > 0$ voxels are considered overlapping if they lie within τ voxels of each other. Then the registration *fidelity* can be defined over all such regions of interest as

$$F_R(\tau) = \frac{1}{M} \sum_{i=1}^{i=M} O_i(\tau).$$

The fidelity allows anatomically specific measures of correspondence error to be monitored during and post registration to a specified tolerance. It allows upper bounds on regional correspondence error to be established provided labeling errors can be measured independently. To illustrate its use, in Fig. 1, a subject's ventricles were labeled automatically by nonrigid registration to a presegmented atlas image (see Holden et al., 2002, for a full account of this work). The automatic labeling is compared with hand segmentation for three nonrigid registrations (affine, nonrigid registration using a B-spline transformation model with 10.0-mm spaced control points and the same model with 5.0-mm spaced control points). In Fig. 2 the regional registration fidelity $F_R(\tau)$ is reported for a range of tolerances for the three attempts at registration. As expected, the affine registration performs relatively poorly compared with the other nonrigid registrations. What would not be apparent by reporting fidelity for $\tau = 0$ alone is that the overlap for the coarse and fine registrations rises from ≈ 0.76 to ≈ 0.95 for $\tau = 1$ and to ≈ 0.99 for $\tau = 2$. Therefore, in this example at least, the regional registration error has an upper bound of about 2 voxels. Although this error necessarily includes contributions from the original segmentations it is informative about the closeness of overlap as well as the degree of overlap. The fidelity does not provide information about what happens within the labeled regions but can easily be measured for subregions if such labeling is possible. Alternatively if internal landmarks can be identified, standard distance metrics can be computed, or if detailed information about changes in shape is required then parametric shape models of the labeled regions can be constructed. Where hypotheses about structurally or functionally homologous regions of the brain are tested using automated voxel-based NRR it makes sense to systematically verify the veracity of the NRR at least around the regions under investigation

[6] Many of these correspondences may be in homogenous lumps of tissue. There is an interesting philosophical question about whether such correspondences are legitimate for validation as they can rarely be fully corroborated in clinical data.

[7] The postregistration distance between corresponding points which were not used for registration.



(a) affine (b) 10.0mm (c) 5.0mm

Fig. 1. Ventricular labels propagated from an atlas onto an individual brain using nonrigid registration (a) affine, (b) B spline with 10.0-mm spacing, and (c) B spline with 5.0-mm spacing. From Holden et al. (2002).

using at least one such approach. This is almost never done in clinical studies.

Surface geometrical approaches offer alternative validation strategies that potentially provide detailed information about the spatial variation of registration error. The root-mean-square registration error on surfaces registered using different transformation models can be mapped into the target space and used to assess residual variance at the registered boundaries (Thompson et al., 2000a). In principle the sensitivity of NRR to the detail of the voxel labeling can be modeled to assess local improvements in boundary matching; where surface information is available we advocate exploiting it in this way. The overlap of classified tissues, correlation of a differential operator sensitive to changes in gyral-sulcal geometry, and shape and overlap of 12 major sulci have been used to compare four NRR techniques by Hellier et al. (2001). This approach is noteworthy for striving to measure the ability of NRR to match shape; however, it remains unclear how feasible it is to use such detailed measures in large studies and how they could be incorporated into an error-analysis framework for studies relying on NRR.

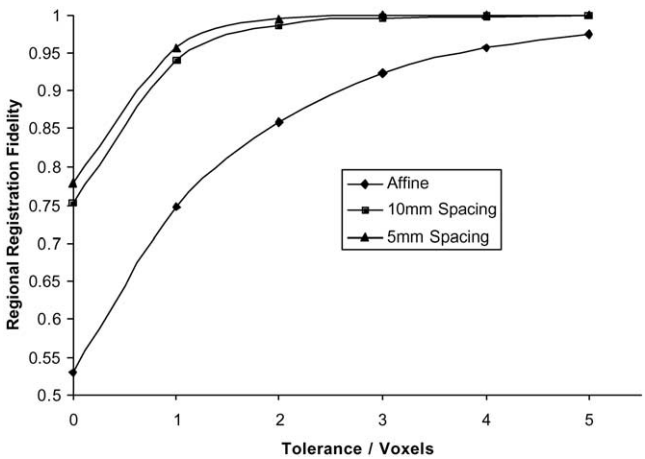


Fig. 2. Nonrigid registration fidelity as a function of registration tolerance evaluated by comparing propagated and manual ventricular labelings.

