

Image Registration and Motion Correction

Mark Jenkinson

Centre for Functional MRI of the Brain (FMRIB), University of Oxford, UK

1 Introduction

Registration is a widely used and essential tool in imaging of all forms, and especially in medical imaging. A great deal of effort and research has gone into developing registration methods to align images of various types. This talk aims to introduce the basic concepts behind registration, particularly the fully automatic registration methods, as applied in medical imaging. Because of the great variety of approaches used, it is necessary to only concentrate on a few here, but many surveys, reviews and textbooks covering both broad [14, 15, 20, 23, 28, 37, 41] and more specialised [2, 9, 10, 16, 24, 25, 26, 31] topics in medical image registration exist for the interested reader.

The basic registration task is to take two images (either 2D or 3D images) and align them. That is, spatially reshape one to match the other. This is done by finding a relationship between the voxel locations in the two images. Once this relationship is known, the information at corresponding locations can be combined, allowing fusion or comparisons of the two images.

For example, the information from different images (such as T1-weighted and T2-weighted MR images) can be combined, allowing features visible in one modality to be seen on top of features visible in the other. Another example is in functional MRI where the functional, low-resolution activations can be related to the anatomical structures seen in a separate high-resolution structural image of the same subject.

Registration is also useful for comparing images of the same modality that are taken at different times. For instance, longitudinal studies of disease (e.g. Multiple Sclerosis, Alzheimer's, etc.), studies looking at anatomical variation between groups (e.g. Schizophrenia), motion correction of functional/dynamic images, and pre- and post-operative comparisons are all cases where registration is essential in the process of analysing the images to look for differences. In fact, the spatial relationships established by registration are often the quantities of most interest as they allow changes in shapes and volumes to be seen and quantified directly.

Finally, consider two areas where registration is particularly important and challenging. The first is the area of functional imaging (fMRI) where geometric and intensity distortions in the functional images and anatomical variability between subjects make both functional to structural and structural to standard-space registrations extremely difficult. The second challenging area is in surgical planning and assistance, where deformation of the patient's anatomy occurs and real-time updating is needed for surgeons in the operating room.

Registration is therefore an essential tool in a wide range of medical imaging scenarios. Some aspects of registration are relatively well understood and methods exist that perform very well (e.g. rigid-body registration of intra-modal brain images) whereas other aspects are only just becoming possible (e.g. real-time surgical assistance) or are currently evolving (e.g. inter-subject comparisons). An understanding of the fundamental principles of registration and what limitations exist in the methods used is important in order to utilise medical imaging in a careful and quantitative way, either in a research or a clinical setting.

2 Basic Registration Components

There are two main categories of registration method: feature-based and intensity-based. The feature-based methods use either manually or automatically extracted features (or landmarks) from the images and align these features. In contrast, the intensity-based methods use all the voxel intensity values to determine the alignment of the images. The intensity-based methods are more common at present, and it is these methods which will be described herein, although interesting developments in feature-based methods are still being made (a recent example is [33]).

For intensity-based methods, the vast majority are based on three simple components: specifying the allowable transformations; choosing a way of measuring the similarity between transformed images; and choosing a way to search for the transformation that gives the best similarity. An understanding of this material is important for people that use registration methods, as well as those that develop them, since many registration applications require the selection of the correct options within these components in order to work well. Therefore these components, and the options within them, are what is described in the next sections.

2.1 Transformations

A spatial transformation is what is applied to an image in order to change the position, orientation or shape of structures (such as the brain) in the image. Mathematically it is expressed as a set of equations relating the old image positions (coordinates) to the new ones. These equations need to be restricted in some way in order to limit the possible deformations of the images. It is important to understand these restricted models of transformations (e.g. rigid-body, affine, viscous-fluid) as they determine the physical model for the deformations – either due to changes in the anatomy or in the imaging process (e.g. geometric distortions). For instance, rigid-body transformations do not allow the size or shape of any structures to change and so are useful for intra-subject registrations but not for inter-subject registrations where size and shape are different.

The transformation model is often described by its Degrees Of Freedom (*DOF*), which is the number of independent ways that the transformation can be changed. For example, consider translations (shifts) in 3D. There are three independent translations (one in x , one in y and one in z) making this a 3 DOF transformation model.

