

TOPICAL REVIEW

Medical image registration**Derek L G Hill, Philipp G Batchelor, Mark Holden and David J Hawkes**Radiological Sciences, King's College London, Guy's Hospital, 5th Floor Thomas Guy House,
St Thomas' Street, London SE1 9RT, UK

E-mail: derek.hill@kcl.ac.uk

Received 12 June 2000

Abstract

Radiological images are increasingly being used in healthcare and medical research. There is, consequently, widespread interest in accurately relating information in the different images for diagnosis, treatment and basic science. This article reviews registration techniques used to solve this problem, and describes the wide variety of applications to which these techniques are applied. Applications of image registration include combining images of the same subject from different modalities, aligning temporal sequences of images to compensate for motion of the subject between scans, image guidance during interventions and aligning images from multiple subjects in cohort studies. Current registration algorithms can, in many cases, automatically register images that are related by a rigid body transformation (i.e. where tissue deformation can be ignored). There has also been substantial progress in non-rigid registration algorithms that can compensate for tissue deformation, or align images from different subjects. Nevertheless many registration problems remain unsolved, and this is likely to continue to be an active field of research in the future.

Summary of notation

A	An image (reference or target)
B	A second image to be aligned with A
x	A point
$\{x_i\}$	A set of points
x_A	A point in image A
$A(x_A)$	The value of image A at position x_A
T	The spatial transformation (mapping) from B to A
L	A linear transformation
\mathcal{T}	The transformation that maps both position and intensity (i.e. takes account of interpolation and sampling)
B^T	Image B transformed into the coordinate space of image A
$\tilde{\Omega}_A$	The imaged volume or field of view i.e. continuous domain of image A
Ω_A	The discrete domain of image A

Ω_B	The discrete domain of image B
$\Omega_{A,B}^T$	The overlap domain of images A and B for a given transformation estimate T
$N = \sum_{\Omega_{A,B}^T} 1$	The number of pixels or voxels in the overlap domain of images A and B
Ω_a	An isointensity set (level set) in image A with intensity a
Ω_a^T	The isointensity set of A in the overlap domain
$\Omega_{a,b}^T$	Overlapping isointensities
p_i, q_i	Points p, q
p	A probability value (between 0 and 1)
$p_{AB}^T(a, b)$	Joint probability density function: the probability that a voxel in the overlap domain has intensity a in A and b in B^T
\mathcal{F}	Intensity mapping function (makes image A look like image B)
$A _{\Omega_{A,B}^T}$	Image A in the overlap domain
$B^T _{\Omega_{A,B}^T}$	Image B in the overlap domain
μ	Mean
σ	Standard deviation
ζ	Sample spacing in a discrete image
Γ_{ζ_A}	Sampling grid for image A
\mathbb{E}	Euclidean space
\mathbb{R}	Real numbers (1.2, π , $\sqrt{2}$, etc)
\in	In (element of)
\forall	For all

1. Introduction

Medical images are increasingly being used within healthcare for diagnosis, planning treatment, guiding treatment and monitoring disease progression. Within medical research (especially neuroscience research) they are used to investigate disease processes and understand normal development and ageing. In many of these studies, multiple images are acquired from subjects at different times, and often with different imaging modalities. In research studies, it is sometimes desirable to compare images obtained from patient cohorts rather than just single subjects imaged multiple times. Furthermore, the amount of data produced by each successive generation of imaging system is greater than the previous generation. This trend is set to continue with the introduction of multislice helical CT scanning and MR imaging systems with higher gradient strengths. There are, therefore, potential benefits in improving the way in which these images are compared and combined. Current clinical practice normally involves printing the images onto radiographic film and viewing them on a light box. Computerized approaches offer potential benefits, particularly by accurately aligning the information in the different images, and providing tools for visualizing the combined images. A critical stage in this process is the alignment or *registration* of the images, which is the topic of this review article. There have been previous surveys of the medical image registration literature (e.g. Maurer and Fitzpatrick 1993, van den Elsen *et al* 1993, Maintz and Viergever 1998). This article aims to complement them both by describing some more recent literature and by using notation that makes clear some of the practical difficulties in implementing robust registration techniques. Furthermore, this article focuses discussion on some of the most widely used registration algorithms rather than attempting to provide a comprehensive survey of all the literature in this field. For this reason, some very recently devised algorithms are not described, as it is not yet clear whether they will become widely used.

In this article we describe the main approaches used for the registration of radiological images. The most widely used application of medical image registration is aligning tomographic images. That is aligning images that sample three-dimensional space with reasonably isotropic resolution. Furthermore, it is often assumed that between image acquisitions, the anatomical and pathological structures of interest do not deform or distort. This ‘rigid body’ assumption simplifies the registration process, but techniques that make this assumption have quite limited applicability. Many organs do deform substantially, for example with the cardiac or respiratory cycles or as a result of change in position. The brain within the skull is reasonably non-deformable provided the skull remains closed between imaging, and that there is no substantial change in anatomy and pathology, such as growth in a lesion, between scans. Imaging equipment is imperfect, so regardless of the organ being imaged, the rigid body assumption can be violated as a result of scanner-induced geometrical distortions that differ between images. Although the majority of the registration approaches reviewed here have been applied to the rigid body registration of head images acquired using tomographic modalities, there is now considerable research activity aimed at tackling the more challenging problems of aligning images that have different dimensionality (for example projection images with tomographic images), aligning images of organs that deform, aligning images from different subjects, or of aligning images in ways that can correct for scanner-induced geometric distortion. This is quite a rapidly moving research field, so the work reviewed in these areas is more preliminary.

For all types of image registration, the assessment of registration accuracy is very important. The required accuracy will vary between applications, but for all applications it is desirable to know both the expected accuracy of a technique and also the registration accuracy achieved on each individual set of images. For one type of registration algorithm, point-landmark registration, the error propagation is well understood. For other approaches, however, the algorithms themselves provide no useful indication of accuracy. The most promising approach to ensuring acceptable accuracy is visual assessment of the registered images before they are used for the desired clinical or research application.

Although this review concentrates on registration of radiological images of the same subject (*intrasubject* registration), there is some discussion of the closely related topics of intersubject registration, including registration of images of an individual to an atlas, and image-to-physical space registration.

1.1. Types of images

The term ‘medical image’ covers a wide variety of types of images, with very different underlying physical principles, and very different applications. The sort of images used in healthcare and medical research vary from microscopic images of histological sections to video images used for remote consultation, and from images of the eye taken with a fundus camera to whole body radioisotope images. In principle, medical image registration could involve bringing all the information from a given patient, whatever the form, together into a single representation of the individual that acted like a multimedia electronic patient record with implicit information about the spatial and temporal relationship between all the image information. The huge variety of spatial and temporal resolutions and fields of view of the different images makes this difficult, and the clinical benefit of such an approach has not yet been demonstrated. Recent developments in medical image registration have been driven less by this dream of unifying image information than by the practical desire to make better use of certain types of image information for specific clinical applications or in medical research.

In this article we primarily consider the main radiological imaging modalities. These include traditional projection radiographs, with or without contrast and subtraction, nuclear medicine projection images, ultrasound images and the cross-sectional modalities of x-ray computed tomography (CT), magnetic resonance imaging (MRI), single photon emission computed tomography (SPECT) and positron emission tomography (PET). We refer to these last four modalities (CT, MRI, SPECT and PET) as the tomographic modalities. In many ways these are the easiest modalities from the point of view of image registration. They provide voxel datasets in which the sampling is normally uniform along each axis, though the voxels themselves tend to have anisotropic resolution. In a projection x-ray, each pixel represents the integral of attenuation along one of a set of converging lines through the patient, and they represent a superposition of structures in the patient with varying magnification. Many nuclear medicine images acquired with gamma cameras are parallel projections, in which each pixel represents the integral along one of a set of parallel lines through the patient. It is, therefore, a superposition of structures in the patient all at the same magnification (though at different resolutions). The majority of ultrasound images are acquired with a free-hand transducer. Each image is a two-dimensional slice through part of the patient. The images are acquired at a high frame rate, but the spatial relationship between the frames is not recorded. If the transducer is moved in a controlled way or tracked, the relative positions of the frames can be recorded, and a three-dimensional dataset obtained, provided the structures being tracked are not moving or deforming during the acquisition.

Video images are often acquired during surgery, for example using endoscopes or microscopes. For the purpose of image guidance, it can be useful to relate the video images to preoperatively acquired diagnostic images. Video images are, like radiographs and many nuclear medicine images, projections. They differ, however, in that they normally only contain information about the surface of structures in the field of view, rather than a superposition of overlying structures. There is a huge computer vision literature on estimating three-dimensional shapes from one or more video camera views. Video images can be aligned with tomographic images either by first extracting surface structures, or directly.

2. Definitions, notation and terminology

In this article we use the term ‘registration’ to mean determining the spatial alignment between images of the same or different subjects, acquired with the same or different modalities, and also the registration of images with the coordinate system of a treatment device or tracked localizer. Other authors distinguish between different categories of alignment using the words registration, co-registration and normalization. The term normalization is usually restricted to the intersubject registration situation, and registration and co-registration are often used interchangeably. The algorithms used for all these applications have many features in common, so we prefer to use the term registration for all cases.

The word *registration* is used with two slightly different meanings. The first meaning is determining a transformation that can relate the *position* of features in one image or coordinate space with the *position* of the corresponding feature in another image or coordinate space. We use the symbol T to represent this type of positional registration transformation. The second meaning of registration both relates the position of corresponding features and enables us to compare the *intensity* at those corresponding positions (e.g. to subtract image intensity values). We use the symbol \mathcal{T} to describe this second meaning of registration, which incorporates the concepts of resampling and interpolation.

Using the language of geometry, the registration transformation is referred to as a mapping. We can consider the mapping T , that transforms a position x from one image to another, or

from one image to the coordinate system of a treatment device (image to physical registration)

$$T : \mathbf{x}_A \mapsto \mathbf{x}_B \Leftrightarrow T(\mathbf{x}_A) = \mathbf{x}_B. \quad (1)$$

Using this notation, T is a spatial mapping. We also need to consider the more complete mapping \mathcal{T} that maps both position and associated intensity value from image A to image B . \mathcal{T} therefore maps an image to an image, whereas T maps between coordinates. If we wish to overlay two images that have been registered, or to subtract one from another, then we need to know \mathcal{T} , not just T . \mathcal{T} is only defined in the region of overlap of the image fields of view, and has to take account of image sampling and resolution. $A(\mathbf{x}_A)$ is the intensity value at the location \mathbf{x}_A , and similarly for image B ¹. It is important to remember that the medical images A and B are derived from a real object, i.e. the patient. The images have a limited field of view that does not normally cover the entire patient. Furthermore, this field of view is likely to be different for the two images.

We can usefully think of the two images themselves as being mappings of points in the patient within their field of view (or *domain* Ω) to intensity values

$$A : \mathbf{x}_A \in \Omega_A \mapsto A(\mathbf{x}_A)$$

$$B : \mathbf{x}_B \in \Omega_B \mapsto B(\mathbf{x}_B).$$

Because the images are likely to have different fields of view, the domains Ω_A and Ω_B will be different. This is a very important factor, which accounts for a good deal of the difficulty in devising accurate and reliable registration algorithms. We will return to this issue in section 2.1.

To compare images A and B , we would like them both to be defined at the location \mathbf{x}_A , so that we can write $B(\mathbf{x}_A)$. This is wrong, however, as B is not defined at location \mathbf{x}_A . We therefore introduce the notation B^T for the image B transformed with a given mapping T . If T accurately registers the images, then $A(\mathbf{x}_A)$ and $B^T(\mathbf{x}_A)$ will represent the same location in the object to within some error depending on T . If we didn't have to worry about interpolation, we could write $B^T(\mathbf{x}_A)$ instead of $B^T(\mathbf{x}_A)$, but due to the discrete nature of medical images, discussed further in section 2.2, interpolation is necessary for any realistic transformation.

As the images A and B represent one object X , imaged with the same or different modalities, there is a relation between the spatial locations in A and B . Modality A is such that position $\mathbf{x} \in X$ is mapped to \mathbf{x}_A , and modality B maps \mathbf{x} to \mathbf{x}_B . The registration process involves recovering the spatial transformation T which maps \mathbf{x}_A to \mathbf{x}_B over the entire domain of interest, i.e. which maps from Ω_A to Ω_B within the overlapping portion of the domains. We refer to this overlap domain as $\Omega_{A,B}^T$. This notation makes it clear that the overlap domain depends on the domains of the original images A and B , and also on the spatial transformation T . The overlap domain is the positions in the domain of image A that are also in the domain of image B after transformation, and can be defined as:

$$\Omega_{A,B}^T = \{\mathbf{x}_A \in \Omega_A \mid T^{-1}(\mathbf{x}_A) \in \Omega_B\}.$$

As stated earlier, the transformation \mathcal{T} maps both position, and intensity at that position, from one image to another, taking account of issues to do with sampling and interpolation. It is important to emphasize, however, that \mathcal{T} is not an intensity mapping: it does not make image B look like image A by giving a position in image $B^T(\mathbf{x}_A)$ the same intensity as $A(\mathbf{x}_A)$. We use the symbol \mathcal{F} to represent that mapping of intensities from one image to another. For two images differing only by noise, \mathcal{F} will be the identity. In general \mathcal{F} will be a spatially varying (non-stationary) function that is not monotonic.

¹ $A(\mathbf{x}_A)$ and $B(\mathbf{x}_B)$ are normally scalars, but in some circumstances can be vectors (e.g. flow) or tensors (e.g. diffusion). Non-scalar voxel values can add further complications to image transformation which are not addressed in this article.

Registration algorithms that make use of geometrical features in the images such as points, lines and surfaces, determine the transformation \mathcal{T} by identifying features such as sets of image points $\{x_A\}$ and $\{x_B\}$ that correspond to the same physical entity visible in both images, and calculating \mathcal{T} for these features. When these algorithms are iterative, they iteratively determine \mathcal{T} , and then infer \mathcal{T} from \mathcal{T} when the algorithm has converged.

Registration algorithms that are based on image intensity values work differently. They iteratively determine the image transformation \mathcal{T} that optimizes a voxel similarity measure. Many of these voxel similarity measures involves analysing iso-intensity sets (or level sets) within the images. For a single image A , an iso-intensity set with intensity value a is the set of voxels within a subdomain of A , such that

$$\Omega_a = \{x_A \in \Omega_A | A(x_A) = a\}. \quad (2)$$

This equation, put into words, states ‘all locations x_A in the field of view of image A for which the intensity value is a ’.

Some algorithms do not work on iso-intensity sets corresponding to a single intensity value, but on iso-intensity sets corresponding to small groups, or bins, of intensities. For example a 12-bit image may have its intensities grouped into 256 4-bit bins. We use a to mean either individual intensities or intensity bins, as appropriate.

It is important to remember that Ω_a is the iso-intensity set within all of image A that is within the domain Ω_A . As has been stated above, for registration using voxel similarity measures, we work within the overlap domain $\Omega_{A,B}^T$. The iso-intensity set within this overlap domain is, of course, a function of \mathcal{T} . To emphasize this \mathcal{T} dependence, we define the iso-intensity set of image A with value a , within $\Omega_{A,B}^T$ as

$$\Omega_a^T = \{x_A \in \Omega_{A,B}^T | A(x_A) = a\}. \quad (3)$$

Similarly, we can consider an iso-intensity set in image B . Image B , of course, is always the image that we consider transformed, so the definition is slightly different from that for image A . We consider the iso-intensity set to be the set of voxels in the space A that have intensity b in image B^T

$$\Omega_b^T = \{x_A \in \Omega_{A,B}^T | B^T(x_A) = b\}. \quad (4)$$

2.1. Consequences of different image fields of view

For the 3D intrasubject registration task that we are focusing on here, the object being imaged is the same for images A and B , and, while the domains Ω_A and Ω_B are different in extent and position, they are both three dimensional. The overlap domain $\Omega_{A,B}^T$ is, in general, smaller than either Ω_A or Ω_B , and is also a function of the transformation \mathcal{T} . The latter point is important and sometimes overlooked. For registration algorithms that make use of corresponding geometrical features, the difference in field of view of images A and B can cause difficulties, as features identified in one image may not be present in the second. The dependence of $\Omega_{A,B}^T$ on \mathcal{T} as the algorithm iterates is, however, less important. Greater difficulties arise for registration algorithms that make use of image intensity values to iteratively determine \mathcal{T} . The iso-intensity sets used by these algorithms are the sets of fixed intensities within $\Omega_{A,B}^T$. Since $\Omega_{A,B}^T$ changes with \mathcal{T} , algorithms that are too sensitive to changes in $\Omega_{A,B}^T$ may be unreliable.

The difficulty caused by the different field of views of images A and B is further illustrated by considering an approach to rigid body registration called the method of moments (Faber and Stokely 1988). When applying this to images of a part of the body, e.g. the head, that part of the body is first delineated from images A and B using a segmentation algorithm, giving the binary volumes O_A and O_B . The images can then be registered by first aligning the

centroids (from the first-order moment) of O_A and O_B , and then aligning the principal axes of O_A and O_B (from the second-order moment). This approach is, however, unsatisfactory for most medical image registration applications because the first- and second-order moments are highly sensitive to change in the image field of view. In order for this method to work accurately, the object used for the calculations must be entirely within $\Omega_{A,B}^T$, and it is frequently difficult to delineate appropriate structures with this property from clinical images.

To emphasize the importance of this overlap domain in image registration, we introduce notation to make this clear. $A|_{\Omega_{A,B}^T}$ and $B^T|_{\Omega_{A,B}^T}$ are the portions of image A and B^T respectively in the overlap domain. The use of the $A|_{\Omega_{A,B}^T}$ and $B^T|_{\Omega_{A,B}^T}$ notation is equivalent to $A(x_A) \forall x_A \in \Omega_{A,B}^T$ and $B^T(x_A) \forall x_A \in \Omega_{A,B}^T$.

2.2. The discrete nature of the images

A further important property of the medical images with which we work is that they are discrete. That is, they sample the object at a finite number of voxels in three dimensions or pixels in two dimensions. In general, this sampling is different for images A and B ; also, while the sampling is commonly uniform in a given direction, it is anisotropic, that is it varies with direction in the image. Discretization has important consequences for image registration, so it is useful to build this concept into our notational framework.

We can define our domain Ω as

$$\Omega = \tilde{\Omega} \cap \Gamma_{\zeta}$$

where $\tilde{\Omega}$ is a *bounded* continuous set, and could be called the *volume* or *field of view* of the image, and Γ is an infinite discrete grid. Γ is our *sampling* grid, which is characterized by the anisotropic sample spacing $\zeta = (\zeta^x, \zeta^y, \zeta^z)$. The sampling is normally different for the images A and B being registered, and we denote this by introducing sampling grids Γ_{ζ_A} and Γ_{ζ_B} for the domains Ω_A and Ω_B .