A related missing link in validation is the gap between image similarity measures and registration fidelity measures. The image similarity measure is a nontrivial function of the correspondence error, image noise, and the transformation model and is (1) uninformative about whether the registration is likely to be "correct," (2) uninformative about the magnitude of errors of correspondence, and (3) uninformative about the spatial distribution of errors of correspondence. Knowledge of image statistics and the likely variation in the appearance and morphology of imaged structures of interest might in the future be used to build similarity measures which provide a contextual basis for registration and allow monitoring as well as post hoc analysis. Such measures would inform us about the a priori likelihood of success, help in avoiding local minima in the parameter space, and guide the choice of registration strategy and

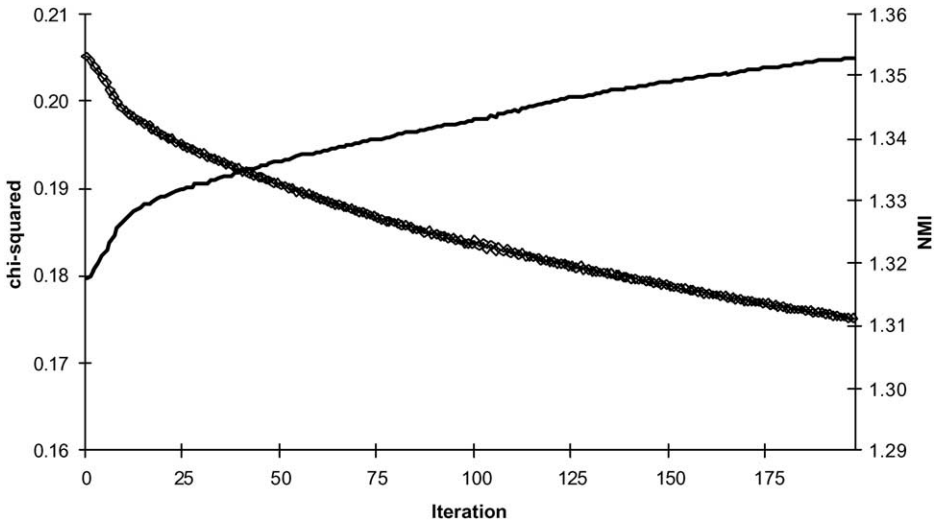


Fig. 3. Using prior knowledge and a χ^2 similarity measure to drive iterative nonrigid registration. The Normalised Mutual Information was computed but not used to drive the registration and is shown for comparison.

would mark a major step forward. When there is prior knowledge available, similarity measures can be constructed which measure the departure of imperfectly registered images from the ideal state. In the case of fluid registration of serial MR, for instance, to measure individual change due to atrophy or adult-onset growth hormone deficiency, it is reasonable to assume that the source image will closely resemble the target image postregistration and therefore the postregistration joint intensity histogram can be estimated and used to drive the registration. Assuming Poisson statistics for error in the intensity histogram, a simple χ^2 similarity measure based on knowledge of the postregistration joint intensity histogram is

$$\chi^2 = \sum_i \sum_j \chi^2_{i,j} = \sum_i \sum_j \frac{(n_{i,j} - f_{i,j})^2}{f_{i,j}} .$$

Here the summation is over the contents n_{ij} of all bins in the joint intensity histogram and $f_{i,j}$ are the estimated contents postregistration. Fig. 3 shows χ^2, together with the normalised mutual information computed concurrently, for an example fluid registration of serial MR brain images. The advantage of formulating the similarity measure in terms of χ^2 is that it then becomes possible to estimate the covariance matrix, C_θ, on the set of transformation[8] parameters θ via

$$C_\theta^{-1} = \sum_i \sum_j (\nabla_\theta \chi_{i,j})^T \otimes (\nabla_\theta \chi_{i,j}) .$$

With such information available, the links between spatial correspondence uncertainty, transformation model parameters, and similarity measure uncertainty can be made explicit. More informative similarity measures and a commitment to true validation on a case-by-case basis rather than generic characterisation of techniques would add considerable value to studies relying on NRR and is urgently required.

[8] An appropriate choice of transformation model is vital for this to yield useful information.

Voxel-based morphometry

Recently the debate on registration and correspondence formed the basis of papers by Bookstein (2001) and Ashburner and Friston (2001) in which the validity of the voxel-based morphometry (VBM) approach to structural brain analysis in large cohorts was debated. Much of the argument revolved around the degree to which correspondence could be achieved between brains of different subjects and to what extent errors of correspondence could affect a statistical analysis of group-wise grey-matter density differences.

In simple terms, VBM looks for significant differences between brain images of two cohorts under some hypothesis of clinical, behavioural, or genetic condition that separates them. An automated image registration and segmentation procedure is used to transform the images into a standard space by nonrigid registration to a standard anatomical template — an example of spatial normalisation. Smoothing of the normalised images selects a spatial scale at which abnormalities will be sought. Statistical tests are carried out on a voxel-by-voxel basis within this space to look for significant differences in grey or white-matter density between the cohorts. The process of spatial normalisation is actually quite involved and is described in detail by Good et al. (2001b) and Ashburner and Friston (2000). As a one-to-one correspondence between brains is not generally present, a relatively coarse nonrigid registration is used for the

spatial normalisation and fine structure is blurred by the smoothing step.