In general, increasing the number of DOF allows the transformations greater scope to make one image match the other. The three most common models of transformations in 3D are: rigid-body (6 DOF), affine (12 DOF) and non-linear (anything from 12 to millions of DOF). These models are now described in more detail.

2.1.1 Rigid Body Transformations

The rigid-body transformation model only permits rotations and translations. In 3D it has 6 DOF: three rotations (one about each axis) and three translations. This fully describes the type of movements that a rigid-body (one that does not change shape) can undergo, and so it is a good model of how a rigid body-part (e.g. brain or pelvis) can move. See figure 1 for some examples of rigid-body transformations of a brain image.

This transformation model does not allow any structures within the image to change size or shape and therefore should only be used when this is known to be true, such as for images of the same subject where no anatomical changes are expected. Note that for articulated body-parts (e.g. the spine), a separate rigid-body transformation might be a good model for each part (e.g. vertebra), but not for the whole object.

2.1.2 Affine Transformations

The affine transformation model permits all linear coordinate changes: that is, rotations, translations, scalings and shears (or shears). In 3D this gives a total of 12 DOF: 3 translation, 3 rotation, 3 scaling and 3 skew parameters. See figure 1 for example transformations. Note that this includes, as a sub-class, all rigid-body transformations.

This transformation model allows both size and shape change for all structures in the image. However, these changes are global across the image and it does not, for example, allow one shape to increase in size while others stay the same size or decrease in size. Therefore this transformation model will not allow accurate registrations of images where, for example, only one has ventricular enlargement. The most common use of affine transformations is in registering images to an approximate common (or standard) space (e.g. Talairach space [35]) or for compensating for geometric image distortions, such as those caused by eddy currents in diffusion imaging [27].

2.1.3 Nonlinear Transformations

The non-linear transformation model includes all transformations that do not fit into the affine transformation model. Hence, it encompasses a very wide range of transformations from those that are nearly linear, with few DOF, to the most general transformations which have a separate displacement (3 DOF) for each voxel, giving well over a million DOF for typical images. In principle, these higher DOF transformations allow any geometric change between images to be modelled although typically some constraints are still applied to prevent non-physical transformations from occurring (such as changes in topology).

Many non-linear transformation models are based on various mathematical or physical models. These models generally control the DOF and the smoothness available to the transformation, and hence control how well they can align or reshape an image to match another. Some common examples of these models are:

Basis Functions: the transformation is made up of sums of simple functions (e.g. polynomials in AIR [40], or sinusoids in SPM [11]). The number of basis functions controls the DOF (usually in the hundreds) and the smoothness, since the lower-order basis functions are usually arranged to be the smoothest.

Splines: the transformation is built from simple piecewise terms and specified by values at knot or control points.

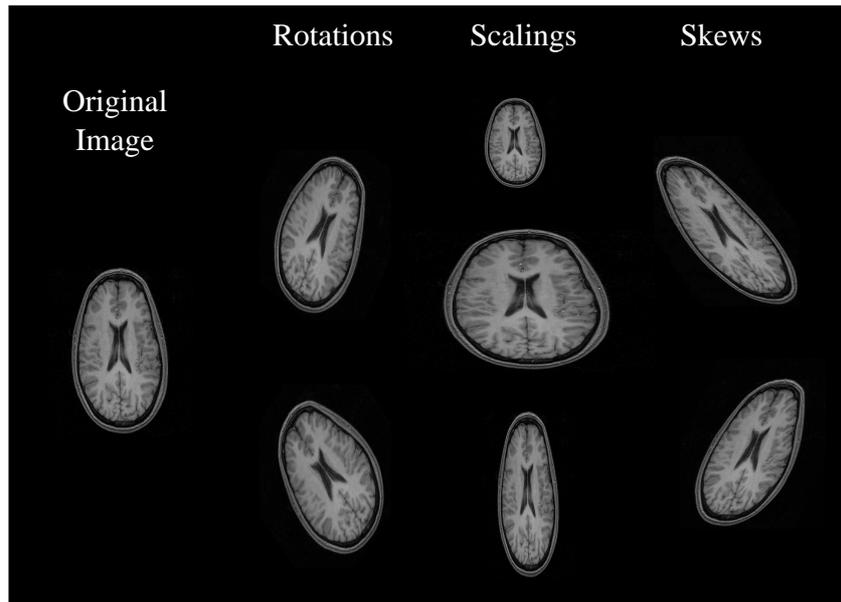


Figure 1: Various examples of affine transformations of an original image (left). Note that, for the affine case, these individual transformation types can be mixed in any order (e.g. scalings and rotations and skews). Rigid-body transformations are a sub-set of affine transformations and include rotations, but not scalings or skews.