For any given T , the intersection of the discrete domains Ω_A and $T(\Omega_B)$ is likely to be the empty set, because no sample points will exactly overlap. In order, therefore, to compare the images A and B for any estimate of T , it is necessary to interpolate between sample positions and to take account of the difference in sample spacing ζ_A and ζ_B . This introduces two problems. Firstly, fast interpolation algorithms are imperfect, introducing blurring or ringing into the image (discussed further in section 9). This changes the image histograms, and hence alters the isointensity sets discussed above. Secondly, we must be careful when the image B being transformed has higher-resolution sampling than image A (i.e. one or more of the elements in ζ_B is less than the corresponding element in ζ_A). In this case, we risk aliasing when we resample B from Ω_B to $\Omega_{A,B}^T$, so we should first blur B with a filter of resolution ζ_A or lower before resampling.

As stated earlier, the transformation \mathcal{T} maps both positions and intensities at these positions. \mathcal{T} , therefore, has to take account of the discrete sampling. The spatial mapping T , in contrast, does not take account of this effect. When we use \mathcal{T} as a superscript of a symbol, as in $B^{\mathcal{T}}$, we are making it clear that the quantity represented by this symbol is dependent on both the spatial mapping T and the interpolation or blurring used during resampling. Figure 1 illustrates the relationship between field of view, domain of the image and the registration transformation.

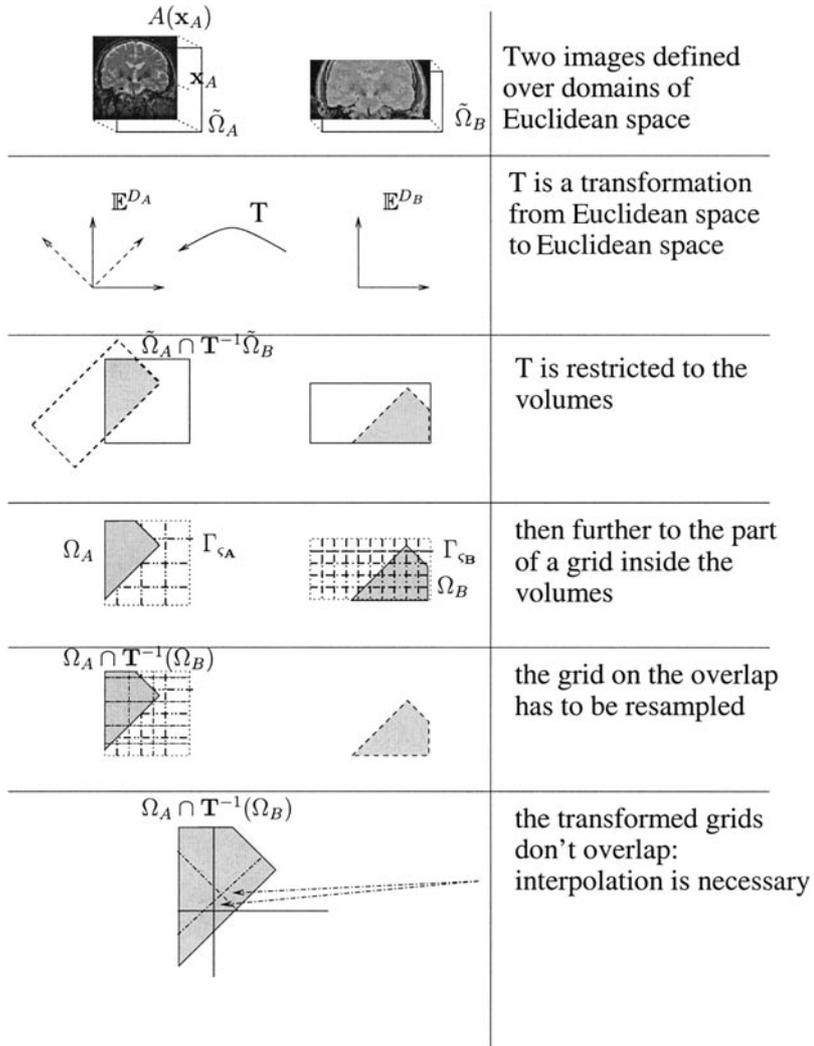


Figure 1. Two images A and B illustrated in the top panel of this figure have different fields of view $\tilde{\Omega}_A$ and $\tilde{\Omega}_B$. For a registration transformation T there will be a region of overlap between the images as illustrated in the third panel. It is important to remember that the images are discrete. The discretization is determined by the sampling grids Γ_{ζ_A} and Γ_{ζ_B} shown in the fourth panel. Even if images A and B have exactly the same sampling grid, the grid points will not normally coincide in the volume of overlap as shown in the fifth panel. Interpolation is therefore necessary. For iterative registration algorithms, this interpolation is necessary at each iteration. The image transformation T described in the text maps both position and the associated intensity in the images within the region of overlap, incorporating the concepts of field of view and discrete sampling.

3. Types of transformation

We have not, so far, discussed the nature of the transformation T . For most current applications of medical image registration, both Ω_A and Ω_B are three dimensional, so T transforms from 3D space to 3D space. In some circumstances, such as the registration of a radiograph to a CT scan, it is useful to consider transformations from 3D space to 2D space (or vice versa). Where both images being registered are two dimensional (e.g. two radiographs before and after contrast is

injected), the appropriate transformation might be from 2D space to 2D space². Most medical image registration algorithms additionally assume that the transformation is ‘rigid body’, i.e. there are six degrees of freedom (or unknowns) in the transformation: three translations and three rotations. The key characteristic of a rigid body transformation is that all distances are preserved.

Some registration algorithms increase the number of degrees of freedom by allowing for anisotropic scaling (giving nine degrees of freedom) and shears (giving 12 degrees of freedom). A transformation that includes scaling and shears as well as the rigid body parameters is referred to as affine, and has the important characteristics that it can be described in matrix form and that all parallel lines are preserved. A rigid body transformation can usefully be considered as a special case of affine, in which the scaling values are all unity and the shears all zero.

Individual bones are rigid at the resolution of radiological imaging modalities, so rigid body registration is widely used in medical applications where the structures of interest are either bone or are enclosed in bone. By far the most important part of the body registered in this way is the head, and in particular the brain. Rigid body registration is used for other regions of the body in the vicinity of bone (e.g. the neck, pelvis, leg or spine) but the errors are likely to be larger.

The use of an affine transformation rather than a rigid body transformation does not greatly increase the applicability of image registration, as there are not many organs that only stretch or shear. Tissues usually deform in more complicated ways. There are, however, several scanner introduced errors that can result in scaling or skew terms, and affine transformations are sometimes used to overcome these problems (see section 4).

For most organs in the body, many more degrees of freedom are necessary to describe the tissue deformation with adequate accuracy. Even in the brain, development of children, lesion growth or resection can make an affine transformation inadequate. Also, it is common in neuroscience research to align images from different subjects. This is called *intersubject* registration and is discussed in section 10.4.3. While an affine transformation is widely used to provide an approximate alignment of different subjects, additional degrees of freedom are needed for more accurate registration. These non-affine registration transformations are not the main focus of this article, but some approaches for both intrasubject and intersubject registration are described in section 10.4.

3.1. Linear transformations

Many authors refer to affine transformations as linear. This is not strictly true, as a linear map is a special map L which satisfies

$$L(\alpha x + \beta x') = \alpha L(x) + \beta L(x') \quad \forall x, x' \in \mathbb{R}^D.$$

The translational parts of affine transformations violate this. An affine map is more correctly described as the composition of linear transformations with translations.

Furthermore, reflections are a linear transformation, but they are normally undesirable in medical image registration. For example, if a registration algorithm used in image-guided neurosurgery calculated a reflection as part of the transformation it might result in a patient having a craniotomy on the wrong side of his or her head. If there is any doubt about whether an algorithm might have calculated a reflection, this should be checked prior to use. Since affine transformations can be represented in matrix form, a reflection can be detected simply from

² Although the images being registered may be 2D, the structures being imaged are often free to translate, rotate and deform in three dimensions.

a negative value of the determinant of this matrix. When we refer to affine transformations elsewhere in this article we exclude reflections as they are undesirable in this application.

3.2. One-to-one transformations

For intrasubject registration, the object being imaged with the same or different modalities is the same patient. It would at first seem likely that the desired transformation \mathcal{T} should be one-to-one (bijective). This means that each location in image B gets transformed to a single point in image A and *vice versa*. There are several situations in which this does not happen. Firstly, if the dimensionality of the images is different such as in the registration of a radiograph to a CT scan, a one-to-one transformation is impossible. The differences in image field of view and sampling discussed in sections 2.1 and 2.2 also complicate this issue.

For various types of non-affine registration, a one-to-one transformation is not desirable. For example in registration of images from different subjects, or of the same subject before and after surgery, there may be structures in image A that are absent from image B .

4. Image distortion

4.1. Geometric distortion

It is important to realize that even when the object being imaged two or more times is rigid, there may be distortions in the imaging process that mean that the correct transformation between the coordinate spaces Ω_A of image A and Ω_B of image B requires additional degrees of freedom. There are many different causes of geometrical distortion in radiological imaging. The types of distortion in images is dependent on the underlying physics of the acquisition, so will vary between the different modalities. It is theoretically possible to correct geometrical distortion. In practice, however, it may be difficult to get the information needed to do this routinely.

The simplest forms of geometric distortion result in scaling and skewing of the coordinate space, and can therefore be described using an affine transformation. Scaling distortion can arise as a result of miscalibration of the voxel dimensions of the images being registered. Skew distortion is most common in CT images, where a tilted gantry rotates the image plane with respect to the axis of the bed, introducing a skew into the images. The gantry tilt angle may not be accurately known, so the skew error, like the scaling error, could be an unknown degree of freedom in the registration transformation \mathcal{T} .

Pin-cushion and barrel distortion lead to a smoothly varying geometrical distortion across the field of view of the images, which is largest at the periphery.

Magnetic resonance images are especially liable to distortion. If the gradients are miscalibrated, then the voxel dimensions will be incorrect, and a scaling error will be present in the data. This can be corrected using phantom measurements (e.g. Hill *et al* 1998, Lemieux and Barker 1998). If the static magnetic field B_0 is not uniform over the subject, then the magnetic field gradients applied during imaging will not result in a linear variation in resonant frequency with position, leading to distortion. The main field inhomogeneity is dependent on the object being imaged, and in particular on local susceptibility changes. The spatial error due to this inhomogeneity is dependent on the field inhomogeneity as a proportion of the imaging gradient strength. The distortion is reduced by increasing the gradient strength, which is equivalent to increasing the bandwidth per pixel. For conventional spin-warp MR imaging (2D multislice or 3D), this distortion is greatest in the read-out direction, and absent in the phase-encode direction(s) where the bandwidth per pixel is essentially infinite because the lines are acquired from different excitations (Chang and Fitzpatrick 1992, Sumanaweera *et al* 1994). For echo

planar imaging, the bandwidth per pixel is *lowest* in the phase encode (or blip) direction, so the distortion is highest in that direction (Jezzard and Balaban 1995). For two-dimensional images, the field inhomogeneity results in the excitation of slices that are curved rather than planar. The magnet-dependent inhomogeneity can be measured with a phantom experiment, but to correct for the object-dependent field inhomogeneity it is necessary to make additional measurements during imaging. For spin-echo images, this can be done by making two measurements with different readout gradient strengths (or opposite gradient directions) (Chang and Fitzpatrick 1992). For gradient echo images (Sumanaweera *et al* 1994) and echo planar images (Jezzard and Balaban 1995), distortion can be inferred from a map of field inhomogeneity.

Patient motion during either CT or multislice MR imaging can result in a variation in slice orientation relative to the patient across the scan. This is currently an unsolved problem, and makes registration of such datasets very difficult. In the rigid body case, there is a different transformation needed for each slice (or in some MR images, groups of slice interleaves), and this transform is hard to find.

Motion during MR imaging within the acquisition of a single slice or an entire 3D volume results in a different problem, namely ghost artefacts. This can result in one or more ‘ghosts’ of the object appearing in the image along with the main image of the object. These ghosts normally have higher spatial frequency content than the main image, but there is a different registration transformation needed for each ghost. Just as the geometric distortion described above can be corrected, these ghost artefacts can be removed (e.g. Atkinson *et al* 1999, McGee *et al* 2000) but this is not routinely performed.

With emission tomography (PET and SPECT), an important cause of distortion errors can be poor alignment of the detector heads in multidetector systems, or uncertainty in the centre of rotation. These lead to a halo artefact that distorts the true distribution. Also, in PET imaging it is important to calibrate the voxel dimensions, as some reconstruction algorithms assume that all photons are detected at the face of the scintillator crystals, but the mean free path of photons in these materials can be 1 cm or so, giving scaling errors in the data.

4.2. Intensity distortion

In addition to the geometric distortion described above, intensity distortion is common in images. This results in the same tissue in different places appearing in the image with varying intensity. Since the intensity distortion is likely to be different between images (even images of the same modality acquired at different times), this effect results in the intensity mapping \mathcal{F} being non-stationary (i.e. changing over the image). The effect is common in MR imaging, where the shading is caused primarily by inhomogeneity in the RF (or B_1) field, and is also present in radiographs, where the heel effect results in a slow variation in intensity across the image. The MR intensity distortion can sometimes be corrected if appropriate assumptions are made about the image (e.g. Guillemaud and Brady 1997, Sled *et al* 1998).

5. Rigid body registration algorithms using geometric features

In this section we describe registration algorithms that determine T using image features that have been extracted from the images either interactively or automatically. The most widely used of these features in this application are points and surfaces, though crest lines and extremal points identified using differential geometric operators are also used. An alternative approach to registration using geometrical features is to generate derived images from A and B that represent the strength of a feature (such as a ridge) at each voxel and then to register these derived images using a voxel similarity measure. Such approaches are described in section 7.1.

5.1. Points and the Procrustes problem

Point-based registration involves identifying corresponding three-dimensional points in the images to be aligned, registering the points and inferring the image transformation from the transformation determined from the points. Using the notation introduced in section 2, we want to find points $\{\mathbf{x}_A^i\}_{i=1\dots N}$ in image A and $\{\mathbf{x}_B^i\}_{i=1\dots N}$ in image B corresponding to the set of features $\{\mathbf{x}^i\}_{i=1\dots N}$ in the object, and find the transformation that aligns them in a least square sense. The corresponding points are sometimes called homologous landmarks to emphasize that they should represent the same feature in the different images.

5.1.1. The orthogonal Procrustes problem. The orthogonal Procrustes problem draws its name from the Procrustes area of statistics. Procrustes was a robber in Greek mythology. He would offer travellers hospitality in his road-side house, and the opportunity to stay the night in his bed that would perfectly fit each visitor. As the visitors discovered to their cost, however, it was the guest who was altered to fit the bed, rather than the bed that was altered to fit the guest. Short visitors were stretched to fit, and tall visitors had suitable parts of their body cut off so they would fit. The result, it seems, was invariably fatal. The hero Theseus put a stop to this unpleasant practice by subjecting Procrustes to his own method.

The term ‘Procrustes’ became a criticism for the practice of unjustifiably forcing data to look like they fit another set. More recently, Procrustes statistics has lost its negative associations and is used in shape analysis. The mathematical problem has relevance in many domains including statistics (Green 1952, Hurley and Cattell 1962, Koschat and Swayne 1991, Schönemann 1966, Rao 1980), gene recognition (Gelfand *et al* 1996), satellite positioning and robotics (Kanatani 1994, Umeyama 1991) in addition to its interest in numerical mathematics as a least square problem (Golub and van Loan 1996, Edelman *et al* 1998, Söderkvist 1993, Stewart 1993). The Procrustes problem is an optimal fitting problem, of least square type: given two configurations of N non-coplanar points $\mathbf{P} = \{\mathbf{p}_i\}$ and $\mathbf{Q} = \{\mathbf{q}_i\}$, one seeks the transformation \mathbf{T} which minimizes $G(\mathbf{T}) = \|\mathbf{T}(\mathbf{P}) - \mathbf{Q}\|^2$. The notation is: \mathbf{P}, \mathbf{Q} are the N -by- D matrices whose rows are the coordinates of the points $\mathbf{p}_i, \mathbf{q}_i$, $\mathbf{T}(\mathbf{P})$ is the corresponding matrix of transformed points, and $\|\dots\|$ is a matrix norm, the simplest being the Frobenius $= (\sum_i \|(\mathbf{T}(\mathbf{p}_i) - \mathbf{q}_i)^2\|)^{1/2}$. The standard case is when \mathbf{T} is a rigid body transformation (Dryden and Mardia 1998, Fitzpatrick *et al* 1998b). One can additionally consider scaling (Dryden and Mardia 1998). If \mathbf{T} is affine, we are faced with a standard least square (Golub and van Loan 1996).

5.1.2. Solutions. The classical Procrustes problem, i.e. $\mathbf{T} \in \{\text{rigid body transformations}\}$ has known solutions. A matrix representation of the rotational part can be computed using singular-value decomposition (SVD) (Dryden and Mardia 1998, Golub and van Loan 1996, Kanatani 1994, Schönemann 1966, Umeyama 1991).

First replace \mathbf{P} and \mathbf{Q} by their demeaned versions, as the optimal transformation is from centroid to centroid

$$\begin{aligned}\mathbf{p}_i &\mapsto \mathbf{p}_i - \bar{\mathbf{p}} \\ \mathbf{q}_i &\mapsto \mathbf{q}_i - \bar{\mathbf{q}}\end{aligned}$$

This reduces the problem to the orthogonal Procrustes problem in which we wish to determine the orthogonal rotation \mathbf{R} . Central to the problem is the D -by- D correlation matrix $\mathbf{K} := \mathbf{P}^t \mathbf{Q}$, as this matrix quantifies how much the points in \mathbf{Q} are ‘predicted’ by points in

P . If $P = [p_1^t, \dots, p_N^t]^t$ is a matrix of row vectors (and the same for Q), $K = \sum_i K_i$ where $K_i := p_i q_i^t$, then

$$K = UDV^t \Rightarrow R = V\Delta U^t \quad \Delta := \text{diag}(1, 1, \det(VU^t))$$

where $K = UDV^t$ is the SVD of K .

It is essential for most medical registration applications that R does not include any reflections. This can be detected from the determinant of VU^t , which should be +1 for a rotation with no reflection, and will be -1 if there is a reflection. In the above equation, Δ takes this into account.

Finally the translation t is given by $t = \bar{q} - R\bar{p}$.

This approach has been widely used in medical image registration, first for multimodality registration (e.g. Evans *et al* 1988, Hill *et al* 1991) and more recently in image-guided surgery (e.g. Maurer *et al* 1997). The points can either be anatomical features that can be identified in 3D, or markers attached to the patient. The theory of errors has been advanced in the medical application domain through the work of Fitzpatrick *et al* (1998b).

5.2. Surface matching

Boundaries, or surfaces, in medical images tend to be more distinct than landmarks, and various segmentation algorithms can successfully locate such high-contrast surfaces. This is especially true of the skin surface—the boundary between tissue and air—which is high contrast in most imaging modalities, with the important exceptions of certain tracers in nuclear medicine emission tomography and some echo planar MR images. If equivalent surfaces can be automatically segmented from two images to be combined, then rigid body registration can be achieved by fitting the surfaces together. The surface matching algorithms described below are normally only used for rigid body registration.