Most publications referring to VBM without further qualification are applying the SPM implementation of VBM. However VBM is a more general methodology for structural analysis that uses some appropriate NRR for spatial normalisation. There is no reason why the particular NRR incorporated in the SPM package needs to be the one used for spatial normalisation and nowhere are we criticising particular implementation details beyond the fact that the spatial distribution of registration error is unknown, the techniques do not discern whether registration has "succeeded" or "failed," and the registration error can only be estimated in subsequent statistical analysis from the residual voxel intensity variation. These criticisms are appropriate to all voxel-level NRR techniques where other contextual information is not used to drive the registration or to check the correspondence postregistration. These are precisely the techniques that are being applied in large imaging studies of normal and abnormal neuroanatomy.

The main criticism of VBM in the recent debate was the presumption that a continuum of analysis exists which is parameterised by the degree of nonrigid registration used for spatial normalisation. At one extreme purely affine registration would lead to poor localisation of significant differences. At the other extreme—unattainable in practice for registration of one human brain to another — each brain in the cohort will be precisely normalised to the anatomical standard and appear identical. In this case information about the structural differences in the cohorts is encoded in the nonrigid transformations rather than the voxel intensities. VBM analysis is generally performed somewhere between these extremes; a certain amount of nonrigid registration is performed to coarsely align the brains and statistical tests are applied to the smoothed residual intensity differences. The number of basis functions chosen for the NRR (typically $7 \times 8 \times 7$ in the x, y, and z directions, respectively) specifies where in the continuum analysis will be performed. The other important parameter choice is the size of the smoothing kernel that is specified to select scales of interest post spatial normalisation; this is frequently set in the range 8–12 mm. There is clearly an interaction between the parameter choices for the NRR and the smoothing and the scales of anatomy and differences in anatomy both within and between groups.

In more recent variants of VBM, the volume-change at each voxel during spatial normalisation is used to modulate its intensity so that the intensity integral across any structure before and after spatial normalisation is preserved.[9] This means that there is potentially a continuum with regard to volume change[10] and that modulation may reduce the dependence of VBM on the number of basis functions used for spatial normalisation. In our opinion an opportunity is missed by Ashburner and Friston (2001) to demonstrate that for a reasonable range of variation in the degree of nonrigid registration, VBM results in the consistent detection and localisation of structural differences. In fact they make a surprising generalisation by suggesting that "Reporting that registration over intersubject variation was done with SPM99 using the default settings is a precise description of how homologies are identified." This is only true to the extent that running the program twice on the same data will result in the same set of correspondences being defined. It does state precisely how to *repeat the identification of such homologies* but it makes no statement — and it cannot, even knowing the detail of the registration algorithm — about *how* the homologies themselves are defined or about the distribution of errors of correspondence. The homology they refer to can only be between voxel intensity profiles in the target image and those in the standard space; this is a much weaker definition of homology than that of true neuroanatomical correspondence.

Furthermore, the VBM user has no way of predicting or assessing the validity of such a set of correspondences and the errors of correspondence are unlikely to be consistent in spatial location or magnitude across the subjects comprising a cohort. All that can be hoped for is that for large enough cohort numbers and appropriately chosen NRR and smoothing parameters, randomly distributed registration errors of a suitably small scale are removed by image smoothing. On the other hand it is possible that significant structural differences between cohorts, which are usually the reason for performing VBM in the first place, will bias the distribution of registration errors in both magnitude and location making interpretation of results difficult. The reason VBM has been comparatively successful is probably due to a mutually beneficial relationship between the degree (i.e., the scale) of the NRR, the scale of the applied blurring kernel, and the spatial extent of structures of interest in the brain.

Some recent work has examined the precision of spatial normalisation in the medial temporal lobe by identifying a series of landmarks across a VBM cohort (Salmond et al., 2002). Here it was found that allowing too much spatial normalisation (i.e., endowing the NRR with a high number of degrees of freedom) resulted in poorer landmark colocalization. This is because improved intensity matching was at the expense of poorer anatomical correspondence. From this limited study it is not obvious how to select the optimal number of degrees of freedom a priori except by doing exactly the kind of manual labeling VBM is supposed to avoid; neither is it obvious that the optimal degrees of freedom for structures of a particular scale at a particular location in the brain will be optimal for other structures of different scales or in different locales. In some other prac-

[9] A conceptually similar step is also applied during spatial normalisation in the RAVENS approach of Davatzikos et al. (2001).

[10] As a voxel of "stuff" is more compressed it gets brighter, as it is more expanded it gets darker so the integral of stuff across the voxel remains constant

tical cases where spatial normalisation is used the resolution of functional maps (e.g., PET with FWHM ~10 mm) may be the accuracy-limiting factor rather than the exact form of the spatial normalisation (Crivello et al., 2002).