The number of control points specifies the DOF (typically thousands) and the spacing effects the smoothness. Common examples include B-splines [32] and thin-plate splines [4].

Elastic-solid: the transformation is modelled as the deformation of an elastic solid [6, 21]. The material properties (e.g. Lamé coefficients) control the smoothness and the effective DOF (note that the effective DOF is less than the nominal DOF, especially if the smoothness is high which implies that the parameters are not independent). These models typically have very high nominal DOF (millions or more) and can model small changes very accurately, but may have problems with large deformations where the internal elastic energy pulls the image away from the best matching transformation. Note that bio-mechanical models [13], often used in surgical applications, are a prime example of this type of model, where the elastic energy also has a physical meaning.

Viscous-fluid: the transformation is modelled as the deformation of a viscous fluid which relaxes over time [7, 8]. The viscosity parameters control the speed of convergence but may not effect the final smoothness, depending on the number of iterations and the stopping criteria. These models have very high DOF (more than the effective DOF of most elastic-solid models) and low smoothness constraints, allowing them to fit to almost any possible image deformation.

Smoothed displacement fields: these model are more general than the others, often allowing 3 DOF per voxel (for a total of millions of nominal DOF). However, a smoothing parameter is used to control the smoothness directly and the effective DOF implicitly. One example in this category is the ‘Demons algorithm’ [36].

Several other categories exist and many more examples of methods in these categories can be found in any of the general references [14, 15, 20, 23, 28, 37, 41].

In general, non-linear transformations with very high DOF are useful for medium to high resolution images with good contrast and little or no artefacts, where the subtle changes in anatomy can be seen. For other situations where the internal details are indistinct or significant intensity changes are induced for other reasons (such as BOLD changes or artefacts) it is often more appropriate to use lower DOF non-linear or affine transformations, as these are less sensitive to poor image quality (including artefacts) and hence more robust. However, the final choice of transformation model must be made according to what specific acquisitions are available and the desired application, where other issues may also be important.

2.1.4 Interpolation

All images are acquired and stored as a collection of discrete points set on a spatial grid. However, in order to carry out the image transformations the intensity at positions between grid points needs to be calculated. This process is called interpolation.

As an illustration, consider interpolating a one-dimensional signal using a linear interpolation method. The aim is to calculate the intensity at any location between the sample points. (One reason to want to do this would be to effect a sub-voxel — i.e. very small — translation.) To do this, the two nearest samples are chosen, and a hypothetical line is constructed between the intensities at these points so that the intensity at any point in between can be found. This is illustrated in figure 2.

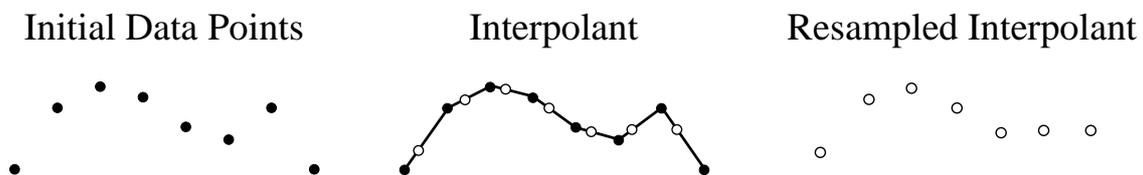


Figure 2: A one-dimensional linear interpolation. The left figure shows the original discrete samples. In the middle figure the original samples are connected with lines (hence linear interpolation) — forming the interpolant. This interpolant is then resampled at the new points (open circles), effectively performing a sub-voxel translation.