5.2.1. The head-and-hat algorithm. Pelizzari and colleagues (Pelizzari *et al* 1989, Levin *et al* 1988) proposed a surface fitting technique for intermodality registration of images of the head that became known as the ‘head-and-hat’ algorithm. Two equivalent surfaces are identified in the images. The first, from the higher-resolution modality, is represented as a stack of discs, and is referred to as the head. The second surface is represented as a list of unconnected 3D points. The registration transformation is determined by iteratively transforming the (rigid) hat surface with respect to the head surface, until the closest fit of the hat onto the head is found. The measure of closeness of fit used is the square of the distance between a point on the hat and the nearest point on the head, in the direction of the centroid of the head. The iterative optimization technique used is the Powell method (Press *et al* 1992). The Powell optimization algorithm performs a succession of one-dimensional optimizations, finding in turn the best solution along each of the six degrees of freedom, and then returning to the first degree of freedom. The algorithm stops when it is unable to find a new solution with a significantly lower cost (as defined by a tolerance factor) than the current best solution. This algorithm has been used with considerable success for registering images of the head (Levin *et al* 1988), and has also been applied to the heart (Faber *et al* 1991). The surfaces most commonly used are the skin surface delineated from both MR images and PET transmission images, or the brain surface delineated from both MR images and PET emission images. The measure of goodness of fit can be prone to error for convoluted surfaces: the distance metric used by the head-and-hat algorithm does not always measure the distance between a hat point and the closest point on the head surface, because the nearest head point will not always lie in the direction of the head centroid, especially if the surface is convoluted.

5.2.2. Distance transforms. A modification to the head-and-hat algorithm is to use a distance transform to preprocess the head images. A distance transform is applied to a binary image in which object pixels (or voxels) have the value 1 and other voxels have the value 0. It labels all voxels in this image with their distance from the surface of the object. By prelabelling all image voxels in this way, the cost per iteration can be substantially reduced (potentially to a single address manipulation and accumulation for each transformed hat surface point). A widely used distance transform is the chamfer filter proposed by Borgefors (1984). This approach was used for rigid body medical image registration (e.g. Jiang *et al* 1992, van den Elsen 1992, van Herk and Kooy 1994). More recently, exact Euclidean distance transforms have been used in place of the chamfer transform (Huang and Mitchell 1994).

Given estimates for the six degrees of freedom of the rigid body transformation, the hat points are transformed and their distances from the head surface are calculated from the values in the relevant voxels in the distance transform. These values are squared and summed to calculate the cost associated with the current transformation estimate. There remains a risk of finding local minima, so a pyramidal multiscale representation and outlier rejection is desirable (Jiang *et al* 1992).

5.2.3. Iterative closest point. The iterative closest point algorithm (ICP) was proposed by Besl and McKay (1992) for the registration of 3D shapes. It was not designed with medical images in mind, but has subsequently been applied to medical images with considerable success, and is now probably the most widely used surface matching algorithm in medical imaging applications (e.g. Cuchet *et al* 1995, Declerck *et al* 1997, Maurer *et al* 1998). The original article is written in terms of registration of collected data to a model. The collected data, \mathcal{P} , could come from any sensor that provides three-dimensional surface information including laser scanners, stereo video and so forth. The model data, \mathcal{X} , could come from a computer-aided design model. In medical imaging applications, both sets of surface data might be delineated from radiological images, or the model might be derived from a radiological image and the data from stereo video acquired during an operation. The algorithm is designed to work with seven different representations of surface data: point sets, line segment sets (polylines), implicit curves, parametric curves, triangle sets, implicit surfaces and parametric surfaces. For medical image registration the most relevant representations are likely to be point sets and triangle sets, as algorithms for delineating these from medical images are widely available.

The algorithm has two stages and iterates. The first stage involves identifying the closest model point for each data point, and the second stage involves finding the least square rigid body transformation relating these points sets. The algorithm then redetermines the closest point set and continues until it finds the local minimum match between the two surfaces, as determined by some tolerance threshold.

Whatever the original representation of the data surface \mathcal{P} , it is first converted to a set of points $\{p_i\}$. The model data remain in their original representation. The first stage involves identifying, for each point p_i on the data surface \mathcal{P} , the closest point on the model surface \mathcal{X} . This is the point x in \mathcal{X} for which the distance d between p_i and x is minimum

$$d(p_i, \mathcal{X}) = \min_{x \in \mathcal{X}} \|x - p_i\|.$$

The resulting set of closest points (one for each p_i) is $\{q_i\}$. For a triangulated surface, which is the most likely model representation from medical image data, the model \mathcal{X} comprises a set of triangles $\{t_i\}$. The closest model point to each data point is found by linearly interpolating across the facets. If triangle t_i has vertices r_1 , r_2 and r_3 , then the distance between the point p_i and the triangle t_i is

$$d(p_i, t_i) = \min_{u+v+w=1} \|ur_1 + vr_2 + wr_3 - p_i\|$$

where $u \in [0, 1]$, $v \in [0, 1]$ and $w \in [0, 1]$. The closest model point to the data point p_i is, therefore, $q_i = (ur_1, vr_2, wr_3)$.

A least squares registration between the points $\{p_i\}$ and $\{q_i\}$ can then be carried out using the method described in section 5.1³. The set of data points $\{p_i\}$ is then transformed to $\{p'_i\}$ using the calculated rigid body transformation, and then the closest points once again identified. The algorithm terminates when the change in mean square error between iterations falls below a threshold.

The optimization can be accelerated by keeping track of the solutions at each iteration. If there is good alignment between the solutions (to within some tolerance), then both a parabola and a straight line are fitted through the solutions, and the registration estimate is updated using one or the other of these estimates based on a slightly *ad hoc* method to ‘be on the safe side’.

As the algorithm iterates to the local minimum closest to the starting position, it may not find the correct match. The solution proposed by Besl and McKay is to start the algorithm multiple times, each with a different estimates of the rotation alignment, and choose the minimum of the minima obtained.

5.3. Registration using crest lines

An alternative to the idea of using surface matching for registration is to use distinctive features on those surfaces instead. For this to work the surfaces must be smooth, that is possible to differentiate up to third order. Using the tools of differential geometry, it is possible to define two principal curvatures k_1 and k_2 at each point on a surface, (the strength of k_1 is greater than k_2) with associated principal directions. *Crest lines* are the loci of the surface where the value k_1 is locally maximal in its principal direction (Monga and Benayoun 1995). For medical images in which intensity threshold values can determine isosurfaces that delineate structures of interest (e.g. bone from CT scans), these crest lines can be identified directly from the image voxel arrays. For other applications, such as registration involving MRI where intensity shading makes isosurfaces less meaningful, prior surface segmentation is needed. In both cases, smoothing of the data is needed along with differentiation in order to reduce sensitivity to noise.

Images can be registered by aligning the crest lines identified in the images. This has greatest applicability when the images are very similar, in which case there will be good correspondence between crest lines (Gueziec and Ayache 1994, Pennec and Thirion 1997). Alternative approaches include using hash tables of geometrical invariants for each curve (Gueziec *et al* 1997) together with the Hough transform, and using a modification of the iterative closest point algorithm described above (Gueziec and Ayache 1994, Pennec and Thirion 1997). The technique has been applied to registration of CT skull images and T_1 -weighted MR image volumes.

6. Intramodality registration using voxel similarity measures

Registration using voxel similarity measures involves calculating the registration transformation \mathcal{T} by optimizing some measure calculated directly from the voxel values (or pixel values) in the images rather than from geometrical structures such as points or surfaces derived from the images. As was stated in section 2, with voxel similarity measures we are almost invariably iteratively determining \mathcal{T} , whereas in the case of point registration or surface matching we first identify corresponding features, determine \mathcal{T} directly or iteratively from these, and finally infer \mathcal{T} .

³ The authors of the iterative closest point algorithm in fact use the equivalent quaternion method.

In sections 5.1 and 5.2 above we did not distinguish between registration where images A and B are of the same modality and registration of A and B when they are of different modalities. For registration using voxel similarity measures this is an important distinction, as will be seen from the following example. A common reason for carrying out same modality, or intramodality, registration is to compare images from a subject taken at slightly different times in order to ascertain whether there have been any subtle changes in anatomy or pathology. If there has been no change in the subject, we might expect that, after registration and subtraction, there will be no structure in the difference image, just noise. Where there is a small amount of change in the structure, we would expect to see noise in most places in the images, with a few regions visible in which there has been some change. If there were a registration error, we would expect to see artefactual structure in the difference image resulting from the poor alignment. In this application, various voxel similarity measures suggest themselves. We could, for example, iteratively calculate \mathcal{T} while minimizing the structure in the difference image on the grounds that at correct registration there will be either no structure or a very small amount of structure in the difference image, whereas with increasing misregistration, the amount of structure would increase. The structure could be quantified, for example, by the sum of squares of difference values, or the sum of absolute difference values, or the entropy of the difference image. An alternative intuitive approach (at least for those familiar with signal processing techniques) would be to find \mathcal{T} by cross correlation of images A and B . In this section, we describe these intramodality techniques in more detail.

6.1. Minimizing intensity difference

One of the simplest voxel similarity measures is the sum of squared intensity differences between images, SSD, which is minimized during registration. For N voxels in the overlap domain $\Omega_{A,B}^{\mathcal{T}}$

$$\text{SSD} = \frac{1}{N} \sum_{x_A \in \Omega_{A,B}^{\mathcal{T}}} |A(x_A) - B^{\mathcal{T}}(x_A)|^2. \quad (5)$$

It is necessary to divide by the number of voxels N in the overlap domain (the cardinality of $\Omega_{A,B}^{\mathcal{T}}$) because $N = \sum_{\Omega_{A,B}^{\mathcal{T}}} 1$ may vary with each estimate of \mathcal{T} . It can be shown that this is the optimum measure when two images only differ by Gaussian noise (Viola 1995). It is quite obvious that for intermodality registration this will never be the case. This strict requirement is seldom true in intramodality registration either, as noise in medical images such as modulus MRI scans is frequently not Gaussian, and also because there is likely to have been change in the object being imaged between acquisitions or there would be little purpose in registering the images!

The SSD measure is widely used for serial MR registration, for example by Hajnal *et al* (1995a, b). It is also used in a slightly different framework by the registration algorithm within Friston's statistical parametric mapping (SPM) software (Friston *et al* 1995, Ashburner and Friston 1999), as described in more detail in section 7.1 below.

The SSD measure is very sensitive to a small number of voxels that have very large intensity differences between images A and B . This might arise, for example, if contrast material is injected into the patient between the acquisition of images A and B or if the images are acquired during an intervention and instruments are in different positions relative to the subject in the two acquisitions. The effect of these 'outlier' voxels can be reduced by using the sum of absolute differences, SAD rather than SSD:

$$\text{SAD} = \frac{1}{N} \sum_{x_A \in \Omega_{A,B}^{\mathcal{T}}} |A(x_A) - B^{\mathcal{T}}(x_A)|. \quad (6)$$

6.2. Correlation techniques

The SSD measure makes the implicit assumption that after registration, the images differ only by Gaussian noise. A slightly less strict assumption would be that, at registration, there is a linear relationship between the intensity values in the images. In this case, the optimum similarity measure is the correlation coefficient, CC

$$CC = \frac{\sum_{\mathbf{x}_A \in \Omega_{A,B}^T} (A(\mathbf{x}_A) - \bar{A})(B^T(\mathbf{x}_A) - \bar{B})}{\{\sum_{\mathbf{x}_A \in \Omega_{A,B}^T} (A(\mathbf{x}_A) - \bar{A})^2 \sum_{\mathbf{x}_A \in \Omega_{A,B}^T} (B^T(\mathbf{x}_A) - \bar{B})^2\}^{1/2}}. \quad (7)$$

where \bar{A} is the mean voxel value in image $A|_{\Omega_{A,B}^T}$ and \bar{B} is the mean of $B^T|_{\Omega_{A,B}^T}$. This similarity measure has been used for intramodality registration (e.g. Lemieux *et al* 1998). The correlation coefficient can be thought of as a normalized version of the widely used cross correlation measure C

$$C = \sum_{\mathbf{x}_A \in \Omega_{A,B}^T} A(\mathbf{x}_A) B^T(\mathbf{x}_A). \quad (8)$$

6.3. Ratio image uniformity

The ratio image uniformity (RIU) algorithm was originally introduced by Woods *et al* (1992) for the registration of serial PET studies, but has more recently been widely used for serial MR registration (Woods *et al* 1998), and is available in the AIR registration package from UCLA. The algorithm can be thought of as working with a derived ratio image calculated from images A and B . An iterative technique is used to find the transformation \mathcal{T} that maximizes the uniformity of this ratio image, which is quantified as the normalized standard deviation of the voxels in the ratio image. The RIU acronym was not introduced when the algorithm was first published, and it is also frequently referred to as the variance of intensity ratios (VIR) algorithm. The RIU algorithm is most easily thought of in terms of an intermediate ratio image R comprising N voxels within $\Omega_{A,B}^T$ ($N = \sum_{\Omega_{A,B}^T} 1$)

$$R(\mathbf{x}_A) = \frac{A|_{\Omega_{A,B}^T}(\mathbf{x}_A)}{B^T|_{\Omega_{A,B}^T}(\mathbf{x}_A)} \quad \bar{R} = \frac{1}{N} \sum_{\mathbf{x}_A \in \Omega_{A,B}^T} R(\mathbf{x}_A) \quad (9)$$

$$RIU = \frac{\sqrt{\frac{1}{N} \sum_{\mathbf{x}_A \in \Omega_{A,B}^T} (R(\mathbf{x}_A) - \bar{R})^2}}{\bar{R}}. \quad (10)$$

7. Intermodality registration using voxel similarity measures

In section 6 we described algorithms that register images of the same modality by optimizing a voxel similarity measure. Because of the similarity of the intensities in the images being registered, the subtraction, correlation and ratio techniques described have an intuitive basis. With intermodality registration, the situation is quite different. There is, in general, no simple relationship between the intensities in the images A and B . Using the notation of section 2, the intensity mapping function \mathcal{F} is a complicated function and no simple arithmetic operation on the voxel values is, therefore, going to produce a single derived image from which we can quantify misregistration.

In section 7.1 we describe some interesting approaches to applying the intramodality measures of subtraction and correlation to images of different modalities. In the remainder of this section we then describe similarity measures designed to work directly on intermodality images.

In theory these measures could be used to calculate an arbitrary mapping \mathcal{T} . In practice, current intermodality registration techniques almost always involve finding rigid body or affine transformations.

7.1. Using intramodality similarity measures for intermodality registration

It is possible to use image subtraction or correlation techniques for intermodality registration by estimating an intensity mapping function \mathcal{F} and applying this to one image (say A) to create a ‘virtual image’ A_v that has similar intensity characteristics to image B , of a different modality. If the estimate of \mathcal{F} is reasonably good, then the virtual image A_v is sufficiently similar to image B that subtraction or correlation algorithms can be used for registration.

7.1.1. Intensity re-mapping for MR–CT registration. One approach, which works well in MR–CT registration, is to transform the CT image intensities, such that high intensities are re-mapped to low intensities. This creates a virtual image from the CT images that has an intensity distribution more like an MR images (in which bone is dark) (van den Elsen *et al* 1994). The MR image and virtual MR image created from CT are then registered by cross correlation.

7.1.2. Registration of intensity ridges. A slightly different approach is to generate virtual images A_v and B_v from A and B , such that, although A and B are very different from each other, A_v and B_v are quite similar. This can be done by applying scale-space derivatives to both images A and B in order to generate virtual images that comprise intensity ridges (van den Elsen *et al* 1995, Maintz *et al* 1996). The operator used by these authors to quantify ‘ridgeness’ in images is L_{vv} . In their terminology L is the intensity distribution (either 3D or 2D) and a gradient-based local coordinate system is then defined in which w is the direction of maximum gradient and v is the normal to that direction. The direction v will change relative to the Cartesian axis of the image depending on the local intensity landscape. L_{vv} is therefore the second derivative of the intensity landscape at a local position and a selected scale. It is a line in 2D and a surface in 3D (Maintz *et al* 1996). Applying this operator to both images to be registered produces two derived images of ridgeness that can be registered by cross correlation.

7.1.3. The registration algorithm used by the statistical parametric mapping (SPM) software. Registration involves finding the optimal transformation \mathcal{T} . In order to run optimization algorithms, these transformation are assumed to be parametrized, for example the rotations are typically parametrized by Euler angles. We call the collection of parameters $\theta = (\theta_0, \dots, \theta_{K-1})$, and we indicate the parametrization using the notation $\mathcal{T} = \mathcal{T}_\theta$. For non-rigid transformations, the number of parameters K can be very large. In this sense, a general similarity measure $S_{A,B}(\mathcal{T})$ becomes a function of these parameters θ : $S_{A,B}(\theta)$.

Friston *et al* (1995) proposed an alternative to running some optimization algorithm on a similarity measure. In the section on similarity measures, it was mentioned that different relationships between the image intensities are possible. At registration, the intensity mapping could be the identity, i.e. $B(\mathcal{T}(x_A)) = A(x_A) + \epsilon(x_A)$. Alternatively the mapping could be linear, in which case $B^T(x_A) = \alpha A(x_A) + \beta + \epsilon(x_A)$. More generally, we could have some global functional relation $B^T(x_A) = \mathcal{F}(A(x_A)) + \epsilon(x_A)$ or local functional (non-stationary) relationship: $B^T(x_A) = \mathcal{F}(A(x_A), x_A) + \epsilon(x_A)$. In all these equations, $\epsilon(x_A)$ is some error term, which is discarded. Friston *et al* note that this provides one equation for every x_A . Just as

the transformation can be parametrized, they assume that the unknown \mathcal{F} can be parametrized too, say by $\mathbf{u} = (u_0, \dots, u_{L-1})$, with L parameters. If we rewrite the N equations in the form

$$\Phi_{x_A}(\theta, \mathbf{u}) = B^{T_\theta}(x_A) - \mathcal{F}_u(A(x_A), x_A) = 0 \quad (11)$$

we get N implicit equations in $K + L$ unknowns. If $N \geq K + L$ such a system can be formally solved. This is a very general approach, and Friston *et al* propose various specific versions for different applications. The general approach is analogous to the sum of squares of difference (SSD) algorithm described in section 6.1, in which registration is accomplished by minimizing the sum of squares of differences in voxel intensities between a virtual image derived from image A and the corresponding locations in image B

$$\text{SSD} = \|\mathcal{F}(A(x_A)) - B^T(x_A)\|^2. \quad (12)$$

The assumptions in Friston *et al*'s algorithm, however, mean that their solution is likely to be different from the solution obtained by iteratively minimizing SSD.

In order to be able to solve explicitly equation (11), Friston *et al* make a series of assumptions, which actually have the effect of ensuring that K and L are not too big. If Φ is smooth, and we know that the solution is close to our starting estimate θ_0, \mathbf{u}_0 , equation (11) can be linearized by taking the first two terms of the Taylor expansion of Φ , in order to compute an explicit equation: $\Phi_{x_A}(\theta, \mathbf{u}) = \Phi_{x_A}(\theta_0, \mathbf{u}_0) + [\partial_\theta \Phi_{x_A}(\theta_0, \mathbf{u}_0), \partial_u \Phi_{x_A}(\theta_0, \mathbf{u}_0)][\theta, \mathbf{u}]^T$. If we call A the matrix of partial derivatives, then equation (11) takes the simple form $A[\theta, \mathbf{u}]^T = -\Phi_{x_A}(\theta_0, \mathbf{u}_0)$ which can be solved by standard least square techniques. This is not iterative, but obviously, due to the assumptions which have been made, iterative improvement might be required. Good starting estimates are essential for this technique.