Many of the same criticisms can be directed at deformation-based morphometry (DBM) where spatial normalisation is performed as above but subsequent analysis is performed directly on the resulting deformation fields rather than on transformed images (e.g., Thompson et al., 1997; Ashburner et al., 1998). This potentially opens up many more possibilities for examining structural morphology and cortical geometry compared with simply looking at changes in local tissue density. There are so many variants of DBM that it is unfair to criticise them in a broad-brush way. However, it is evident that DBM techniques that use deformation fields derived from voxel-intensity based registration suffer from the same problems of validation and error management as VBM techniques. In fact they may prove more susceptible to registration problems as measures computed from the local geometry of the deformation fields may be far more sensitive to the detail of those fields than volumetric measures of tissue density. Some attempts to validate DBM have been made by comparing the mean Jacobian determinant computed over a ventricular region with volume change computed from manual delineation of the same region (Gaser et al., 2001). Deformation fields were determined at a resolution "two times coarser" than the voxel size of 1 mm^3, converted to a field of displacement vectors, one per voxel, and then smoothed with a Gaussian of FWHM 8 mm. Therefore as in VBM, the information encoded at a point represents a weighted average from a locale obtained by smoothing the normalised images. While interesting, it remains to be seen whether similar approaches to validation are useful for more sophisticated applications of DBM. Ultimately the spatial resolution of deformation fields in DBM and VBM should be determined by the geometry of the local neuroanatomy so that local estimates of the limits of achievable correspondence can be obtained. Recent work by Shen and Davatzikos (2003) computes a geometric attribute vector for each voxel that is matched during registration. Such approaches summarise the structural context of each voxel and are a promising approach to more robust registration and spatial normalisation.

Conclusions

Nonrigid registration is being used as a tool for imaging neuroscience despite lacking the necessary framework to explicitly estimate and localise error. As studies become more sophisticated it is becoming increasingly important to understand and measure the degree, regional variation, and confidence in the correspondences established by registration. The kind of correspondence, the manner of achieving it, and the acceptable accuracy are application dependent; the images define the kinds of correspondences that could in principle be sought but the scientific question defines the kinds of correspondence that *should* be sought as in our example of lesion detection versus resection earlier. Specifying registration goals more explicitly enables the algorithms to be tailored to specific applications, performance characterised early in development, and success objectively measured. There is a void in current registration technology between sophisticated but well-calibrated techniques which require prior knowledge of surfaces, volumes, and landmarks, and automated techniques which work purely with voxel intensities but are almost impossible to validate. It is vital that convergence occurs to provide the neuroimaging community with registration tools that can monitor their own performance and estimate correspondence error with minimal intervention. A simple analogy of the current state of registration is fitting a curve to a set of data points, perhaps the most common task in scientific data analysis. The naïve student might look at the distribution of points and make a guess as to the functional form of the curve that should be fitted (Fig. 4). If the points lie "close" to the fitted curve the result is judged to be "good." This is akin to the state of registration at present. With the addition of error bars to the points a more informed choice guided by a priori knowledge of the likely errors and an appropriate fitting function can be made and the likely error in the fit can be estimated too (Fig. 5).

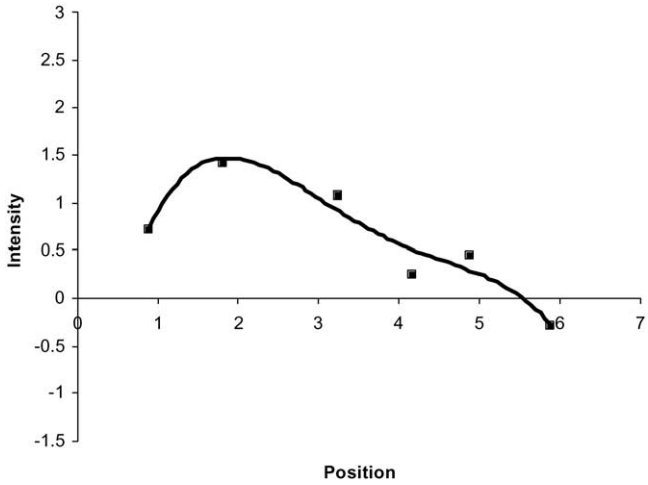


Fig. 4. Analogy of current registration practice—fit a likely looking curve to a set of points.