Several different methods of interpolation exist and each produces slightly different results. Some of the most common methods in three dimensions are: nearest neighbour (take the value of the nearest original neighbour), trilinear (the three-dimensional extension of the one-dimensional linear interpolation example given above), sinc (an interpolation function commonly applied in image processing) and spline-based (smooth forms made from simple piecewise segments). Note that both sinc and spline-based interpolations use information from outside the immediately neighbouring voxels. As a consequence they can include more information about the image structure and can produce sharper images, but take longer to calculate.

When the precise value of the intensity at a point in the image is important, as it is in motion correction of fMRI volumes, the choice of interpolation can be significant. However, for many registration applications it is sufficient that the interpolated intensity be similar to surrounding tissue, allowing for some small interpolation inaccuracies. Therefore, the choice of interpolation method is usually not critical as long as it is smooth or continuous. Hence, for registration, trilinear interpolation is commonly used, due to its simplicity and speed, while nearest-neighbour, because it is not smooth, is generally avoided. Furthermore, it is common to apply a more accurate interpolation method (e.g. spline) in the final image transformation (resampling) while using a faster method (e.g. trilinear) for the main registration calculations.

2.2 Similarity and Cost Functions

Once the transformation model and interpolation method have been selected, it is necessary to define a global measure of similarity between the images. That is, a function which will quantify how similar the images are *after* a spatial transformation has been applied. This is done by defining a *similarity function* where better-aligned images give larger values (alternatively, a *cost function* can be used, where better aligned images give smaller values). Given a similarity function, the registration problem is solved by systematically trying different transformations in turn, to find the one which gives the maximum similarity value.

It is not essential for users of registration methods to understand the precise details of the similarity function calculations. However, it is essential to understand which ones are appropriate in a given situation (e.g. that least squares, mean absolute difference, mean square difference and normalised correlation should *only* be used for *intra-modal* registrations). Therefore the details presented in the next sections can be skipped on a first reading, but note should be taken of the appropriate usage conditions that are required for each similarity function.

One of the greatest challenges for fully automatic registration methods is the definition of a sensitive but robust similarity function. The first consideration when choosing a similarity function is whether registration is between two images of the same type (intra-modal) or different types (inter-modal). With intra-modal registration, one can normally assume that the two images will look fairly similar after correct registration — maybe just a change in overall brightness and contrast is necessary. However, in the latter case, one cannot assume that a given tissue type will have the same image intensity in each image. For instance, consider registering a T1-weighted MR image with a T2-weighted image. In the T1-weighted image the CSF has a lower intensity than the white matter,

whereas in the T2-weighted image its intensity is higher. The unknown intensity transformation then becomes a confounding factor to be dealt with by the similarity function.

2.2.1 Intra-modal

For intra-modal image pairs, a given tissue should map to the same (or similar) intensity in each image. Therefore, the similarity can be measured by looking at the difference of intensities at corresponding voxels. Consequently, cost functions such as the mean absolute difference or mean squared difference (often called ‘least squares’) can be used – see appendix A. Note that since both positive and negative differences represent non-similar intensities, the absolute value or squared value of the difference is used. Figure 3 shows an image pair in several alignments together with the difference image.

One problem with using the difference between images as a similarity measure is that the overall intensity level and contrast can vary (see figure 3) if, for instance, changes in scanner calibration occur. This means that there will be an overall bias in the difference image which can result in a poor registration being found. A way of overcoming this problem is to use a similarity function which is normalised so that global changes in brightness and contrast (global offset and scaling of intensity) do not change the similarity value. An example of this type of similarity function is normalised correlation. This is equal to the usual statistical covariance measure ($E\{I_A I_B\} - E\{I_A\}E\{I_B\}$) divided by the standard deviation of the intensities in each individual image. It is the division by the two standard deviations that compensates for global scaling while the correlation itself is unchanged by global offsets.