7.1.4. Example: MR-PET registration. Here $A = \text{MR}$, $B = \text{PET}$. Friston *et al*'s choice was: T is a rigid body transformations, and the parametrization of $\mathcal{F}(\text{MR}(x_{\text{MR}}), x_{\text{MR}})$ was obtained under the following assumptions: the signal from many PET images comes predominantly from grey matter, so a segmentation of the grey matter in MR should give an approximation to a virtual PET image (Friston *et al* 1995). This image might still have a non-stationary scaling, which is called $u_0(x_{\text{MR}})$. This segmented and scaled MR still has the wrong resolution, which is corrected by low-pass filtering by convolution with a Gaussian kernel. If v_g is the estimated intensity, the segmentation is done by the transformation $\mathcal{F}(\text{MR}(x_{\text{MR}}), x_{\text{MR}}) = u_0(x_{\text{MR}})e^{-(\text{MR}(x_{\text{MR}}) - v_g)^2 / 2\sigma^2}$. The real grey-value intensity is going to vary, so v_g is replaced by $v_g + v(x_{\text{MR}})$. The parameters of (we don't write the argument for clarity) $\mathcal{F}(\text{MR}(\cdot), \cdot) = u_0(\cdot)e^{-(\text{MR}(\cdot) - (v_g + v(\cdot)))^2 / 2\sigma^2}$ are then u_0 and v . Friston *et al* make the parameter transformation $u_1 = u_0v$ and assumes smooth variation over the image to ensure the system is overdetermined.

7.2. Partitioned intensity uniformity

The first widely used intermodality registration algorithm that used a voxel similarity measure was proposed by Woods and colleagues for MR-PET registration soon after they proposed their RIU algorithm (Woods *et al* 1993). Here, we refer to this intermodality algorithm as partitioned intensity uniformity (PIU). This algorithm involved a fairly minor change to the source code of their previously published RIU technique (see section 6.3), but transformed its functionality. This algorithm makes an idealized assumption that 'all pixels with a particular MR pixel value represent the same tissue type so that values of corresponding PET pixels should also be similar to each other'. The algorithm therefore partitions the MR image into 256 separate bins based on the value of the MR voxels, then seeks to maximize the uniformity

of the PET voxel values within each bin. Once again, uniformity within each bin is maximized by minimizing the normalized standard deviation.

It is possible to consider this algorithm also using the concept of ratio images, as in the intramodality RIU algorithm. In the intermodality case, however, instead of generating a single ratio image, one ‘ratio image’ is generated for each of the MR intensity bins. This produces 256 sparse images, one for each isointensity set Ω_a . No explicit division is needed to generate these ‘ratio images’, however, because the denominator image corresponds to a single MR bin. The normalized standard deviation of each of these sparse ‘ratio images’ is then calculated, and the overall similarity measure calculated from a weighted sum of the normalized standard deviations.

In the discussion above, we have described the algorithm in terms of MR and PET registration only. We can now formulate the algorithm more generally in terms of images A and B . It is important to note that the two images are treated differently, so there are two different versions of the algorithm, depending on whether image A or image B is partitioned.

For registration of the images A and B , the partitioned image uniformity measure (PIU) can be calculated in two ways. Either as the sum of the normalized standard deviation of intensities in B for each intensity a in A (PIU_B) or the sum of the normalized standard deviation of intensities in A for each intensity b in B (PIU_A)

$$\text{PIU}_B = \sum_a \frac{n_a}{N} \frac{\sigma_B(a)}{\mu_B(a)} \quad \text{and} \quad \text{PIU}_A = \sum_b \frac{n_b}{N} \frac{\sigma_A(b)}{\mu_A(b)} \quad (13)$$

where

$$\begin{aligned} n_a &= \sum_{\Omega_a^T} 1 & n_b &= \sum_{\Omega_b^T} 1 \\ \mu_B(a) &= \frac{1}{n_a} \sum_{\Omega_a^T} B^T(\mathbf{x}_A) & \mu_A(b) &= \frac{1}{n_b} \sum_{\Omega_b^T} A(\mathbf{x}_A) \\ \sigma_B^2(a) &= \frac{1}{n_a} \sum_{\Omega_a^T} (B^T(\mathbf{x}_A) - \mu_B(a))^2 & \sigma_A^2(b) &= \frac{1}{n_b} \sum_{\Omega_b^T} (A(\mathbf{x}_A) - \mu_A(b))^2. \end{aligned}$$

In words, we can say that n_a is the number of voxels of the isointensity set a in $A|_{\Omega_{A,B}^T}$, and $\mu_B(a)$ and $\sigma_B(a)$ are the mean and standard deviation values of the voxels in $B^T|_{\Omega_{A,B}^T}$ that co-occur with this set. The PIU algorithm is widely used for MR–PET registration, requiring that the scalp is first removed from the MR image to avoid a breakdown of the idealized assumption described above. The technique has never been widely used for registration of other modalities, but its success has inspired considerable research activity aimed at identifying alternative voxel similarity measures for intermodality registration.

7.3. Information theoretic techniques

It can be useful to think of image registration as trying to maximize the amount of shared information in two images. In a very qualitative sense, we might say that if two images of the head are correctly aligned then corresponding structures will overlap, so we will have two ears, two eyes, one nose and so forth. When the images are out of alignment, however, we will have duplicate versions of these structures from A and B .

Using this concept, registration can be thought of as reducing the amount of information in the combined image, which suggests the use of a measure of *information* as a registration metric. The most commonly used measure of information in signal and image processing is the

Shannon–Wiener entropy measure H , originally developed as part of communication theory in the 1940s (Shannon 1948, 1949)

$$H = - \sum_i p_i \log p_i. \quad (14)$$

H is the average information supplied by a set of n symbols whose probabilities are given by $p_1, p_2, p_3, \dots, p_n$.

This formula, save for a multiplicative constant, is derived from three conditions that a measure of choice or uncertainty in a communication channel should satisfy. These are:

- (a) The functional should be continuous in p_i .
- (b) If all p_i equal $\frac{1}{n}$, where n is the number of symbols, then H should be monotonically increasing in n .
- (c) If a choice is broken down into a sequence of choices then the original value of H should be the weighted sum of the constituent H . That is $H(p_1, p_2, p_3) = H(p_1, p_2 + p_3) + (p_2 + p_3)H(\frac{p_2}{p_2+p_3}, \frac{p_3}{p_2+p_3})$.

Shannon proved that the $-\sum p_i \log p_i$ form was the only functional form satisfying all three conditions.

Entropy will have a maximum value if all symbols have equal probability of occurring (i.e. $p_i = \frac{1}{n} \forall i$), and have a minimum value of zero if the probability of one symbol occurring is 1, and the probability of all the others occurring is zero.

An important observation made by Shannon is that any change in the data that tends to equalize the probabilities of the symbols $\{p_1, p_2, p_3, \dots, p_n\}$ increases the entropy. Blurring the symbols is one such operation. For a single image, the entropy is normally calculated from the image intensity histogram in which the probabilities $p_1 \dots p_n$ are the histogram entries⁴. If all voxels in an image have the same intensity a , the histogram contains a single non-zero element with probability of 1, indicating that $A(x_A) = a$ for all x_A . The entropy of this image is $-1 \log 1 = 0$. If this uniform image were to include some noise, then the histogram will contain a cluster of non-zero entries around a peak at the average (mode) intensity value, which will be approximately a . The addition of noise to the image, therefore, tends to equalize the probabilities by ‘blurring’ the histogram which *increases* the entropy. The dependence of entropy on noise is important. One consequence is that interpolation of an image may smooth the image (see section 9 for more detail) which can reduce the noise, and consequently ‘sharpen’ the histogram. This sharpening of the histograms *reduces* entropy.

An application of entropy for intramodality image registration is to calculate the entropy of a difference image. If two identical images, perfectly aligned, are subtracted the result is an entirely uniform image that has zero entropy (as stated above). For two images that differ by noise, the histogram will be ‘blurred’, giving higher entropy. Any misregistration, however, will lead to edge artefacts that further increase the entropy. Very similar images can, therefore, be registered by iteratively minimizing the entropy of the difference image (Buzug and Weese 1998).

7.3.1. Joint entropy. In image registration we have two images A and B to align. We therefore have two values at each voxel location for any estimate of the transformation \mathcal{T} . Joint entropy measures the amount of information we have in the two images combined (Shannon 1948).

⁴ In some applications, image entropy is calculated by treating the value of each pixel or voxel $x_A \in \Omega_A$ as a probability, so entropy is given by $\sum_{x_A \in \Omega_A} A(x_A) \log A(x_A)$, with suitable normalization such that all probabilities sum to 1. In this article, we consider entropy calculated from the histogram of the image, not directly from the voxels in the image.

If A and B are totally unrelated, then the joint entropy will be the sum of the entropies of the individual images. The more similar (i.e. less independent) the images are, the lower the joint entropy compared with the sum of the individual entropies

$$H(A, B) \leq H(A) + H(B). \quad (15)$$

The concept of joint entropy can be visualized using a joint histogram calculated from image A and B^T (Hill *et al* 1994). For all voxels in the overlapping regions of the images ($\mathbf{x}_A \in \Omega_{A,B}^T$), we plot the intensity of this voxel in image A , $A(\mathbf{x}_A)$ against the intensity of the corresponding voxel in image B^T . The joint histogram can be normalized by dividing by the total number of voxels N in $\Omega_{A,B}^T$, and regarded as a joint probability density function (PDF) p_{AB}^T of images A and B . We use the superscript T to emphasize that p_{AB}^T changes with T . Due to the quantization of image intensity values, the PDF is discrete, and the values in each element represent the probability of pairs of image values occurring together. The joint entropy $H(A, B)$ is therefore given by

$$H(A, B) = - \sum_a \sum_b p_{AB}^T(a, b) \log p_{AB}^T(a, b). \quad (16)$$

The number of elements in the PDF can either be determined by the range of intensity values in the two images, or from a partitioning of the intensity space into ‘bins’. For example MR and CT images being registered could have up to 4096 (12 bits) intensity values, leading to a very sparse PDF with 4096 by 4096 elements. The use of between 64 and 256 bins is more common. In the above equation a and b either represent the original image intensities or the selected intensity bins. Joint entropy was simultaneously proposed for intermodality image registration by Studholme *et al* (1995) and Collignon *et al* (1995) at the 1995 Information Processing in Medical Imaging Conference.

As can be seen from figure 2, the joint histograms disperse or ‘blur’ with increasing misregistration such that the brightest regions of the histogram gets less bright, and the number of dark regions is reduced. This arises because misregistration leads to joint histogram entries that correspond to different tissue types in the two images. This increases the entropy. Conversely, when registering images we want to find a transformation that will produce a small number of histogram elements with high probabilities, and give us as many zero-probability elements in the histogram as possible, which will minimize the joint entropy. Registration can, therefore, be thought of as trying to find the transformation that maximizes the ‘sharpness’ of the histogram, thereby minimizing the joint entropy.

The simple form of the equation for joint entropy (equation (16)) can hide an important limitation of this measure. As we have emphasized with the T superscript on the joint probabilities, joint entropy is dependent on T . In particular, p_{AB}^T is very dependent on the overlap $\Omega_{A,B}^T$, which is undesirable. For example, a change in T may alter the amount of air surrounding the patient overlapping in the images A and B . Since the air region contains noise that will tend to occupy the lowest value intensity bins (e.g. $a = 0, b = 0$), changing this overlap will alter the joint probability $p_{AB}^T(0, 0)$. If the overlap of air increases, $p_{AB}^T(0, 0)$ will increase, reducing the joint entropy $H(A, B)$. If the overlap of air decreases, $p_{AB}^T(0, 0)$ will reduce, increasing $H(A, B)$. A registration algorithm that seeks to minimize joint entropy will tend, therefore, to maximize the amount of air in $\Omega_{A,B}^T$, which may result in an incorrect solution. More subtly, interpolation which is needed for both subvoxel translation and any rotation will blur the image, altering the PDF values p_{AB}^T .

7.3.2. Mutual information. A solution to the overlap problem from which joint entropy suffers is to consider the information contributed to the overlapping volume by each image

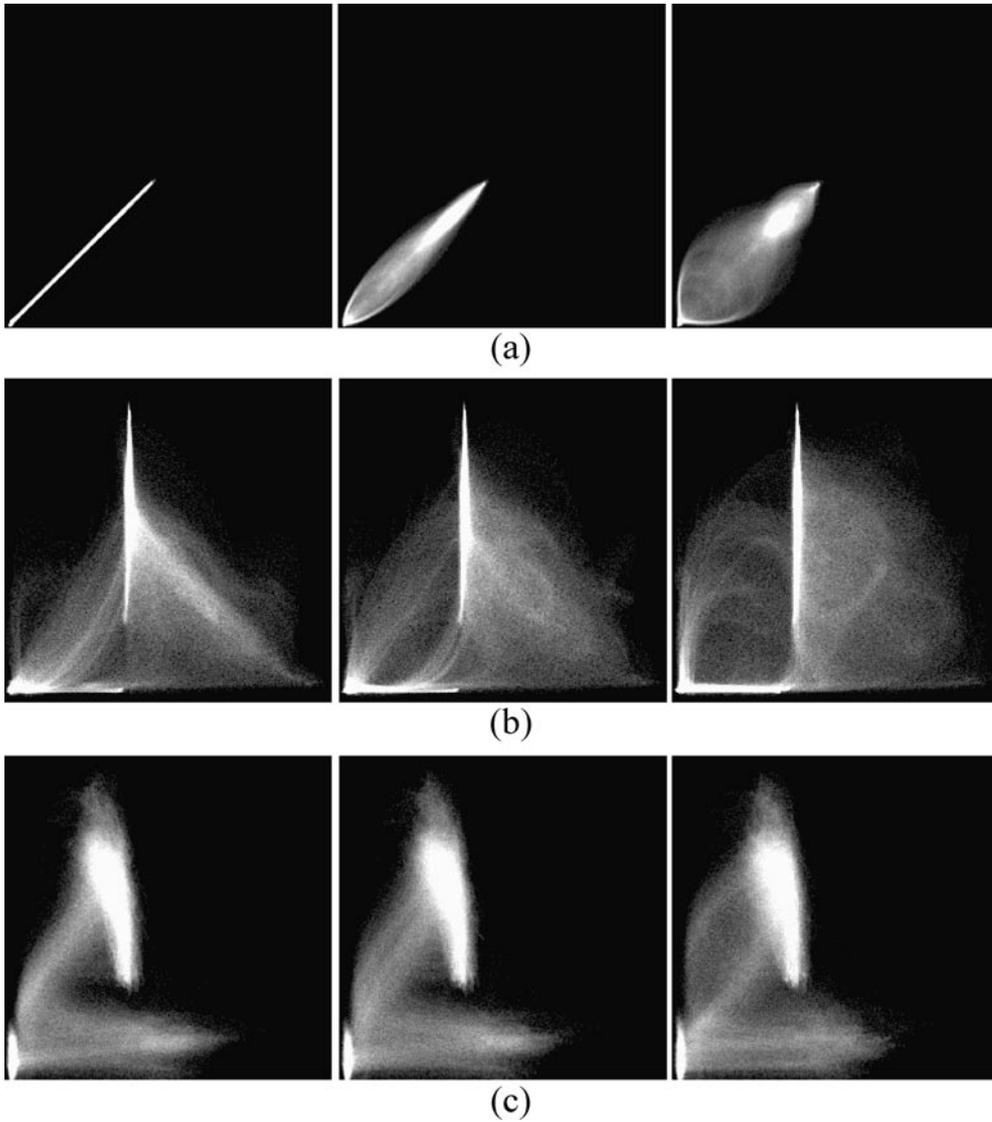


Figure 2. Example 2D histograms from Hill *et al* (1994) (with permission) for (a) identical MR images of the head, (b) MR and CT images of the head and (c) MR and PET images of the head. For all modality combinations, the left panel is generated from the images when aligned, the middle panel when translated by 2 mm, and the right panel when translated by 5 mm. Note that while the histograms are quite different for the different modality combinations, misregistration results in a dispersion or blurring of the signal. Although these histograms are generated by lateral translational misregistration, misregistration in other translation or rotation directions has a similar effect.

being registered together with the joint information. The information contributed by the individual images is simply the entropy of the portion of the image that overlaps with the other image volume:

$$H(A) = - \sum_a p_A^T(a) \log p_A^T(a) \quad \forall A(x_A) = a | x_A \in \Omega_{A,B}^T \quad (17)$$

$$H(B) = - \sum_b p_B^T(b) \log p_B^T(b) \quad \forall B^T(x_A) = b | x_A \in \Omega_{A,B}^T \quad (18)$$

where p_A^T and p_B^T are the marginal probability distributions, which can be thought of as the projection of the joint PDF onto the axes corresponding to intensities in image A and B respectively. It is important to remember that the marginal entropies are not constant during the registration process. Although the information content of the images being registered is constant (subject to slight changes caused by interpolation during transformation), the information content of the portion of each image that overlaps with the other image will change with each change in estimated registration transformation. The superscript T in p_B^T once again emphasizes the dependence of the probabilities on T . The probabilities p_A^T have a superscript T rather than T because image A is the target image which is not interpolated during registration, but the overlap domain nevertheless changes with T .

Communication theory provides a technique for measuring the joint entropy with respect to the marginal entropies. This measure, introduced by Shannon (1948) as ‘rate of transmission of information’ in his article that founded information theory, has become known as mutual information $I(A, B)$. It was independently and simultaneously proposed for intermodality medical image registration by researchers in Leuven, Belgium (Collignon *et al* 1995, Maes *et al* 1997) and MIT in the USA (Viola 1995, Wells *et al* 1996). In maximizing mutual information, we seek for solutions that have a low joint entropy together with high marginal entropies

$$I(A, B) = H(A) + H(B) - H(A, B) = \sum_a \sum_b p_{AB}^T(a, b) \log \frac{p_{AB}^T(a, b)}{p_A^T(a) \cdot p_B^T(b)}. \quad (19)$$

The difference between joint entropy and mutual information is illustrated for serial MR images in figure 3. These plots were obtained from MR images that were acquired perfectly aligned⁵. The correct registration transformation should correspond to zero rotation and zero translation, so we want the optimum value of the similarity measure to be at this position. Figure 3 plots the value of joint entropy, marginal entropies and mutual information for misalignments of between 0 and 6 mm in 0.2 mm increments. Subvoxel translation was achieved using trilinear interpolation which can introduce interpolation errors. The plots therefore show the change in entropies with translation both using the original data (full curve), and using the data pre-filtered with a Gaussian of variance 0.5 voxels to smooth the images and reduce interpolation artifacts (broken curve). These plots demonstrate three important points. Firstly, the marginal entropies change with translation due to change in the overlap domain $\Omega_{A,B}^T$. Secondly, that mutual information ($I(A, B)$) varies more smoothly with misregistration than joint entropy ($H(A, B)$). Thirdly, subvoxel interpolation can blur the images, resulting in reduced entropy that introduces local extrema into parameter space, and the consequences of this are greatly reduced by preblurring the data.