Much of the criticism in this paper has been aimed at VBM as one of the most widely used applications of automated NRR. Proponents of VBM often emphasise that a prior hypothesis about the location of structural differences is not required as a regional analysis is performed without bias over the entire brain. Therefore VBM is in principle less labour-intensive than traditional delineation of regions of interest followed by volumetric analysis as no manual intervention is required during processing. However these automated approaches are only of value if their performance can be evaluated, preferably in a way that allows correspon-

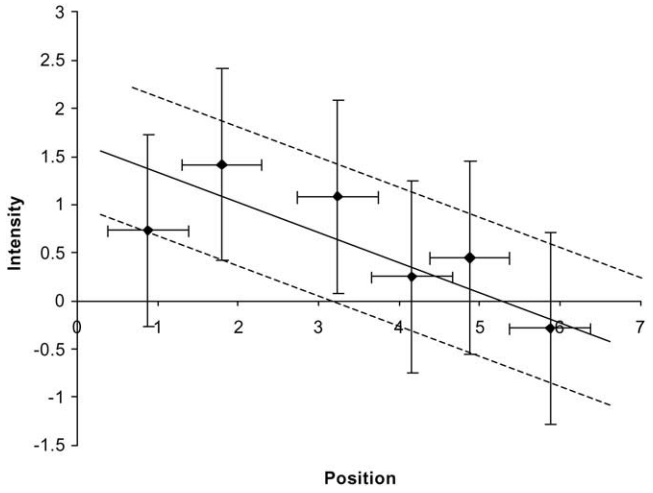


Fig. 5. Analogy of the desired situation. There are estimated errors in both the position and intensity at each point. A properly fitted function to these data requires an explicit hypothesis about its functional form and has confidence intervals describing the likely error. In cohort studies, such an analysis may identify the presence of confounding group effects.

dence error to be propagated through subsequent statistical analysis. This is currently not the case and in a recent publication, Good et al. (2002) accept the shortcomings of techniques that rely on automated registration and segmentation techniques by stating that "with automated techniques such as VBM, visual inspection should be performed at each stage of image preprocessing in order to improve interpretation". To understand this comment, remember that the core of VBM is simply a mechanism for detecting differences between groups of voxels. The membership of these groups is determined by the segmentation and spatial normalisation preprocessing steps. Therefore when VBM detects a significant difference, identifying the cause requires a careful consideration of the underlying pathophysiology and how this interacts with segmentation and NRR. Visual inspection is then vital to allow the investigator to accept or reject the inference of local structural difference in the context of the experimental hypothesis. In most instances, where the anatomical differences are of a small spatial scale, significant differences are caused by cortical thinning and related structural or microstructural changes. However, macroscopic abnormalities can interact with the spatial normalisation procedure to localise cortical areas in different regions. This may produce significant inferences; however, the interpretation will be fundamentally different and rely upon visual inspection. Other recent work comparing VBM with manual volumetry concludes that "although semiautomated and voxel-based methods can provide a reasonable estimate of regional brain volume, they cannot serve as a substitute for manual volumetry" (Tisserand et al., 2002). VBM is undoubtedly a powerful framework that successfully removes the need for expert labour-intensive segmentation but currently replaces it with a complicated problem of interpretation and validation which significantly reduces its efficacy. The validity of current published studies relying on NRR in this way is in most cases limited and in some cases suspect due to indiscriminate application of these poorly understood techniques. Until the technology can be made demonstrably more robust one possible solution is to accept that VBM should only be used as a prompting system to highlight regions of the brain worthy of further analysis by more manually intensive techniques. This approach combines the hypothesis-free advantages of VBM with the benefits of expert manual intervention but is by no means an ideal solution.

To properly quantify correspondence error will ultimately require models of image formation, biomechanical and statistical models of deformation and statistical models of anatomical variation. These models could in principle be combined to determine the range of anatomical variation, image quality, and deformation over which a registration algorithm produces point or regional correspondences to within known tolerances. For new registration problems, the quality and content of the input data and the resulting correspondences could be checked to ensure they lie within the validated parameter range of the algorithm.

Medical image analysis has often been inspired by advances in other fields and is certain to benefit from future developments in areas like computer vision. However, with increasing specialisation, medical image analysis is now established as a field in its own right. Some of the advances required to solve the problems outlined in this paper may only appear from within the medical image analysis and computational anatomy communities because medical image registration has additional specific challenges compared with many mainstream computer vision applications. The wide variability in image characteristics and quality combined with a practical need for high reliability and robustness make medical image registration difficult and demanding. Therefore these problems need to be addressed now rather than waiting for solutions to permeate through from other fields.

The message of this article can be summed up in one word—quality. At present we do not usually know the quality of nonrigid registration in terms of how good it is. Just as importantly we usually do not know its quality in terms of its character, i.e., the magnitude and spatial distribution of registration errors and their relationship to the correspondence of imaged anatomy. Quality should be present at all stages of a nonrigid registration task, that is, from prescribing a success criterion, quantifying the technical quality of the images, the quality of the relationship between the underlying biology and the imaged features, and the quality of the measured errors of the result. Properly understanding the influence of quality will maximise the construct validity of the next generation of nonrigid registration algorithms and may increase their face validity. In the words of Pirsig's character Phaedrus, "The sun of quality does not revolve around the subjects and objects of our existence. It does not just passively illuminate them. It is not

subordinate to them in any way. It has created them. They are subordinate to it!"