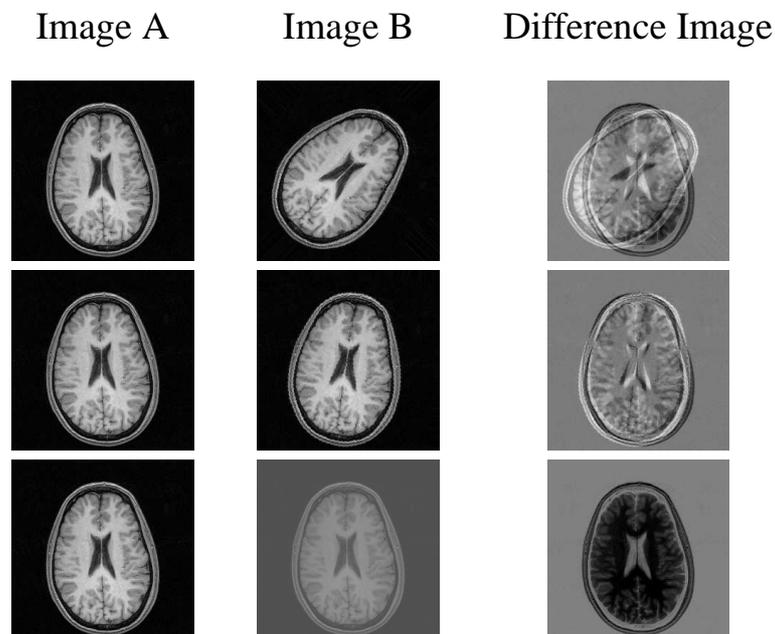


Figure 3: An example showing the difference image formed by several image pairs. Each row contains an image pair (Image A and Image B) together with the corresponding difference image (Image B - Image A), where light and dark areas in the difference image represent positive and negative values, respectively. The top two rows show the effect of a large and small rotation while the bottom row shows the effect of a change in intensity brightness and contrast. Of these, the middle difference image has the least deviation (positive or negative).

2.2.2 Inter-modal

For inter-modal image pairs, the intensities associated with the various tissue types can be completely different from one image to another. However, within an area that contains a single tissue type the intensity should be nearly constant. This is true for each image, although the mean intensity of this area in each image is likely to be very different. For example, consider some tissue, say grey matter. The intensity in one image might be 1000 ± 10 , so that any intensity near 1000 in the first image would be most likely to correspond to grey matter, whilst in the other image it might be 400 ± 5 , a very different intensity, but again the intensity of all the grey matter within this image is nearly constant.

This observation is the basis of the Correlation Ratio [30], which initially performs a segmentation of one image into areas of similar intensity. This segmentation is done by binning the intensity values as would be done to create a histogram. A bin number is assigned to each location (voxel), so that all locations with the same bin number should correspond to the same tissue type. For example, choosing a bin width of 50 would assign all locations with an intensity value between 1 and 50 to bin 1, between 51 and 100 to bin 2, between 101 and 150 to bin 3, and so on. The word *areas* will be used here to denote the set of locations with the same bin number (see figure 4). These areas represent spatial regions.

Each area is then mapped onto the second image, and the intensities inside each area examined. These intensities in the second image should be approximately constant if the image is aligned well, as each area should contain the same tissue type. However, if the images are poorly aligned there are likely to be several tissue types in any single area. Therefore, the similarity function is based on how much the intensity of the second image varies within each of the areas. Note that although the first image may be over-segmented (that is, a given tissue type may be split into more than one area) this does not matter since, if the alignment is good, the corresponding intensities in the other image will still be approximately constant, while if the alignment is poor, they will fluctuate greatly since the areas will cross tissue boundaries.

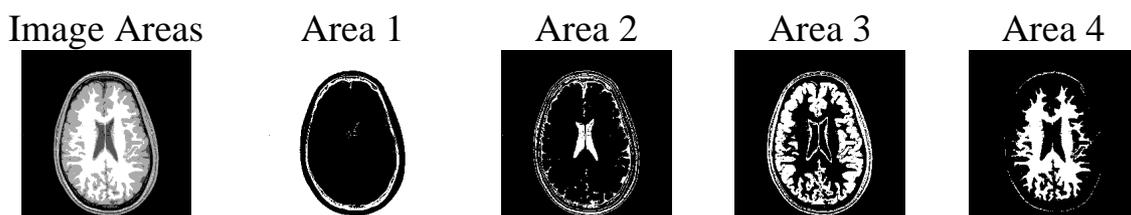


Figure 4: An illustration of the spatial areas formed by binning the intensity values. On the left all the bins are shown in the one image, represented by different shades of grey, while in the four images on the right, each area is shown separately in white. For this example only four bins were used, whereas in practice often hundreds are used.