Mutual information can qualitatively be thought of as a measure of how well one image explains the other, and is maximized at the optimal alignment. We can make our description more rigorous if we think more about probabilities. The conditional probability $p(b|a)$ is the probability that B will take the value b given that A has the value a . The conditional entropy is, therefore, the average of the entropy of B for each intensity in A , weighted according to

⁵ A two average gradient echo volume sequence was acquired with isotropic voxels of dimension 2 mm. The raw data were exported from the scanner, and echoes contributing to the two averages were separated and reconstructed as two different images. Because the echoes contributing to the two images are interleaved, the acquisitions are essentially simultaneous, so are registered. However, as the echoes in the two images are obtained from different excitations, random noise and artefact should be different in the two images.

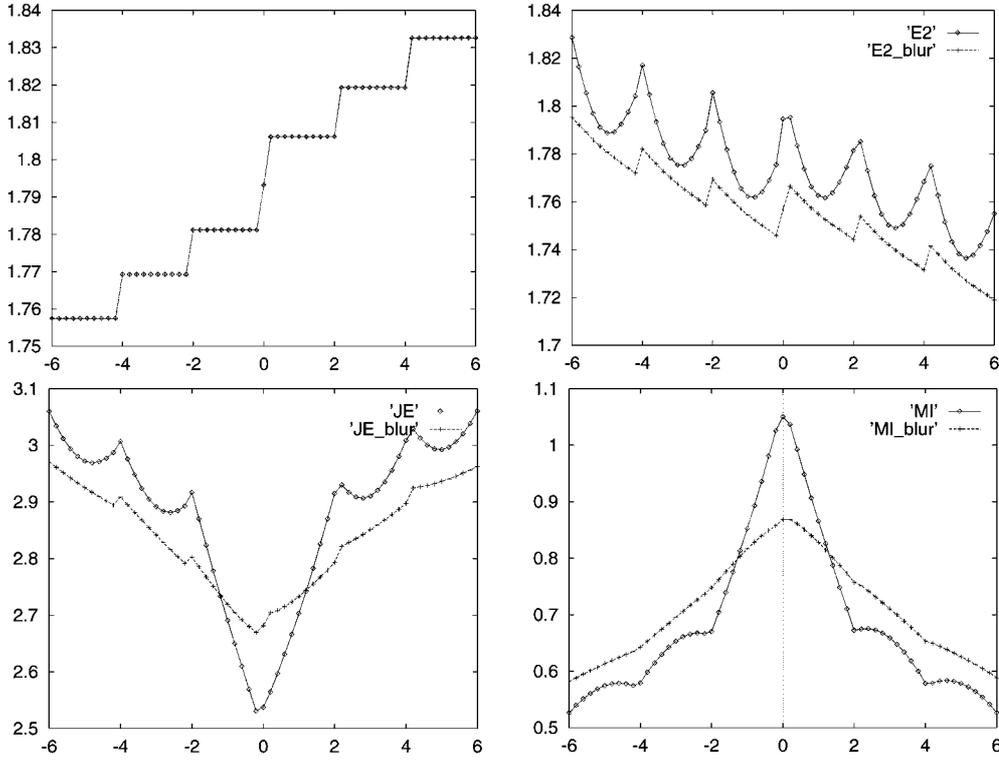


Figure 3. A comparison of change in marginal and joint entropies with cranial–caudal translation for registration of two MR images that differ only by noise. Top left $H(A)$, top right $H(B)$, bottom left $H(A, B)$ and bottom right $I(A, B)$. Each plot has two traces. The full curve is calculated with the images at original resolution, with subvoxel translation achieved using linear interpolation. Note the local extrema of the measures with the period of the voxel separation (2 mm). The broken curve is obtained from images that have been preblurred with a Gaussian kernel with variance $\sigma^2 = 0.5$ voxels. This has no effect on $H(A)$, as image A is not interpolated. For $H(B)$, the preblurring with a Gaussian kernel reduces the interpolation artefacts, and results in a smooth trace for mutual information.

the probability of getting that intensity in A

$$H(B|A) = - \sum_{a,b} p(a, b) \log p(b|a) = H(A, B) - H(A). \quad (20)$$

Using equation (20), we can rewrite our previous expression for mutual information (equation (19))

$$I(A, B) = H(A) - H(B|A) = H(B) - H(A|B). \quad (21)$$

The maximization of mutual information, therefore, involves minimizing the conditional entropy with respect to the entropy of one of the images. The conditional entropy will be zero if, knowing the intensity value $A(x_A)$, we can exactly predict the intensity value in B^T . The conditional entropy will be high if a given intensity value $A(x_A)$ in image A is a very poor predictor of the intensity value $B^T(x_A)$ in the corresponding location in image B .

Given that the images are of the same object, then when they are correctly registered, corresponding voxels in the two images will be of the same anatomical or pathological structure. Knowing the value of a voxel in image A should, therefore reduce the uncertainty (and hence entropy) for the value of the corresponding location in image B . This can be thought of as

a generalization of the assumption made by Woods in his PIU measure. The PIU measure assumes that at registration the uniformity of values in B corresponding to a given value a in A should be minimal. The information theoretic approaches assume that, at alignment, the value of a voxel in A is a good predictor of the value at the corresponding location in B . As misregistration increases, one image becomes a less good predictor of the second. In practical terms, the advantage of mutual information over PIU is that two structures which have the same intensity in image A may have very different intensities in image B . For example, in an MR image, cortical bone and air will both have very low intensities, whereas in CT, air will have a very low intensity but cortical bone a high intensity. If we have a low-intensity voxel in image A , then at correct alignment we know that this should either be air or bone in the CT image⁶. The histogram of CT intensity values corresponding to low intensities in MR will, therefore, have sharp peaks at both low-intensity values and high-intensity values, and the sharp peaks in the histogram give us low entropy. Even though the MR intensity is a good predictor of the CT intensity, however, the PIU would give a very low uniformity value. Mutual information has been compared with other voxel similarity measure for MR–CT and MR–PET registration (e.g. Studholme *et al* 1996, 1997).

7.3.3. Normalized mutual information. Mutual information does not entirely solve the overlap problem described above. In particular, changes in overlap of very low-intensity regions of the image (especially air around the patient) can disproportionately contribute to the mutual information. As was stated earlier, Shannon (1948) was the first to present the functional form of mutual information, calling it the ‘rate of transmission of information’ in a noisy communication channel between source and receiver. In his application in telecommunications, the time over which the different measurements of the source and receiver are made are constant, by definition. In image registration, however, the quantity analogous to time is the total number of image data in the overlap domain $\Omega_{A,B}^T$, and this changes with the transformation estimate T . To remove this dependence on volume of overlap we should normalize to the combined information in the overlapping volume.

Three normalization schemes have so far been proposed in journal articles to address this problem. Equations (22) and (23) below were mentioned in passing in the discussion section of Maes *et al* (1997), though no results were presented comparing them with standard mutual information (equation (19))

$$\tilde{I}_1(A, B) = \frac{2I(A, B)}{H(A) + H(B)} \quad (22)$$

$$\tilde{I}_2(A, B) = H(A, B) - I(A, B). \quad (23)$$

Studholme *et al* (1999) have proposed an alternative normalization devised to overcome the sensitivity of mutual information to change in image overlap. This measure involves normalizing mutual information with respect to the joint entropy of the overlap volume

$$\tilde{I}_3(A, B) = \frac{H(A) + H(B)}{H(A, B)} = \frac{I(A, B)}{H(A, B)} + 1 = \frac{1}{\tilde{I}_1(A, B) - 2}. \quad (24)$$

This third version of normalized mutual information has been shown to be considerably more robust than standard mutual information for intermodality registration in which the overlap volume changes substantially (Studholme *et al* 1999). For serial MR registration, when images A and B have virtually identical fields of view, however, mutual information

⁶ We are assuming in this example that the MR sequence being used will give us no other structures with very low intensity.

and normalized mutual information (equation (24)) have been shown to perform equivalently (Holden *et al* 2000).

8. Optimization and capture ranges

With the exception of registration using the Procrustes technique described in section 5.1, and in certain circumstances the registration algorithm in SPM described in section 7.1, all the registration algorithms reviewed in this article require an iterative approach, in which an initial estimate of the transformation is gradually refined by trial and error. In each iteration, the current estimate of the transformation is used to calculate a similarity measure. The optimization algorithm then makes another (hopefully better) estimate of the transformation, evaluates the similarity measure again, and continues until the algorithm converges, at which point no transformation can be found that results in a better value of the similarity measure, to within a preset tolerance. A review of optimization algorithms can be found in (Press *et al* 1992).

One of the difficulties with optimization algorithms is that they can converge to an incorrect solution called a 'local optimum'. It is sometimes useful to consider the parameter space of values of the similarity measure. For rigid body registration there are six degrees of freedom, giving a six-dimensional parameter space. Each point in the parameter space corresponds to a different estimate of the transformation. Non-rigid registration algorithms have more degrees of freedom (often many thousands), in which case the parameter space has correspondingly more dimensions. The parameter space can be thought of as a high-dimensionality image in which the intensity at each location corresponds to the value of the similarity measure for that transformation estimate. If we consider dark intensities as good values of similarity, and high intensities as poor ones, an ideal parameter space image would contain a sharp low intensity optimum with monotonically increasing intensity with distance away from the optimum position. The job of the optimization algorithm would then be to find the optimum location given any possible starting estimate.

Unfortunately, parameter spaces for image registration are frequently not this simple. There are often multiple optima within the parameter space, and registration can fail if the optimization algorithm converges to the wrong optimum. Some of these optima may be very small, caused either by interpolation artefacts (discussed further in section 9) or a local good match between features or intensities. These small optima can often be removed from the parameter space by blurring the images prior to registration. In fact, a hierarchical approach to registration is common, in which the images are first registered at low resolution, then the transformation solution obtained at this resolution is used as the starting estimate for registration at a higher resolution, and so on.

Multiresolution approaches do not entirely solve the problem of multiple optima in the parameter space. It might be thought that the optimization problem involves finding the globally optimal solution within the parameter space, and that a solution to the problem of multiple optima is to start the optimization algorithm with multiple starting estimates, resulting in multiple solutions, and choose the solution which has the lowest value of the similarity measure. This sort of approach, called 'multistart' optimization, can be effective for surface matching algorithms. For voxel similarity measures, however, the problem is more complicated. The desired optimum when registering images using voxel similarity measures is frequently *not* the global optimum, but is one of the local optima. The following example serves to illustrate this point. When registering images using joint entropy or mutual information, an extremely good value of the similarity measure can be found by transforming the images such that only air in the images overlaps. This will give a few pixels in the joint histogram with very high probabilities, surrounded by pixels with zero probability. This is a very low entropy situation, and will tend

to have lower entropy than the correct alignment. The global optimum in parameter space will, therefore, tend to correspond to an obviously incorrect transformation. The solution to this problem is to start the algorithm within the ‘capture range’ of the correct optimum, that is within the portion of the parameter space in which the algorithm is more likely to converge to the correct optimum than the incorrect global one. In practical terms, this requires that the starting estimate of the registration transformation is reasonably close to the correct solution. The size of the capture range depends on the features in the images, and cannot be known *a priori*, so it is difficult to know in advance whether the starting estimate is sufficiently good. This is not, however, a very serious problem, as visual inspection of the registered images, described further in section 11, can easily detect convergence outside the capture range. In this case, the solution is clearly and obviously wrong (e.g. relevant features in the image do not overlap at all). If this sort of failure of the algorithm is detected, the registration can be re-started with a better starting estimate obtained, for example, by interactively transforming one image until it is approximately aligned with the other.

9. Image transformation

Image registration using voxel similarity measures involves determining the transformation \mathcal{T} that relates the domain of image A to image B . This transformation can then be used to transform one image into the coordinates of the second within the region of overlap of the two domains $\Omega_{A,B}^{\mathcal{T}}$. As discussed in section 2 above, this process involves interpolation, and needs to take account of the difference in sample spacing in images A and B .

9.1. A consideration of sampling and interpolation theory

Shannon (1949) showed that a bandlimited signal sampled with an infinite periodic sampling function can be perfectly interpolated using the sinc function interpolant previously proposed by the mathematician Whittaker (1935). Many medical images are, however, not bandlimited. For example, multislice datasets are not bandlimited in the through-slice direction, as the field of view is truncated with a top-hat function. Even in MR image volumes reconstructed using a 3D Fourier transform, the condition is not usually satisfied because the image data provided by the scanner are often truncated to remove slices at the periphery at the field of view. Also the data provided are usually the modulus of the signal, and taking the modulus is a nonlinear operation that can increase the spatial frequency content.

Even if the images being transformed were strictly bandlimited, it would not be possible to carry out perfect interpolation using a sinc function because a sinc function is infinite in extent.

For many purposes, this problem is entirely ignored during medical image analysis. The most widely used image interpolation function is probably trilinear interpolation, in which a voxel value in the transformed coordinates is estimated by taking a weighted average of the nearest eight neighbours in the original dataset. The weightings, which add up to one, are inversely proportional to the distance of each neighbour to the new sample point. For the accurate comparison of registered images, for example by subtracting one image from another, the errors introduced by trilinear interpolation become important. It can be shown that trilinear interpolation applies a low-pass filter to the image and introduces aliasing (Parker *et al* 1983). For transformations that contain rotations, the amount of low-pass filtering varies with position in the image. If subtracting one image from another to detect small change, for example in serial MR imaging, the low-pass filtering in this process can lead to substantial artefacts. Subtracting a low-pass filtered version of an image from the original is a well known

edge enhancement method, so even in the case of identical images differing only by a rigid body transformation, using linear interpolation followed by subtraction does not result in the expected null result but instead results in an edge enhanced version of the original.

Hajnal *et al* (1995a) recently brought this issue to the attention of MR image analysts and proposed that the solution is to interpolate using a sinc function truncated with a suitable window function such as a Hamming window. Care must be taken when truncating the interpolation kernel to ensure that the integral of the weights of the truncated kernel is unity, or an artefactual intensity modulation can result (Thacker *et al* 1999).

Various modifications to sinc interpolation have recently been proposed. These fall into three categories. Firstly, the use of sinc functions with various radii truncated with various window functions (Lehmann *et al* 1999). Secondly, approximations to windowed sinc functions such as cubic or B-spline interpolants (Lehmann *et al* 1999, Unser 1999). Thirdly, the shear transform, which involves transforming the image using a combination of shears (Eddy *et al* 1996, Cox and Jesmanowicz 1999). This third approach is fast, though it does result in artefacts in the corners of the image which must be treated with caution.

An assumption implicit in the discussion above is that the original data being interpolated are uniformly sampled. This is not always the case in medical images. MR physics researchers are used to the problem of non-uniform sampling in the acquisition, or k -space domain (Robson *et al* 1997, Atkinson *et al* 2000), but this problem is less often considered in the spatial domain. The most common circumstances when non-uniform sampling arises are in free-hand 3D ultrasound acquisition and certain types of CT acquisition where the slice spacing changes during the acquisition. The correct way of interpolating from non-uniformly sampled data onto a uniform grid is the reverse of sinc interpolation. This methodology, sometimes used in k -space regridding (Robson *et al* 1997, Atkinson *et al* 2000), involves calculating the sinc coefficients to go from the desired uniform sampling points to the non-uniform locations acquired, and inverting the matrix of coefficients in order to do the correct interpolation. In the cases of 3D ultrasound and CT variable slice sample spacing, the data are a long way from being bandlimited, so the benefits of inverse sinc interpolation may be small in any case.

9.2. Interpolation during registration

Many registration algorithms involve iteratively transforming image A with respect to image B while optimizing a similarity measure calculated from the voxel values. Interpolation errors can introduce modulations in the similarity measure with \mathcal{T} . This is most obvious for transformations involving pure translations of datasets with equal sample spacing, where the period of the modulation is the same as the sample spacing (Pluim *et al* 2000). Figure 3 illustrates this effect. This periodic modulation of the similarity measure introduces local minima that can lead to the incorrect registration solution being determined.

The computational cost of 'correct' interpolation is normally too high for this approach to be used in each iteration, so lower-cost interpolation techniques must be used. There are several possible approaches. The first is to use low-cost interpolation, such as trilinear or nearest neighbour, until the transformation is close to the desired solution, then carry out the final few iterations using more expensive interpolation. An alternative strategy is to take advantage of the spatial-frequency dependence of interpolation errors. Trilinear interpolation low-pass filters the data, and therefore if the images are blurred prior to registration (high spatial frequency components are removed), the interpolation errors are smaller so errors in the registration are less. Although the loss of resolution that results from blurring is a disadvantage, the registration errors caused by the interpolation errors can be greater than the loss of precision resulting from blurring.

9.3. Transformation for intermodality image registration

It should be emphasized that these interpolation issues are more critical for intramodality registration where accuracy of considerably better than a voxel is frequently desired, than for intermodality image registration. In intermodality registration, one image is frequently of substantially lower resolution than the other, and the desired accuracy of the order of a single voxel at the higher resolution. Furthermore, it is common for the final registration solution to be used to transform the lower resolution image to the sample spacing of the higher resolution modality. Interpolation errors are still likely to be present if trilinear interpolation is used without care, and may slightly reduce the registration accuracy, or degrade the quality of the transformed images.

10. Registration applications

The most widely used applications of image registration involve determining the rigid body transformation that aligns 3D tomographic images of the same subject acquired using different modalities (intermodality registration), or the same subject imaged with a single modality at different times (intramodality registration). There is increasing interest in non-rigid registration of the same or different subjects, registration of 2D images with 3D images, and registration of images with the physical coordinates of a treatment system. In this section we give examples of some applications of these approaches.

10.1. Rigid-body and affine intermodality registration

The motivation behind the majority of work in medical image registration during the 1980s was the desire to combine complementary information about the same patient from different imaging modalities (e.g. Maguire *et al* 1986, Chen *et al* 1985, Peters *et al* 1986, Schad *et al* 1987, Levin *et al* 1988, Faber and Stokely 1988, Evans *et al* 1988, Meltzer *et al* 1990, Hill *et al* 1991). With the introduction of MR imaging into many hospitals, and increasing use of CT and tomographic nuclear medicine modalities, it was becoming common for patients to be imaged with more than one tomographic modality during diagnosis or planning of treatment. Intermodality image registration provided a solution to the problem of relating images from different modalities that differ in field of view, resolution, and slice orientation. The head, and the brain in particular, has always been the major area of application for these techniques partly because the skull makes the rigid body assumption valid. During the 1990s, automatic algorithms that optimize a voxel similarity measure were devised for these applications, and these have been shown to be more accurate and robust than feature based registration algorithms for rigid body MR–CT registration and to a lesser extent also for rigid body MR–PET registration (West *et al* 1999). The information theoretic approaches described in section 7.3 have been shown to be the most generally applicable and robust (Studholme *et al* 1996, 1997). Figure 4 shows example MR and CT images before and after registration using an affine transformation. An affine transformation was used in this case to compensate for scaling and skew errors in the data as well as finding the translations and rotations needed to align the images.