Acknowledgments

Bill Crum is a Senior Research Fellow funded by the EPSRC-MRC Medical Images and Signals IRC. The authors thank Dr. Mark Holden for image data and discussion relating to Figs. 1 and 2 and Dr. Neil Thacker for pointing out the relationship between the inverse covariance matrix and χ^2. Three reviewers each provided detailed and insightful criticisms that have immeasurably improved this paper.

References

Abouheif, E., Akam, M., Dickinson, W.J., Holland, P.W.H., Meyer, A., Patel, N.H., Raff, R.A., Roth, V.L., Wray, G.A., 1997. Homology and developmental genes. Trends Genet. 13, 432–433.

Andreasen, N.C., Rajarethinam, R., Cizadlo, T., Arndt, S., Swayze, V.W., Flashman, L.A., O'Leary, D.S., Ehrhardt, J.C., Yuh, W.T.C., 1996. Automatic atlas-based volume estimation of human brain regions from MR images. J. Comput. Assist. Tomogr. 20, 98–106.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—the methods. NeuroImage 11, 805–821.

Ashburner, J., Friston, K.J., 2001. Why voxel-based morphometry should be used. NeuroImage 14, 1238–1243.

Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K., 1998. Identifying global anatomical differences: deformation-based morphometry. Hum. Brain Mapp. 6, 348–357.

Ashburner, J., Neelin, P., Collins, D.L., Evans, A., Friston, K., 1997. Incorporating prior knowledge into image registration. NeuroImage 6, 344–352.

Balaban, R.S., Hampshire, V.A., 2001. Challenges in small animal noninvasive imaging. Inst. Lab. Anim. Res. J. 42, 248–263.

Bjaalie, J.G., 2002. Localization in the brain: new solutions emerging. Nature Rev. Neurosci. 3, 322–325.

Bookstein, F.L., 2001. "Voxel-based morphometry" should not be used with imperfectly registered images. NeuroImage 14, 1454–1462.

Brett, M., Johnsrude, I.S., Owen, A.M., 2002. The problem of functional localization in the human brain. Nature Rev. Neurosci. 3, 243–249.

Brigandt, I., 2002. Homology and the origin of correspondence. Biol. Philos. 17, 389–407.

Christensen, G.E., Rabbitt, R.D., Miller, M.I., 1996. Deformable templates using large deformation kinematics. IEEE Trans. Image Process. 5, 1435–1447.

Collignon, A., 1998. Multi-modality medical image registration by maximization of mutual information. Ph.D. thesis. Catholic University of Leuven, Leuven, Belgium.

Cootes, T.F., Beeston, C., Edwards, G.J., Taylor, C.J., 1999. A unified framework for atlas matching using active appearance models. IPMI 1999, Proceedings. Lecture Notes in Computer Science 1613, 322–333.

Crivello, F., Schormann, T., Roland, P.E., Zilles, K., Mazoyer, B.M., 2002. Comparison of spatial normalisation procedures and their impact on functional maps. Hum. Brain Mapp. 16, 228–250.

Crum, W.R., Scahill, R.I., Fox, N.C., 2001. Automated hippocampal segmentation by regional fluid registration of serial MRI: validation and application in Alzheimer's disease. NeuroImage 13, 847–855.

Davatzikos, C., 2001. Measuring biological shape using geometry-based shape transformations. Image Vision Comput. 19, 63–74.

Davatzikos, C., Genc, A., Xu, D., Resnick, S.M., 2001. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. NeuroImage 14, 1361–1369.

Davies, R.H., Cootes, T.F., Taylor, C.J., A Minimum description length approach to statistical shape modelling. IPMI 2001, Proceedings, Lecture Notes in Computer Science, 2082, 50–63.

Fitch, W.M., 2000. Homology: a personal view on some of the problems. Trends Genet. 16, 227–231.

Fitzpatrick, J.M., West, J.B., 2001. The distribution of target registration error in rigid-body point-based registration. IEEE Trans. Med. Imaging 20, 917–927.

Fitzpatrick, J.M., West, J., Maurer Jr., C., 1998. Predicting error in rigid-body, point-based registration. IEEE Trans. Med. Imag. 17, 694–702.

Freeborough, P.A., Fox, N.C., 1997. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. IEEE Trans. Med. Imag. 16, 623–629.

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

Garlick, P.B., 2002. Simultaneous PET and NMR—initial results from isolated perfused rat hearts. Br. J. Radiol. 75, S60–S66.

Gaser, C., Nenadic, I., Buchsbaum, B.R., Hazlett, E.A., Buchsbaum, M.S., 2001. Deformation-based morphometry and its relation to conventional volumetry of brain lateral ventricles in MRI. NeuroImage 13, 1140–1145.

Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N.A., Friston, K.J., Frackowiak, R.S.J., 2001b. A voxel-based mophometric study of ageing in 465 normal adult human brains. NeuroImage 14, 21–36.