More precisely, it is the variance of the intensities for each area in the second image that is of interest. For the Correlation Ratio (a similarity function — see appendix A) the variance in each area is first weighted by the size of each area and summed together to give a total (weighted) variance score. This variance is then normalised by dividing it by the total variance of the second image, treated as a single area, and then subtracted from one. Once again this normalisation ensures that the score is unaffected by any global intensity scalings.

Another commonly used inter-modal similarity function is Mutual Information [22, 38], or Normalised Mutual Information [34], which are based on measuring the *joint entropy* of the intensities. Entropy, a quantity used in physics and communication engineering, is a measure of disorder; a substance with high entropy is one that is very disordered. Therefore, low entropy represents an ordered situation, and it is the relation between the corresponding intensities across the two images which should be orderly when the alignment is good.

In practice, the entropy is measured from the joint histogram of the two images. This histogram is formed by assigning a bin number to each voxel, in each image separately, based on the intensity at that voxel. Then a two-dimensional array of bins is formed with the bin numbers for the first image along the vertical axis, and those for the second image along the horizontal axis. This array is the (unfilled) joint histogram and to fill it requires looking at a each voxel position in turn, finding the bin numbers from each image at this position, then adding one to the cell corresponding to the pair of bin numbers found.

For example, consider a voxel position of (10,3,7). If the bin number is 4 at this position in the first image, and is 6 at this position in the second image, then the cell (4,6) in the joint histogram has one added to it. So, by starting with zero in all cells, the joint histogram is built by examining each voxel location in turn (for both images together) and adding one to the appropriate bin. Figure 5 shows some joint histograms formed from an inter-modal pair of images in various alignments. Note how the dispersion of the histogram entries decreases as the images become well aligned.

Given the joint histogram, the entropy quantifies how orderly the entries are. This is defined mathematically with both joint and marginal (individual) entropies, which are combined to form the mutual information measures. Appendix A contains the mathematical definitions of this and the other similarity/cost functions discussed above.

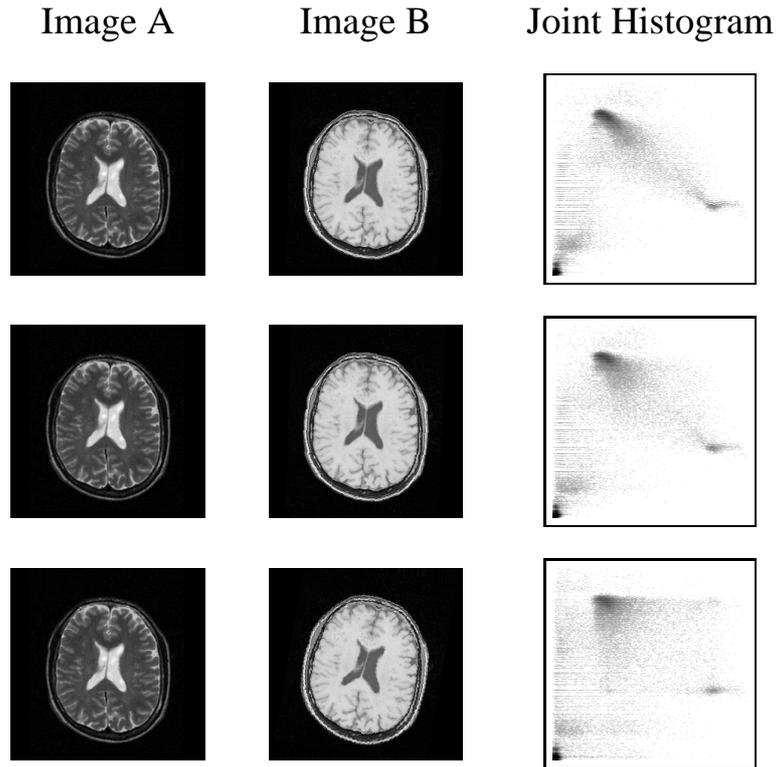


Figure 5: An illustration of some typical joint histograms formed from an inter-modal image pair in various alignments. The images used have T2 (first column) and T1 (second column) weightings and are of the same individual. Alignment of the images varies from nearly perfect (top row) to poor (bottom row). In each case, the vertical axis of the joint histogram corresponds to bin numbers of the first image, while the horizontal axis corresponds to bin numbers in the second image. Note that as the alignment gets worse the dispersion increases, but even for the nearly perfect alignment there is some dispersion. In this case the dispersion is due to the fact that the relationship between tissue type and intensity is not perfect but only approximate.