10.2. Rigid-body intramodality registration

It is becoming common for a subject to be imaged two or more times with the same modality for diagnosis, treatment monitoring or for research. The images may be separated in time by seconds (e.g. in functional experiments), minutes (e.g. pre- and postcontrast images), or

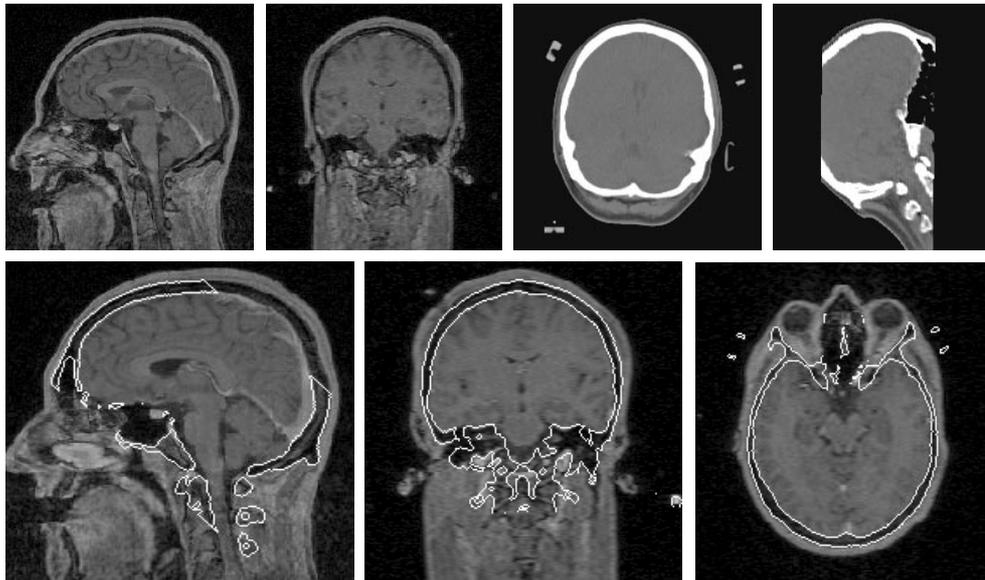


Figure 4. Top row: unregistered MR (left) and CT (right) images. The MR images are shown in the original sagittal plane, and reformatted coronal plane. The CT images in the original oblique plane, and reformatted sagittal plane. Note the different field of view of the images. Bottom panel, MR images in sagittal, coronal and axial planes with the outline of bone, thresholded from the registered CT scan, overlaid. The registration transformation has 10 degrees of freedom, to correct for errors in the machine supplied voxel dimensions and gantry tilt angle.

weeks (e.g. monitoring tumour growth). In all these applications it is desirable to have high sensitivity to small changes in the images. In functional experiments the signal in some voxels may change by a few per cent between the resting and activated state. In contrast or perfusion studies, it is desirable to identify regions that enhance, or quantify intensity change in a region of interest. In longer-term studies, it is desirable to detect small changes in lesion volume or degenerative change in order to plan treatment or monitor response to therapy. Visual inspection of images on a light-box has been shown to be less sensitive to these changes than looking at difference images (Denton *et al* 2000). Since patients often move during examinations, and cannot be repositioned perfectly on subsequent visits, registration is an essential prerequisite for subsequent analysis. Techniques to register intramodality images of the brain using a rigid-body transformation were an area of active research during the 1990s (e.g. Woods *et al* 1992, 1998, Hajnal *et al* 1995a,b, Freeborough *et al* 1996, Lemieux *et al* 1998, Holden *et al* 2000). It might at first seem that intramodality registration is a much simpler problem than intermodality registration because the images being aligned are very similar to one another. It turns out, however, that registration accuracy of much less than a voxel is necessary, so great care must be taken to handle the image transformation issues discussed in section 9 (Hajnal *et al* 1995a). While similarity measures such as SSD, RIU and CC described in section 6 are successfully used for intramodality registration of brain images, care must be taken in the optimization and interpolation to ensure high-quality results. Furthermore, although the brain can be accurately aligned using a rigid body transformation, deformable regions such as the scalp and neck, or regions that change substantially between acquisitions (e.g. due to contrast uptake), can bias the result, leading to significant errors. Presegmentation of the images to exclude these regions can be necessary (Hajnal *et al* 1995a). Alternatively, it has been shown

that the information theoretic similarity measures discussed in section 7.3 can be less sensitive to these changes, and may have an advantage over the more obvious intramodality measures described in section 6 for this application (Holden *et al* 2000).

10.3. 2D–3D registration

There is increasing interest in the registration of 2D projection images with 3D tomographic images. One example of a 2D–3D registration problem is aligning video images with tomographic modalities. This can be achieved using surface reconstruction from video followed by surface matching (e.g. Colchester *et al* 1996, Grimson *et al* 1996). Registration of video images with rendered views calculated from tomographic images has also been proposed (e.g. Viola 1995, Clarkson *et al* 1999, 2000), but these approaches have not yet been shown to work reliably on clinical data.

The predominant 2D–3D registration application is registration of fluoroscopy images and CT images in order to align preinterventional CT scans with interventional x-ray images (Lemieux *et al* 1994a, Weese *et al* 1997, Penney *et al* 1998, Lavallée and Szeliski 1995). The most widely used approach to this problem is to simulate projection x-rays from the CT scans to produce digitally reconstructed radiographs (DRRs), and to iteratively estimate the unknown pose of the x-ray set relative to the CT volume by optimizing a voxel similarity measure calculated from the fluoroscopy image and DRR. If the DRR is a good approximation to the fluoroscopy image, then cross correlation is an appropriate measure (Lemieux *et al* 1994b). A particular problem with interventional fluoroscopy images is that they often contain high-contrast objects such as instruments and prostheses that have been added to the field of view. While bone and soft tissue in accurately simulated DRRs may have similar intensity values to the corresponding pixels in aligned fluoroscopy images, the instruments and other added objects are, of course, absent in the preinterventional CT scan. Many voxel similarity measures are extremely sensitive to a small number of pixels that have large differences between the modalities, and both cross correlation and mutual information have been found to be unreliable for this task (Penney *et al* 1998).

10.3.1. Pattern intensity. Pattern intensity is a similarity measure proposed by Weese *et al* (1997) to register fluoroscopy and CT images. It differs from other measures in that it calculates the value of the similarity measure at each pixel using a neighbourhood of other pixels, rather than just the corresponding pixels themselves. Considering image A to be the x-ray fluoroscopy image, and image B^T to be the DRR generated from the CT scan B , pattern intensity is defined in terms of a difference image D . We describe the pattern intensity measure below in 2D, as originally proposed, although it can also be extended in 3D

$$D(\mathbf{x}_A) = A(\mathbf{x}_A) - \lambda B^T(\mathbf{x}_A)$$

where λ is an intensity scaling factor. The pattern intensity, PI, is then calculated from the difference image over a neighbourhood of N pixels within a radius r each pixel \mathbf{x}_A

$$PI_{r,\sigma} = \sum_{\mathbf{x}_A \in \Omega_{A,B}^T} \sum_{u,v} \frac{\sigma^2}{\sigma^2 + (D(\mathbf{x}_A) - D(u,v))^2}$$

where $\{u, v\}$ are pixels in the neighbourhood of \mathbf{x}_A such that $|\mathbf{x}_A - (u, v)| \leq r$. There are two parameters required by this similarity measure. The first is the radius of the neighbourhood r . Increasing r can improve the reliability of the algorithm, but increases the computational requirement. Values of between 3 and 5 are proposed. The parameter σ controls the sensitivity

of the measure to image features. Weese suggests it should be larger than the standard deviation of the noise in the fluoroscopy image but smaller than the contrast of structures of interest.

To understand how pattern intensity works, it is useful to compare it with the sum of squares of intensity difference measure introduced in section 6.1 above. If pixels $A(x_A)$ and $B^T(x_A)$ are identical, and part of uniform patches of radius $R > r$, then pixel x_A will contribute 0 to the sum of squares of difference measure and $N \frac{\sigma^2}{\sigma^2+0} = N$ to the pattern intensity measure. The corresponding contribution of pixel x_A if it is part of a uniform region in the presence of noise δ would be δ^2 for the sum of squares of difference measure, and approximately $N \frac{\sigma^2}{\sigma^2+\delta^2} \approx N$ to pattern intensity. For a uniform region of very large difference in intensity $\Delta \gg \sigma$, the pixel x_A will contribute the large value of Δ^2 to the sum of squares of difference measure, and approximately $N \frac{\sigma^2}{\sigma^2+\Delta^2} \approx 0$ to the pattern intensity measure. As Δ increases, the sum of squares of intensity measure will increase as Δ^2 , but pattern intensity will asymptotically approach zero, with the consequence that regions of very great difference in intensity contribute proportionately less to pattern intensity than to the sum of squares of intensity difference. Furthermore, if x_A is a single pixel with a large difference Δ between modalities, surrounded by a uniform patch that is the same in both modalities, then this pixel will still give a contribution Δ^2 to the sum of squares of difference measure, but a contribution of $N - 1 \approx N$ to pattern intensity. Pattern intensity is, therefore, almost totally insensitive to individual pixels, or small numbers of pixels $n \ll N$ where there is a very large difference between modalities. Since high-contrast instruments in the fluoroscopy image have exactly this characteristic, pattern intensity is much less sensitive to these objects than sum of square of intensity differences.

10.4. Non-rigid registration: beyond the affine transformation

As was stated in the introduction, aligning images using registration transformations with more degrees of freedom than the affine transformation is an active area of research in which rapid progress is being made. This review does not, therefore, attempt a comprehensive examination of the area (which would inevitably become out of date extremely rapidly). Instead we give a brief introduction to the field.

Non-affine registration algorithms normally either include an initial rigid body or affine transformation, or are run after a rigid-body or affine algorithm has provided a starting estimate. Many different approaches are used. When point landmarks are available, thin-plate splines are often used to determine the transformation (Bookstein 1989). Using intensity-based algorithms (with the measures described in sections 6 and 7), the non-rigid component of the transformation can be determined using a linear combination of polynomial terms (Woods *et al* 1998), basis functions (Friston *et al* 1995) or B-spline surfaces defined by a regular grid of control points (Rueckert *et al* 1999). Alternatively, pseudophysical models can be used in which the deformation of one image with respect to another is modelled as a physical process such as elastic deformation or fluid flow. It is also possible to mix rigid and non-rigid transformations in the same image (Little *et al* 1997).

The transformation T determined by a non-affine registration algorithm can be regarded as a deformation field that records the displacement vector at each voxel in image B needed to align it with the corresponding location in image A . For some non-affine registration applications, T is the most useful outcome of the registration. In a recent study of brain development, serially acquired MR images of children were registered using a non-affine algorithm in order to gain a better understanding of the rate of growth of different structures in the brain (Thompson *et al* 2000). Also, T can be used to propagate segmented structures between images that have been registered using a non-affine algorithm, for example to measure change in volume of structures over time (Dawant *et al* 2000).

10.4.1. Thin plate spline warps from point landmarks. The point registration technique described in section 5.1 above can be used to determine the rigid body or affine mapping that aligns the points P and Q in a least squares sense. If the structures being aligned are deformable, then it may be more appropriate to warp one of the images so that the landmarks are aligned. In the absence of any other information (for example about the mechanical properties of the tissue), the most appropriate transformation might be the one that matches the landmarks exactly, and bends the rest of space as little as possible. This requires an appropriate definition of bending, or bending energy. Intuitively, the squared norm of the second derivative is a good choice of energy. This can also be justified more rigorously from plate theory (Marsden and Hughes 1994). Also, by analogy with differential geometry, the curvature is related to second derivatives of the metric tensor. In more than one dimension, this is to be interpreted as sum of squares of all second order derivatives of all components of the mapping T .

Variational problems of this type are usually solved by finding a corresponding PDE equation, whose solution are minimizers. Here the PDE operator is the Laplacian of the Laplacian ($\Delta\Delta$) (Harder and Desmarais 1972, Goshtasby 1988, Bookstein 1989). Solutions are built by superposition of fundamental solutions, called thin plate splines, again by analogy with plate theory. The reader might be more familiar with fundamental solutions of the normal Laplacian $\Delta = \sum_i \partial_{x_i}^2$. In both cases, it is important to notice that these solutions have a different form in different dimensions: in one dimension, the thin plate splines are cubic splines (Dryden and Mardia 1998), in higher dimensions they are functions of the distance to the landmarks. In two dimensions there is a logarithmic term

$$\mathbf{x}' = \mathbf{L}\mathbf{x} + \mathbf{t} + \sum_i^N c_i r_i^2 \ln r_i^2 \quad (25)$$

whereas in three dimensions it takes the simpler form

$$\mathbf{x}' = \mathbf{L}\mathbf{x} + \mathbf{t} + \sum_i^N c_i r_i \quad (26)$$

where \mathbf{L} is a linear transformation, \mathbf{t} is a translation, and r_i is the distance to a landmark c_i .

10.4.2. Non-affine registration by optimising a similarity measure. The similarity measures described in sections 6 and 7 can be used for affine or non-affine registration. When calculating non-affine transformations, it is normal to optimize a cost function that is the combination of a term due to the similarity measure, and a regularization term. The regularization term assigns high cost to undesirable transformations, such as high local stretching or bending, or folding of the transformation (Rueckert *et al* 1999, Ashburner and Friston 1999). In this section we describe some examples of non-affine registration algorithms.

The approach used in the AIR software package (Woods *et al* 1998) is to initially calculate an affine transformation, then refine this to calculate a non-affine transformation in which the position of each voxel in image B^T is a polynomial expansion of its location in image B . For a second-order polynomial, there are 10 degrees of freedom for each axis (30 in total), rising to 60 degrees of freedom for third order and 105 degrees of freedom for fifth order.

An alternative to using polynomials is to break the image up into regions that are able to move independently of one another. Following an approximate alignment of images A and B using a rigid or affine transformation, the images can be split up into an array of nodes. These nodes may be the control points of a spline such as a B spline (Rueckert *et al* 1999), or may be regarded as the centre of a local region of interest (ROI). Each node in

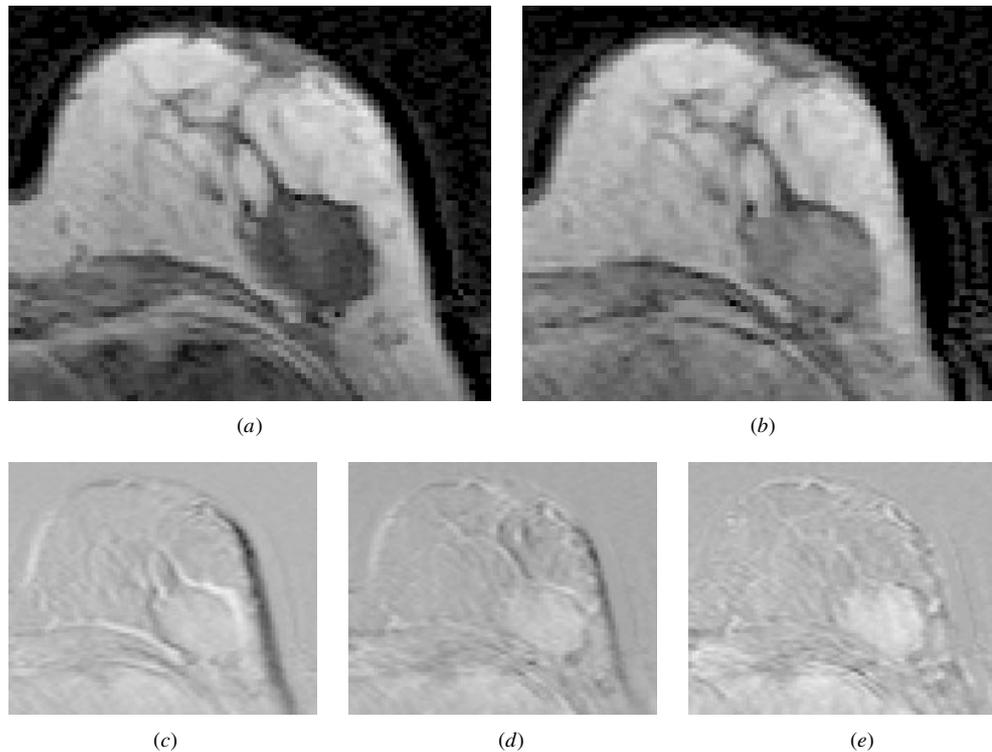


Figure 5. Example slice from (a) pre-contrast and (b) post-contrast MR mammogram. Without registration, the difference image contains distracting artefacts (c). Affine registration results in some improvement (d). Non-rigid registration (e), using a 10 mm grid of B-spline control points (Rueckert *et al* 1999) results in reduced artefact and improved diagnostic value (Denton *et al* 2000).

image B is iteratively translated, or a local affine transformation is iteratively determined for each ROI, in order to optimize the chosen cost function. The similarity measures SSD and CC can be evaluated locally on each ROI. The information theoretic measures such as MI and NMI, however, are statistical in nature, and the estimate determined from a small number of voxels is poor. In this case, the similarity measure should be calculated from the entire image for each node displacement. B-spline interpolation between nodes has the useful characteristic that they have ‘local support’, which means that only voxels within a radius of about two node spacings are affected by translation of that node, so when optimizing information theoretic measures on an array of B-splines, only a small portion of the image needs to be transformed with each iteration, even though the similarity measure is evaluated over the whole image.

Figure 5 shows an example non-affine registration of pre- and postcontrast MR mammograms obtained by optimizing normalized mutual information (NMI) on a 10 mm array of B-spline control points (Rueckert *et al* 1999).

For some regions of the body, it is important to distinguish between different tissues that have very different mechanical properties, such as bone and soft tissue. The algorithm of Little *et al* (1997) provides a technique to treat multiple rigid bodies within a deformable matrix.