Good, C.D., Kuntsi, J., Akers, R., Elgar, K., Price, C., Ashburner, J., Friston, K.J., Frackowiack, R.S.J., Skuse, D.H., 2001a. Gene deletion mapping of the X chromosome. NeuroImage 13, S793 (Proceedings of the 7th Annual Meeting of the Organization for Human Brain Mapping).

Good, C.D., Scahill, R.I., Fox, N.C., Ashburner, J., Friston, K.J., Chan, D., Crum, W.R., Rossor, M.N., Frakowiak, R.S.J., 2002. Automatic differentiation of anatomical patterns in the human brain: validation with studies of degenerative dementias. NeuroImage 17, 29–46.

Grachev, I.D., Berdichevsky, D., Rauch, S.L., Heckers, S., Kennedy, D.N., Caviness, V.S., Alpert, N.M., 1999. A method for assessing the accuracy of intersubject registration of the human brain using anatomic landmarks. NeuroImage 9, 250–268.

Hagemann, A., Rohr, K., Stiehl, H.S., 2003. Coupling of fluid and elastic models for biomechanical simulations of brain deformations using FEM. Med. Image Anal. 6, 375–388.

Hellier, P., Barillot, C., Corouge, I., Gibaud, B., Le Goualher, G., Collins, L., Evans, A., Malandain, G., Ayache, N., 2001. Retrospective evaluation of inter-subject brain registration. MICCAI 2001, Proceedings. Lecture Notes in Computer Science 2208, 258–265.

Hill, D.L.G., Batchelor, P.G., Holden, M., Hawkes, D.J., 2001. Medical image registration. Phys. Med. Biol. 46, R1–R45.

Holden, M., Schnabel, J.A., Hill, D.L.G., 2002. Quantification of small cerebral ventricular volume changes in treated growth hormone patients using non-rigid registration. IEEE Trans. Med. Imaging 21, 1292–1301.

Jacobs, R.E., Cherry, S.R., 2001. Complementary emerging techniques: high-resolution PET and MRI. Curr. Opin. Neurobiol. 11, 621–629.

Johnson, G.A., Cofer, G.P., Fubrara, B., Gewalt, S.L., Hedlund, L.W., Maronpot, R.R., 2002. Magnetic resonance histology for morphologic phenotyping. J. Magn. Reson. Imaging 16, 423–429.

Kooy, R.F., Reyniers, E., Verhoye, M., Sijbers, J., Bakker, C.E., Willems, P.J., Van Der Linden, A., 1999. Neuroanatomy of the fragile X knockout mouse brain studied using in vivo high resolution magnetic resonance imaging. Eur. J. Hum. Genet. 7, 526–532.

Kooy, R.F., Verhoye, M., Lemmon, V., Van Der Linden, A., 2001. Brain studies of mouse models for neurogenetic disorders using in vivo magnetic resonance imaging. Eur. J. Hum. Genet. 9, 153–159.

Le Goualher, G., Argenti, A.M., Duyme, M., Baare, W.F.C., Hulshoff Pol, H.E., Boomsma, D.I., Zouaoui, A., Barillot, C., Evans, A.C., 2000.