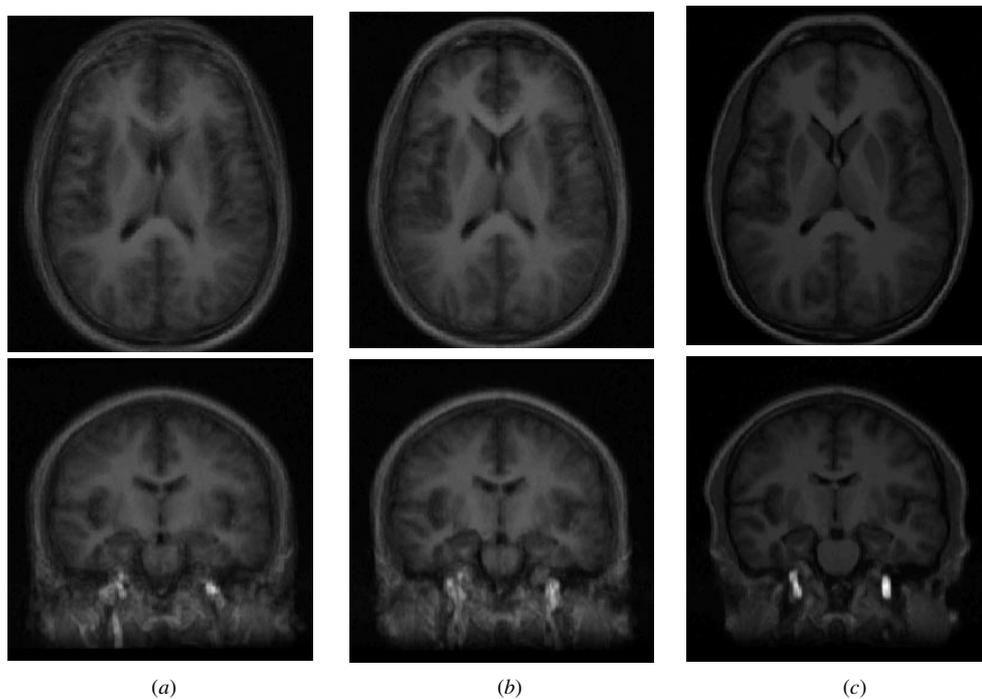


Figure 6. Axial (top) and coronal (bottom) slices from average images produced from MR scans of seven different normal controls, registered using (a) rigid registration, (b) affine registration and (c) non-rigid registration using a 10 mm grid of B-spline control points (Rueckert *et al* 1999). In all cases the registration was achieved by maximizing normalized mutual information, as described in section 7.3.3. Note that after rigid registration the average image is quite blurred, indicating that rigid registration does not line up the different brain scans very well. After affine registration, the images are a lot less blurred, especially around basal structures. The non-rigid registration, however, produces the sharpest image, indicating that this sort of algorithm is better at lining up brain features between subjects.

10.4.3. Intersubject registration. In neuroscience research it is standard practice to register images from different individuals into a standard space, in order to study variability between subjects, or to compare normal subjects with volunteers. This type of registration is frequently called *intersubject normalization*. The most common space to align images is in the Talairach stereotactic space. Aligning images with this space involves registering them with an electronic version of the Talairach atlas (Talairach and Tournoux 1988), generated from one post-mortem female brain. The alignment is most commonly carried out with an affine transformation (e.g. Collins *et al* 1994), but more degrees of freedom are sometimes used (Friston *et al* 1995, Ashburner and Friston 1999, Collins *et al* 1995, Woods *et al* 1998, Christensen *et al* 1996). This is normally treated as an intramodality registration problem, as it is normal to register the images of a particular subject to the images of another subject of the same modality whose images have already been put into Talairach space. The approaches described in section 10.4.2 above are, therefore, appropriate for this application.

Figure 6 compares the mean image obtained after registering seven MR images of normal subjects using a rigid body, affine and non-affine transformation by optimizing normalized mutual information using the algorithm of Rueckert *et al* (1999). Note the increasing sharpness of the average image as the number of degrees of freedom increases.

10.5. Image-to-physical space registration

Image-to-physical space registration is important when it is desirable to align images with a patient in the treatment position. The main applications are in image-guided neurosurgery and radiation therapy. The earliest approach was to use a stereotactic frame, rigidly attached to the patient. This frame could be fitted with imaging markers visible in MR, CT or x-ray angiography, and these markers could be used to relate image coordinates with treatment coordinates defined by the frame (Peters *et al* 1986, Kelly 1991). More recently, frameless techniques have been devised. These use point landmarks or surface features for registration. For neurosurgical applications, markers rigidly attached to the skull are quite invasive, but can provide high accuracy (Maurer *et al* 1997, Edwards *et al* 2000). Skin-affixed markers have less accuracy, but are also much less invasive (Roberts *et al* 1986). Bite blocks or dental stents (Howard *et al* 1995, Hauser *et al* 1997, Fenlon *et al* 2000) rigidly attached to the teeth provide a minimally invasive and relocatable system for registration of images to physical space. The patient is usually scanned wearing the stent with the imaging markers attached. During the intervention, LEDs can be attached to the stent to track head movement. These devices have been shown to provide relocation accuracy of better than 2 mm and tracking accuracy of less than 1 mm if carefully constructed (Fenlon 2000). An even less invasive approach is to use surface anatomical points, but this is very user dependent. Whichever points are used, registration can be achieved using the point registration algorithm described in section 5.1.

An alternative approach is to use surfaces for registration. The skin or exposed bone surface can be recorded in physical space using optical methods (Colchester *et al* 1996, Grimson *et al* 1996) or by manual tracking of a localizer over the skin surface (Tan *et al* 1993). These techniques do not require any markers to be attached to the patient prior to imaging, though points can also be used to improve robustness (Maurer *et al* 1998). In these cases, registration is achieved using one of the surface matching algorithms described in section 5.2.

11. Quantifying registration accuracy

There are two main reasons why it is desirable to quantify registration accuracy. The first is to calculate the expected accuracy of an algorithm in order to ascertain whether it is good enough for a particular clinical application, or in order to compare one algorithm with another. The second reason is to assess the accuracy of registration for a particular subject, for example prior to using the registered images to make a decision about patient management.

The accuracy of a registration transformation T cannot easily be summarized by a single number, as it is spatially varying over the image. If T is the calculated transformation, and T_g is the true ‘gold standard’ transformation, then the registration error will vary with position x_A in the image. If we think of each point in the image as a potential target of some treatment, then we can define the error at this point as the target registration error (TRE) as follows:

$$\text{TRE}(x_A) = |T(x_A) - T_g(x_A)|. \quad (27)$$

The TRE will normally vary with position. For example, in a rigid body transformation there will typically be rotational components. It may be that at some position in the image, by chance, the rotational component of the transformation cancels out the error in the translation component, giving $\text{TRE} = 0$. Elsewhere, however, the TRE will be greater. In the extremely unlikely case that the only error in a transformation is a translation error, then TRE will be the same everywhere. If T_g is known, then TRE can be calculated everywhere in the image. In this case, an image of TRE could be produced. In practice, it is more common to summarize the TRE distribution for example by considering the mean or maximum value. For most practical situations, however, T_g is not known, so TRE cannot be calculated.

11.1. Point landmark registration

11.1.1. Errors in rigid body point registration. For rigid body registration using fiducial points, a theory of errors has been devised that can predict the expected error distribution based on errors in localizing individual points, and the distributions of points.

It is clear that the localization of the fiducial points is never perfect. Fitzpatrick *et al* (1998b) call this error the fiducial localization error (FLE). The residual error in fitting of point sets is called the fiducial registration error (FRE). The distribution of FRE given the distribution of FLE has been described by Sibson (1978). Fitzpatrick *et al* (1998b) stress that what really matters is not the FRE, as the fiducials are not located at the places of interest in the subject, but the target registration error (TRE) we introduced earlier, i.e. the error induced by FLE at a given target. Explicitly, if the mean FLE is ε , the rotation and translations $\mathbf{T}_\varepsilon := (\mathbf{R}_\varepsilon, \mathbf{t}_\varepsilon)$ which solve the Procrustes problem are going to contain an error. The target registration error (TRE) at the target \mathbf{x} is then $|\mathbf{T}_\varepsilon(\mathbf{x}) - \mathbf{T}(\mathbf{x})|$. Fitzpatrick *et al* (1998b) provide formulae in first order in ε for TRE as a function of FLE, explaining a decrease in $1/\sqrt{N}$, where N is the number of points. Summarized: the squared expectation value of TRE at position \mathbf{x} , (coordinates (x_1, \dots, x_K)) is going to be to first order

$$\langle \text{TRE}(\mathbf{x})^2 \rangle \approx \langle \text{FLE} \rangle^2 \left(\frac{1}{N} + \frac{1}{K} \sum_i^K \sum_{j \neq i}^K \frac{x_i^2}{\Lambda_i^2 + \Lambda_j^2} \right)$$

where K is the number of spatial dimensions (normally three for medical applications) and Λ_i are the singular values of the configuration matrix of the markers.

11.2. Determining registration accuracy using a gold standard

If images are available for which the correct registration transformation is known, then registration accuracy can be calculated by comparing the transformation calculated by any registration algorithm with the known, 'gold standard' solution.

In the case of simulated data, the gold standard can have arbitrary accuracy, but the images are often not very realistic. An alternative is to use a gold standard method to register the images very accurately, and then compare any other algorithm with the solution obtained with the gold standard algorithm. This is a satisfactory approach if a gold standard algorithm can be applied to a group of test images, but it would not be appropriate for other registration tasks. For intermodality registration, this can be achieved by using invasive markers to provide the gold standard. These invasive markers can be attached to cadavers (Hemler *et al* 1995), or to surgical patients (West *et al* 1997), and the markers removed from the images prior to using these images as test data for algorithms that could be used when no invasive markers are available. Because the invasive markers can determine the transformation with a TRE at locations of interest of about 0.5 mm (Maurer *et al* 1997), the resulting images are useful for testing other algorithms. This approach has so far only been shown to be effective for rigid body intermodality registration. For non-rigid registration, or intramodality registration when much higher accuracy is required, the approach is not satisfactory.

11.3. Determining registration accuracy using consistency measurements

Because there is no realistic gold standard for intermodality registration, authors frequently test their algorithms by measuring the consistency of transformations (Freeborough *et al* 1996, Woods *et al* 1998, Holden *et al* 2000). Given three images of the same subject, A , B and C , there are three transformations that can be compared $\mathbf{T}_{A \rightarrow B}$, $\mathbf{T}_{B \rightarrow C}$ and $\mathbf{T}_{C \rightarrow A}$. Applying all

three transformations in turn completes a circuit, and should give the identity transformation for a perfect algorithm.

$$\mathbf{T}_c = \mathbf{T}_{A \rightarrow B} \mathbf{T}_{B \rightarrow C} \mathbf{T}_{C \rightarrow A}.$$

For any real algorithm, of course, \mathbf{T}_c will not be the identity. This will give a transformation that provides an estimate of the errors. If the registration errors in the process are uncorrelated, then the RMS registration error for one application of the algorithm will be $\frac{1}{\sqrt{3}}$ times the error for the whole circuit. Because one image is common in each of $\mathbf{T}_{A \rightarrow B}$, $\mathbf{T}_{B \rightarrow C}$ and $\mathbf{T}_{C \rightarrow A}$, however, the errors are not uncorrelated, so the errors estimated in this way will tend to underestimate the true error of the algorithm. As an extreme example, an algorithm that always produces an erroneous transformation close to the identity would incorrectly be found to perform well according to this measure.

11.4. Visual assessment

In many situations, the only practical means of estimating registration accuracy for a single patient is to visually inspect the images. An observer looks at the registered images, using various tools such as colour overlays and difference images, and classifies the registration solution as either a ‘success’ or a ‘failure’ depending on whether they judge the registration accuracy to be above or below the required application accuracy. This is very effective if an algorithm normally registers images well within the required accuracy, but occasionally fails in an obvious way. For algorithms that produce a fairly uniform range of errors on either side of the required accuracy, however, there is a risk that an observer will generate too many false negatives (images that are sufficiently well registered but are classified as failures), or false positives (images that are not well enough registered, but are classified as successes). This approach has been carefully studied for rigid body registration of MR and CT images (Fitzpatrick *et al* 1998a). Preliminary work in serial MR registration has suggested that observers have high sensitivity to errors greater than 0.2 mm when viewing difference images (Holden *et al* 2000).

12. Conclusions

In this review we have introduced the topic of medical image registration and discussed the main approaches described in the literature. The main emphasis of this article is intrasubject registration of tomographic modalities, which is predominantly used to find the rigid-body or affine transformation needed to align images of the head. We have also considered in less detail the closely related topics of non-affine registration (for intrasubject registration of deformable regions, and intersubject registration) 2D–3D registration, and image-to-physical space registration. Non-affine registration, in particular, is a rapidly developing area with many potential applications in healthcare and medical research.

Because most current algorithms for medical image registration calculate a rigid body or affine transformation, their applicability is restricted to parts of the body where tissue deformation is small compared with the desired registration accuracy, and in practice they are used most commonly for registration of images of the head. The most accurate algorithms for intermodality registration of the head are based on optimizing a voxel similarity measure. The most generally applicable of these algorithms are currently the ones based on information theory. These algorithms can be applied automatically to a variety of modality combinations for intermodality and intramodality registration, without the need for presegmentation of the images. They can also be extended to non-affine transformations.

One further appeal of these information theoretic approaches is the mystique that surrounds the word entropy. An interesting anecdote to emphasize this point comes from a conversation between Shannon and Von Neumann (quoted in Applebaum (1996)). Apparently, Shannon had asked Von Neumann which name he should give to his measure of uncertainty. Von Neumann answered: 'You should call it "entropy", and for two reasons: first, the function is already in use in thermodynamics under that name; second, and more importantly, most people don't know what entropy really is, and if you use the word 'entropy' in an argument, you will win every time!' There is, as yet, no proof that the information theory measures are in any way optimal for image registration, and better measures are likely to be devised in due course.

The first algorithms for medical image registration were devised in the early 1980s, and fully automatic algorithms have been available for many intermodality and intramodality applications since the mid 1990s. Despite this, at the time of writing, image registration is still seldom carried out on a routine clinical basis. The most widely used registration applications are probably image-to-physical space registration in neurosurgery and registration of functional MR images to correct for interscan patient motion. Intermodality registration, which accounts for the majority of the literature in this area, is still unusual in the clinical setting.

Image registration is, however, being widely used in medical research, especially in neuroscience where it is used in functional studies, in cohort studies and to quantify changes in structure during development and ageing.

One barrier to the routine clinical use of image registration may be the logistical difficulties in getting the images to be registered onto the same computer. Medical research labs doing a lot of imaging tend to have a more integrated infrastructure than hospitals, so do not suffer from this problem. The healthcare sector is now moving towards integrating text-based and image information about the patient to produce multimedia electronic patient records, and when this infrastructure is in place, the logistics of image registration will be much easier. Another reason for the lack of clinical use of image registration might be that traditional radiological practice can provide all the necessary information for patient management, and registration is unnecessary. Even if this second argument is currently valid, the increasing data generated by successive generations of scanners (including new multislice helical CT scanners) will steadily increase the need for registration to assist the radiologist carry out his or her task.

It is worthwhile briefly considering how image registration is likely to evolve over the next few years. Increasing volumes of data and multimedia electronic patient records have already been referred to, and these practical developments may see registration entering routine clinical use at many centres. Also, increasing use of dynamic acquisitions such as perfusion MRI will necessitate use of registration algorithms to correct for patient motion. In addition, non-affine registration is likely to find increasing application in the study of development, ageing and monitoring changes due to disease progression and response to treatment. In these latter applications, the transformation itself may have more clinical benefit than the transformed images, as this will quantify the changes in structure in a given patient. New developments in imaging technology may open up new applications of image registration. It has recently been shown that very high field whole-body MR scanners can produce high signal to noise ratio images of the brain with 100 μm resolution (Robitaille *et al* 2000). Intramodality registration of these images may open up new applications such as monitoring change in small blood vessels. Also, while ultrasound images have been largely ignored by image registration researchers up until now, the increasing quality of ultrasound images and its low cost makes this a fertile area for both intramodality and intermodality applications (e.g. Roche *et al* 2000, King *et al* 2000).

In some ways, medical image registration is a mature technology that has been around for nearly two decades and has attracted considerable research activity in devising and validating

algorithms, and in demonstrating clinical application. We believe, however, that there will be substantial additional innovation in this area over the next few years, especially in non-affine registration, and applications outside the brain. In the next decade, registration algorithms are likely to enter routine clinical use, and research applications will play a key role in improving our understanding of physiology and disease processes, especially in neuroscience.

Acknowledgments

We are grateful to colleagues in the Computational Imaging Science group at King's College London for useful discussions and assistance, especially Dr Julia Schnabel for figure 6 and Christine Tanner for figure 5.

References

- Applebaum D 1996 *Probability and Information: An Integrated Approach* (Cambridge: Cambridge University Press)
- Ashburner J and Friston K J 1999 Nonlinear spatial normalization using basis functions *Human Brain Mapping* **7** 254–66
- Atkinson D, Hill D L G, Stoye P N R, Summers P E, Clare S, Bowtell R and Keevil S F 1999 Automatic compensation of motion artefacts in MRI *Magn. Reson. Med.* **41** 163–70
- Atkinson D, Porter D A, Hill D L G, Calamante F and Connelly A 2000 Sampling and reconstruction effects due to motion in diffusion-weighted interleaved echo planar imaging *Magn. Reson. Med.* **44** 101–9
- Besl P J and McKay N D 1992 A method for registration of 3-D shapes *IEEE Trans. Pattern Anal. Mach. Intell.* **14** 239–56
- Bookstein F L 1989 Principal warps: thin-plate splines and the decomposition of deformations *IEEE Trans. Pattern Anal. Mach. Intell.* **11** 567–85
- Borgefors G 1984 Distance transformations in arbitrary dimensions *Comput. Vision Graph. Image Process.* **27** 321–45
- Buzug T M and Weese J 1998 Image registration for DSA quality enhancement *Computerized Imaging Graphics* **22** 103
- Chang H and Fitzpatrick J M 1992 A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities *IEEE Trans. Med. Imaging* **11** 319–29
- Chen G T Y, Kessler M and Pitluck S 1985 Structure transfer between sets of three dimensional medical imaging data *Computer Graphics 1985* (Dallas: National Computer Graphics Association) pp 171–5
- Christensen G E, Rabbitt R D and Miller M I 1996 Deformable templates using large deformation kinematics *IEEE Trans. Med. Imaging* **5** 1435–47
- Clarkson M J, Rueckert D, Hill D L G and Hawkes D J 1999 Registration of multiple video images to pre-operative CT for image guided surgery *Proc. SPIE* **3661** 14–23
- 2000 A multiple 2D video–3D medical image registration algorithm *Proc. SPIE* **3979** 342–52
- Colchester A C F, Zhao J, Holton-Tainter K S, Henri C J, Maitland N, Roberts P T E, Harris C G and Evans R J 1996 Development and preliminary evaluation of VISLAN, a surgical planning and guidance system using intra-operative video imaging *Med. Image Anal.* **1** 73–90
- Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P and Marchal G 1995 Automated multi-modality image registration based on information theory *Information Processing in Medical Imaging 1995* ed Y Bizais, C Barillot and R Di Paola (Dordrecht: Kluwer Academic) pp 263–74
- Collins D L, Holmes C J, Peters T M and Evans A C 1995 Automatic 3D model-based neuroanatomical segmentation *Human Brain Mapping* **3** 190–208
- Collins D L, Neelin P, Peters T M and Evans A C 1994 Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space *J. Comput. Assist. Tomogr.* **18** 192–205
- Cox R W and Jesmanowicz A 1999 Real-time 3D image registration for functional MRI *Magn. Reson. Med.* **42** 1014–18
- Cuchet E, Knoploch J, Dormont D and Marsault C 1995 Registration in neurosurgery and neuroradiotherapy applications *J. Image Guided Surg.* **1** 198–207
- Dawant B M, Hartmann S L, Thirion J-P, Maes F, Vandermeulen D and Demaerel P 2000 Automatic 3D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: part 1. Methodology and validation on normal subjects *IEEE Trans. Med. Imaging* **18** 909–16