Statistical sulcal shape comparisons: application to the detection of genetic encoding of the central sulcus shape. NeuroImage 11, 564–574.
Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multi-modality image registration by maximization of mutual information. IEEE Trans. Med. Imaging 16, 187–198.
Marsden, P.K., Strul, D., Keevil, S.F., Williams, S.C.R., Cash, D., 2002. Simultaneous PET and NMR. Br. J. Radiol. 75, S53–S59.
Maudgil, D.D., Free, S.L., Sisodiya, S.M., Lemieux, L., Woermann, F.G., Fish, D.R., Shorvon, S.D., 1998. Identifying homologous anatomical landmarks on reconstructed magnetic resonance images of the human cerebral cortical surface. J. Anat. 193, 559–571.
Nikou, C., Bueno, G., Heitz, F., Armspach, J.P., 2001. A joint physics-based statistical deformable model for multimodal brain image analysis. IEEE Trans. Med. Imaging 20, 1026–1037.
Owen R. 1843. Lectures on the Comparative Anatomy and Physiology of the Vertebrate Animals (Delivered at the Royal College of Surgeons, London).
Pirsig, R., Zen and the Art of Motorcycle Maintenance: An Enquiry into Values. Paperback Reprint, Perennial (HarperCollins); ISBN: 0060958324.
Roche, A., Malandain, G., Pennec, X., Ayache, N., 1998. The correlation ratio as a new similarity measure for multimodal image registration. MICCAI'98, Proceedings. Lecture Notes in Computer Science 1496, 1115–1124.
Roland, P.E., Geyer, S., Amunts, K., Schormann, T., Schleicher, A., Malikovic, A., Zilles, K., 1997. Cytoarchitectural maps of the human brain in standard anatomical space. Hum. Brain Mapp. 5, 222–227.
Rueckert, D., Frangi, A.F., Schnabel, J.A., 2001. Automatic construction of 3D statistical deformation models using non-rigid registration, MICCAI'99. Proceedings. Lecture Notes in Computer Science 1679, 77–84.
Salmond, C.H., Ashburner, J.A., Vargha-Khadem, F., Connelly, A., Gadian, D.G., Friston, K.J., 2002. The precision of anatomical normalization in the medial temporal lobe using spatial basis functions. NeuroImage 17, 507–512.
Schnabel, J.A., Tanner, C., Castellano-Smith, A.D., Degenhard, A., Leach, M.O., Hose, D.R., Hill, D.L.G., Hawkes, D.J. 2003, Validation of non-rigid registration using finite element methods: application to breast MR images. IEEE Trans. Med. Imaging, to appear.
Shen, D., Davatzikos, C., 2003. Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration. NeuroImage 18, 28–41.
Studholme, C., Hill, D.L.G., Hawkes, D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recogn. 32, 71–86.
Thompson, P., Cannon, T.D., Narr, K.L., van Erp, T., Poutanen, V-P., Huttunen, M., Lonnqvist, J., Standertskjold-Nordenstam, C.-G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C.I., Toga, A.W., 2001. Genetic influences on brain structure. Nature Neurosci. 4, 1253–1258.
Thompson, P.M., Giedd, J.N., Woods, R.P., MacDonald, D., Evans, A.C., Toga, A.W., 2000b. Growth patterns in the developing brain detected by using continuum mechanical tensor maps. Nature 404, 190–193.
Thompson, P.M., MacDonald, D., Mega, M.S., Holmes, C.J., Evans, A.C., Toga, A.W., 1997. Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. J. Comput. Assist. Tomogr. 21, 567–581.
Thompson, P.M., Toga, A.W., 1996. A surface-based technique for warping 3-dimensional images of the brain. IEEE Trans. Med. Imaging 15, 1–16.
Thompson, P.M., Woods, R.P., Mega, M.S., Toga, A.W., 2000a. Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. Hum. Brain Mapp. 8, 81–92.
Tisserand, D.J., Pruessner, J.C., Sanz Arigita, E.J., van Boxtel, M.P.J., Evans, A.C., Jolles, J., Uylings, H.B.M., 2002. Regional frontal cortical volumes decrease differentially in aging: an MRI study to compare volumetric approaches and voxel-based morphometry. NeuroImage 17, 657–669.
Toga, A.W., Thompson, P.M., Mega, M.S., Narr, K.L., Blanton, R.E., 2001. Probabilistic approaches for atlasing normal and disease-specific brain variability. Anat. Embryol. 204, 267–282.
Van Essen, D.C., Drury, H.A., Joshi, S., Miller, M.I., 1998. Functional and structural mapping of human cerebral cortex: solutions are in the surfaces. Proc. Natl. Acad. Sci. USA 95, 788–795.
Viola, P., 1995. Alignment by maximization of mutual information. Ph.D. thesis, Massachusetts Institute of Technology, Boston, MA.
Wagner, G.P., 1994. Homology and Development, in: Hall, B.K. (Ed.), "Homology: The Hierarchical Basis of Comparative Biology, Academic Press, San Diego, pp. 273–299.
Warfield, S.K., Rexilius, J., Huppi, P.S., Inder, T.E., Miller, E.G., Wells III, W.M., Zientara, G.P., Jolesz, F.A., Kikinis, R., 2001. A binary entropy measure to assess nonrigid registration algorithms. MICCAI 2001. Proceedings, Lecture Notes in Computer Science 2208, 266–274.
West, J., Fitzpatrick, J.M., Wan, M.Y., Dawan, B.M., Maurer, C.R., Kessler, R.M., Maciunas, R.J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Suetens, P., Vandermeulen, D., van den Elsen, P.A., Napel, S., Sumanaweera, T.S., Harkness, B., Hemler, P.F., Hill, D.L.G., Hawkes, D.J., Studholme, C., Maintz, J.B.A., Viergever, M.A., Malandain, G., Pennec, X., Noz, M.E., Maguire, G.Q., Pollack, M., Pelizzari, C.A., Robb, R.A., Hanson, D., Woods, R.P., 1997. Comparison and evaluation of retrospective intermodality brain image registration techniques. J. Comput. Assist. Tomogr. 21, 554–566.
Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C., 1998a. Automated image registration. I. General methods and intrasubject, intramodality validation. J. Comput. Assist. Tomogr. 22, 139–152.
Woods, R.P., Grafton, S.T., Watson, J.D.G., Sicotte, N.L., Mazziotta, J.C., 1998b. Automated image registration. II. Intersubject validation of linear and nonlinear models. J. Comput. Assist. Tomogr. 22, 139–152.