- Declerck J, Feldmar J, Goris M L and Betting F 1997 Automatic registration and alignment on a template of cardiac stress and rest reoriented SPECT images *IEEE Trans. Med. Imaging* **16** 727–37
- Denton E R E, Holden M, Christ E, Jarosz J M, Russell-Jones D, Goodey J, Cox T C S and Hill D L G 2000 The identification of cerebral volume changes in treated growth hormone deficient adults using serial 3D MR image processing *J. Comput. Assist. Tomogr.* **24** 139–45
- Dryden I and Mardia K 1998 *Statistical Shape Analysis* (New York: Wiley)
- Eddy W F, Fitzgerald M and Noll D C 1996 Improved image registration by using Fourier interpolation *Magn. Reson. Med.* **36** 923–31
- Edelman A, Arias T and Smith S 1998 The geometry of algorithms with orthogonality constraints *SIAM J. Matrix Anal. Appl.* **20** 303–53
- Edwards P J *et al* 2000 Design and evaluation of a system for microscope-assisted guided interventions (MAGI) *IEEE Trans. Med. Imaging* **19** at press
- Evans A C, Beil C, Marrett S, Thompson C J and Hakim A 1988 Anatomical–functional correlation using an adjustable MRI-based region of interest atlas with positron emission tomography *J. Cereb. Blood Flow Metab.* **8** 513–30
- Faber T L, McColl R W, Opperman R M, Corbett J R and Peshock R M 1991 Spatial and temporal registration of cardiac SPECT and MR images: methods and evaluation *Radiology* **179** 857–61
- Faber T L and Stokely E M 1988 Orientation of 3-D structures in medical images *IEEE Trans. Pattern Anal. Mach. Intell.* **10** 626–33
- Fenlon M R, Jusczyzck A S, Edwards P J and King A P 2000 Locking acrylic resin dental stent for image guided surgery *J. Prosthetic Dentistry* **83** 482–5
- Fitzpatrick J M, Hill D L G, Shyr Y, West J, Studholme C and Maurer C R Jr 1998a Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain *IEEE Trans. Med. Imaging* **17** 571–85
- Fitzpatrick J, West J and Maurer C R Jr 1998b Predicting error in rigid-body, point-based registration *IEEE Trans. Medical Imaging* **17** 694–702
- Freeborough P A, Woods R P and Fox N C 1996 Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders *J. Comput. Assist. Tomogr.* **20** 1012–22
- Friston K J, Ashburner J, Poline J B, Frith C D, Heather J D and Frackowiak R 1995 Spatial registration and normalisation of images *Human Brain Mapping* **2** 165–89
- Gelfand M, Mironov A and Pevzner P 1996 Gene recognition via spliced sequence alignment *Proc. Natl. Acad. Sci. USA* **93** 9061–6
- Golub G H and van Loan C F 1996 *Matrix Computations* 3rd edn (Baltimore, MD: Johns Hopkins University Press)
- Goshtasby A 1988 Registration of images with geometric distortions *IEEE Trans. Geosci. Remote Sensing* **26** 60–4
- Green B F 1952 The orthogonal approximation of an oblique structure in factor analysis *Psychometrika* **17** 429–40
- Grimson W E L, Ettinger G J, White S J, Lozano-Perez T, Wells W M III and Kikinis R 1996 An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization *IEEE Trans. Med. Imaging* **15** 129–40
- Guezic A and Ayache N 1994 Smoothing and matching of 3-D space curves *Int. J. Comput. Vision* **12** 79–104
- Guezic A, Pennec X and Ayache N 1997 Medical image registration using geometric hashing *IEEE Comput. Sci. Eng.* **4** 29–41
- Guillemaud R and Brady M 1997 Estimating the bias field of MR images *IEEE Trans. Med. Imaging* **16** 238–51
- Hajnal J B, Saeed N, Oatridge A, Williams E J, Young I R and Bydder G M 1995b Detection of subtle brain changes using subvoxel registration and subtraction of serial MR images *J. Comput. Assist. Tomogr.* **19** 677–91
- Hajnal J B, Saeed N, Soar E J, Oatridge A, Young I R and Bydder G M 1995 A registration and interpolation procedure for subvoxel matching of serially acquired MR images *J. Comput. Assist. Tomogr.* **19** 289–96
- Harder R L and Desmarais R N 1972 Interpolation using surface splines *J. Aircraft* **9** 189–91
- Hauser R, Westermann B and Probst R 1997 Non-invasive tracking of patients' head movements during computer-assisted intranasal microscopic surgery *Laryngoscope* **211** 491–9
- Hemler P F, van den Elsen P A, Sumanaweera T S, Napel S, Drace J and Adler J R 1995 A quantitative comparison of residual error for three different multimodality registration techniques *Information Processing in Medical Imaging 1995* ed Y Bizais, C Barillot and R Di Paola (Dordrecht: Kluwer Academic) pp 251–62
- Hill D L G, Hawkes D J, Crossman J E, Gleeson M J, Cox T C S, Bracey E E C M L, Strong A J and Graves P 1991 Registration of MR and CT images for skull base surgery using point-like anatomical features *Br. J. Radiol.* **64** 1030–5
- Hill D L G, Maurer C R Jr, Studholme C, Fitzpatrick J M and Hawkes D J 1998 Correcting scaling errors in tomographic images using a nine degree of freedom registration algorithm *J. Comput. Assist. Tomogr.* **22** 317–23
- Hill D L G, Studholme C and Hawkes D J 1994 Voxel similarity measures for automated image registration *Proc. SPIE* **2359** 205–16

- Holden M, Hill D L G, Denton E R E, Jarosz J M, Cox T C S, Goodey J, Rohlfing T and Hawkes D J 2000 Voxel similarity measures for 3D serial MR image registration *IEEE Trans. Med. Imaging* **19** 94–102
- Howard M A, Dobbs M B, Siminson T M, LaVelle W E and Granner M A 1995 A non-invasive reattachable skull fiducial marker system *J. Neurosurg.* **83** 372–6
- Huang C T and Mitchell O R 1994 A Euclidean distance transform using grayscale morphology decomposition *IEEE Trans. Pattern Anal. Mach. Intell.* **16** 443–8
- Hurley J R and Cattell R B 1962 The Procrustes program: producing direct rotation to test a hypothesized factor structure *Behav. Sci.* **7** 258–62
- Jezzard P and Balaban R S 1995 Correction for geometric distortion in echo planar images from B_0 field variations *Magn. Reson. Med.* **34** 65–73
- Jiang H, Robb R A and Holton K S 1992 A new approach to 3-D registration of multimodality medical images by surface matching *Proc. SPIE* **1808** 196–213
- Kanatani K 1994 Analysis of 3-D rotation fitting *IEEE Trans. Pattern Anal. Mach. Intell.* **16** 543–9
- Kelly P J 1991 *Tumor Stereotaxis* (Philadelphia: Saunders)
- King A P, Blackall J M, Penny G P, Edwards P J, Hill D L G and Hawkes D J 2000 Bayesian estimation of intra-operative deformation for image-guided surgery using 3D ultrasound *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2000 (Lecture Notes in Computer Science 1935)* (Berlin: Springer) pp 588–97
- Koschat M A and Swayne D F 1991 A weighted Procrustes criterion *Psychometrika* **56** 229–39
- Lavallée S and Szeliski R 1995 Recovering the position and orientation of free-form objects from image contours using 3-D distance maps *IEEE Trans. Pattern Anal. Mach. Intell.* **17** 378–90
- Lehmann T M, Gonner C and Spitzer K 1999 Survey: interpolation methods in medical image processing *IEEE Trans. Med. Imaging* **18** 1049–75
- Lemieux L and Barker G J 1998 Measurement of small inter-scan fluctuations in voxel dimensions in magnetic resonance images using registration *Med. Phys.* **25** 1049–54
- Lemieux L, Jagoe R, Fish D R, Kitchen N D and Thomas D G 1994 A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs *Med. Phys.* **21** 1749–60
- Lemieux L, Kitchen N D, Hughes S W and Thomas D G T 1994 Voxel-based localization in frame-based and frameless stereotaxy and its accuracy *Med. Phys.* **21** 1301–10
- Lemieux L, Wiesmann U C, Moran N F, Fish D R and Shorvon S D 1998 The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization *Med. Image Anal.* **2** 227–42
- Levin D N, Pelizzari C A, Chen G T Y, Chen C-T and Cooper M D 1988 Retrospective geometric correlation of MR, CT, and PET images *Radiology* **169** 817–23
- Little J A, Hill D L G and Hawkes D J 1997 Deformations incorporating rigid structures *Comput. Vision Image Understanding* **66** 223–32
- Maes F, Collignon A, Vandermeulen D, Marchal G and Suetens P 1997 Multimodality image registration by maximization of mutual information *IEEE Trans. Med. Imaging* **16** 187–98
- Maguire G Q Jr, Noz M E, Lee E M and Schimpf J H 1986 Correlation methods for tomographic images using two and three dimensional techniques *Information Processing in Medical Imaging 1985* ed S L Bacharach (Dordrecht: Martinus Nijhoff) pp 266–79
- Maintz J B A, van den Elsen P A and Viergever M A 1996 Evaluation of ridge seeking operators for multimodality medical image matching *IEEE Trans. Pattern Anal. Mach. Intell.* **18** 353–65
- Maintz J B A and Viergever M A 1998 A survey of medical image registration *Med. Image Anal.* **2** 1–36
- Marsden J and Hughes T 1994 *Mathematical Foundations of Elasticity* (New York: Dover)
- Maurer C R Jr and Fitzpatrick J M 1993 A review of medical image registration *Interactive Image-Guided Neurosurgery* ed R J Maciunas (Park Ridge, IL: American Association of Neurological Surgeons) pp 17–44
- Maurer C R Jr, Fitzpatrick J M, Wang M Y, Galloway R L Jr, Maciunas R J and Allen G S 1997 Registration of head volume images using implantable fiducial markers *IEEE Trans. Med. Imaging* **16** 447–62
- Maurer C R Jr, Maciunas R J and Fitzpatrick J M 1998 Registration of head CT images to physical space using a weighted combination of points and surfaces *IEEE Trans. Med. Imaging* **17** 753–61
- McGee K P, Felmlee J P, Manduca A, Riederer S J and Ehman R L 2000 Rapid autocorrection using prescan navigator echoes *Magn. Reson. Med.* **43** 583–8
- Meltzer C C, Bryan R N, Holcomb H H, Kimball A W, Mayberg H S, Sadzot B, Leal J P, Wagner H N Jr and Frost J J 1990 Anatomical localization for PET using MR imaging *J. Comput. Assist. Tomogr.* **14** 418–26
- Monga O and Benayoun S 1995 Using partial derivatives of 3D images to extract typical surface features *Comput. Vision Image Understanding* **61** 171–89
- Parker J, Kenyon R V and Troxel D 1983 Comparison of interpolating methods for image resampling *IEEE Trans. Med. Imaging* **2** 31–9

- Pelizzari C A, Chen G T Y, Spelbring D R, Weichselbaum R R and Chen C-T 1989 Accurate three-dimensional registration of CT, PET, and/or MR images of the brain *J. Comput. Assist. Tomogr.* **13** 20–6
- Pennec X and Thirion J-P 1997 A framework for uncertainty and validation of 3-D registration methods based on points and frames *Int. J. Comput. Vision* **25** 203–29
- Penney G, Weese J, Little J A, Desmedt P, Hill D L G and Hawkes D J 1998 A comparison of similarity measures for use in 2D–3D medical image registration *IEEE Trans. Med. Imaging* **17** 586–95
- Peters T M, Clark J A, Olivier A, Marchand E P, Mawko G, Dieumegarde M, Muresan L and Ethier R 1986 Integrated stereotaxic imaging with CT, MR imaging, and digital subtraction angiography *Radiology* **161** 821–6
- Pluim J, Maintz J B A and Viergever M A 2000 Interpolation artefacts in mutual information-based image registration *Computer Vision Image Understanding* **77** 211–32
- Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in C: The Art of Scientific Computing* 2nd edn (Cambridge: Cambridge University Press)
- Rao C 1980 Matrix approximations and reduction of dimensionality in multivariate statistical analysis *Multivariate Analysis* (Amsterdam: North Holland)
- Roberts D W, Strohbehn J W, Hatch J F, Murray W and Kettenberger H 1986 A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope *J. Neurosurg.* **65** 545–9
- Robitaille P M, Abduljalil A M and Kangarlu A 2000 Ultra high resolution imaging of the human head at 8 Tesla: 2K x 2K for Y2K *J. Comput. Assist. Tomogr.* **24** 2–8
- Robson M D, Anderson A W and Gore J C 1997 Diffusion-weighted multiple shot echo planar imaging of humans without navigation *Magn. Reson. Med.* **38** 82–8
- Roche A, Pennec X, Rudolph M, Auer D P, Malandain G, Ourselin S, Auer L M and Ayache N 2000 Generalised correlation ratio for rigid registration of 3D ultrasound with MR images *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2000 (Lecture Notes in Computer Science 1935)* (Berlin: Springer) pp 567–77
- Rueckert D, Sonoda L, Hayes C, Hill D, Leach M and Hawkes D 1999 Non-rigid registration using free-form deformations: application to breast MR images *IEEE Trans. Med. Imaging* **18** 712–21
- Schad L R, Boesecke R, Schlegel W, Hartmann G H, Sturm V, Strauss L G and Lorenz W J 1987 Three dimensional image correlation of CT, MR, and PET studies in radiotherapy treatment planning of brain tumors *J. Comput. Assist. Tomogr.* **11** 948–54
- Schönemann P H 1966 A generalized solution of the orthogonal Procrustes problem *Psychometrika* **31** 1–10
- Shannon C E 1948 The mathematical theory of communication (parts 1 and 2) *Bell Syst. Tech. J.* **27** 379–423, 623–56 (reprint available from <http://www.lucent.com>)
- 1949 Communication in the presence of noise *Proc. IRE* **37** 10–21 (reprinted in *Proc. IEEE* **86** 447–57)
- Sibson R 1978 Studies in the robustness of multidimensional scaling: Procrustes statistics *J. R. Statist. Soc. B* **40** 234–8
- Sled J, Zijdenbos A and Evans A 1998 A nonparametric method for automatic correction of intensity nonuniformity in MRI data *IEEE Trans. Med. Imaging* **17** 87–97
- Söderkvist I 1993 Perturbation analysis of the orthogonal procrustes problem II: Numerical analysis *BIT* **33** 687–95
- Stewart G 1993 On the early history of the singular value decomposition *SIAM Rev.* **35** 551–66
- Studholme C, Hill D L G and Hawkes D J 1995 Multiresolution voxel similarity measures for MR–PET registration *Information Processing in Medical Imaging 1995* ed Y Bizais, C Barillot and R Di Paola (Dordrecht: Kluwer Academic) pp 287–98
- 1996 Automated 3D registration of MR and CT images of the head *Med. Image Anal.* **1** 163–75
- 1997 Automated 3D registration of MR and PET brain images by multi-resolution optimization of voxel similarity measures *Med. Phys.* **24** 25–35
- 1999 An overlap invariant entropy measure of 3D medical image alignment *Pattern Recognition* **32** 71–86
- Sumanaweera T S, Glover G H, Song S M, Adler J R and Napel S 1994 Quantifying MRI geometric distortion in tissue *Magn. Reson. Med.* **31** 40–7
- Talairach J and Tournoux P 1988 *Co-Planar Stereotaxic Atlas of the Human Brain* (Stuttgart: George Thieme)
- Tan K K, Grzeszczuk R, Levin D N, Pelizzari C A, Chen G T, Erickson R K, Johnson D and Dohrmann G J 1993 A frameless stereotaxic approach to neurosurgical planning based on retrospective patient-image registration *J. Neurosurg.* **79** 296–303
- Thacker N A, Jackson A, Moriarty D and Vokurka E 1999 Improved quality of re-sliced MR images using re-normalized sinc interpolation *J. Magn. Reson. Imaging* **10** 582–8
- Thompson P M, Gledd J N, Woods R P, MacDonald D, Evans A C and Toga A W 2000 Growth patterns in the developing brain detected by using continuum mechanics tensor maps *Nature* **404** 190–3
- Umeyama S 1991 Least-squares estimation of transformation parameters between two point patterns *IEEE Trans. Pattern Anal. Mach. Intell.* **13** 376–80

- Unser M 1999 Splines: a perfect fit for signal and image processing *IEEE Signal Process. Mag.* **16** 22–38
- van den Elsen P A, Maintz J B A, Pol E-J D and Viergever M A 1995 Automatic registration of CT and MR brain images using correlation of geometrical features *IEEE Trans. Med. Imaging* **14** 384–96
- van den Elsen P A, Maintz J B A and Viergever M A 1992 Geometry driven multimodality image matching *Brain Topogr.* **5** 153–8
- van den Elsen P A, Pol E-J D, Sumanaweera T S, Hemler P F, Napel S and Adler J R 1994 Grey value correlation techniques used for automatic matching of CT and MR brain and spine images *Proc. SPIE* **2359** 227–37
- van den Elsen P A, Pol E-J D and Viergever M A 1993 Medical image matching: a review with classification *IEEE Eng. Med. Biol.* **12** 26–39
- Van Herk M and Kooy H M 1994 Automated three-dimensional correlation of CT–CT, CT–MRI and CT–SPECT using chamfer matching *Med. Phys.* **21** 1163–78
- Viola P A 1995 Alignment by maximization of mutual information *PhD Thesis* Massachusetts Institute of Technology
- Weese J, Penney G P, Desmedt P, Buzug T M, Hill D L G and Hawkes D J 1997 Voxel-based 2-D/3-D registration of fluoroscopy images and CT scans for image-guided surgery *IEEE Trans. Inform. Technol. Biomed.* **1** 284–93
- Wells W M III, Viola P, Atsumi H, Nakajima S and Kikinis R 1996 Multi-modal volume registration by maximization of mutual information *Med. Image Anal.* **1** 35–51
- West J B, Fitzpatrick J M, Wang M Y, Dawant B M, Maurer C R Jr, Kessler R M and Maciunas R J 1999 Retrospective intermodality techniques for images of the head: surface-based versus volume-based *IEEE Trans. Med. Imaging* **18** 144–50
- West J B *et al* 1997 Comparison and evaluation of retrospective intermodality brain image registration techniques *J. Comput. Assist. Tomogr.* **21** 554–66
- Whittaker J M 1935 Interpolatory function theory *Cambridge Tracts in Mathematics and Mathematical Physics* no 33 (Cambridge: Cambridge University Press) ch 4
- Woods R P, Cherry S R and Mazziotta J C 1992 Rapid automated algorithm for aligning and reslicing PET images *J. Comput. Assist. Tomogr.* **16** 620–33
- Woods R P, Grafton S T, Holmes C J, Cherry S R and Mazziotta J C 1998 Automated image registration: I. General methods and intrasubject, intramodality validation *J. Comput. Assist. Tomogr.* **22** 139–52
- Woods R P, Grafton S T, Watson J D G, Sicotte N L and Mazziotta J C 1998 Automated image registration: II. Intersubject validation of linear and nonlinear models *J. Comput. Assist. Tomogr.* **22** 153–65
- Woods R P, Mazziotta J C and Cherry S R 1993 MRI-PET registration with automated algorithm *J. Comput. Assist. Tomogr.* **17** 536–46