2.2.3 Summary

Each similarity function makes some assumptions about how the intensities are related between the matched image pair. These assumed intensity relationships are:

Similarity/Cost Function	Assumed Intensity Relationship	Appropriate Usage
Mean Square Difference 'Least Squares'	$I_B = I_A + \epsilon$	Intra-modal (fixed intensity scaling)
Mean Absolute Difference	$I_B = I_A + \epsilon$	Intra-modal (fixed intensity scaling)
Normalised Correlation	$I_B = mI_A + c + \epsilon$	Intra-modal (intensities can be scaled)
Correlation Ratio	$I_B = f(I_A) + \epsilon$	Inter-modal (e.g. MR-MR)
Mutual Information Normalised Mutual Information	I_B and I_A statistically related	Inter-modal (more general)

where ϵ represents image noise (also considered present in the Mutual Information cases). This table can be used to select an appropriate similarity function for given images. For example, if the two images are of the same modality (e.g. T1-weighted) and the sequence is identical then any similarity function can be used, although Mean Square Difference and Mean Absolute Difference may be more sensitive to change and more accurate. However, if the sequence parameters are slightly different (or the scanner gain changes) then Normalised Correlation would be the most appropriate. If, on the other hand, the modalities are different (e.g. T1-weighted and T2-weighted) then *only* Correlation Ratio and the Mutual Information-based similarity functions can be used. Note that the difference between these two types is that Mutual Information does not insist on a strict relationship between intensities and can be used when the image modalities are extremely different (e.g. CT and MR) whereas Correlation Ratio assumes that if there is no distinction between tissue intensities in one image (I_A) then there cannot be one in the other image, which is generally true for MR to MR image registration, but not for CT to MR.

2.3 Finding the Transformation

Once the similarity function has been chosen, the transformation that maximises this similarity must be found. Due to the complexity of the similarity functions there is usually no analytical (that is, directly calculable) solution available, and so the solution must be found by searching. The problem of searching for the transformation parameters that give the best function value (global maximum) is a standard problem in mathematics, called the optimisation problem. Consequently, there are many different algorithms available for solving the optimisation problem.

A difficulty with optimisation in the context of registration is that only the global maximum is of interest, not local maxima, but many optimisation algorithms only aim to find local maxima. The difference between a local maximum and the global maximum is shown in figure 6. As can be seen, there is only one global maximum for a function (the single biggest value), which may be hidden in amongst many local maxima. In practice, some registration methods simply find the nearest local maximum, and so rely on starting close enough to the global maximum, but this can lead to non-robust registration. To solve this problem in general requires more sophisticated optimisation algorithms.

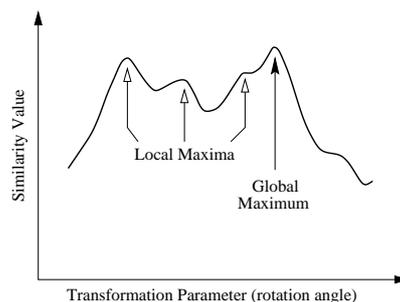


Figure 6: An illustration, using similarity function values calculated for a real image pair, showing both local maxima and the global maximum.

2.3.1 Optimisation

Searching for the transformation that gives the best similarity value (the global maximum) is the task of the optimisation algorithm. Conceptually, the problem seems simple, as it is easy to spot the global maximum in a graph like that shown in figure 6. However, in practice this is difficult because only isolated values can be calculated, one at a time. Therefore, the overall shape and general trends that make it easy to find in the example figure, are not available to the algorithm.

To illustrate this problem, consider an old sailing ship in coastal waters which wants to find the deepest point within some region. At any particular position that the ship is in, it can measure the depth by letting out a chain and measuring how far it has to go before hitting the sea-bed. Initially, the depth is unknown at all positions, and the depth can only be measured at particular positions. Thus it is necessary to start somewhere and then progressively try new positions until the captain is satisfied that the deepest point has been found. What makes this search difficult is that it is largely blind. Given a handful of depth measurements, the difficult steps are deciding where to look next and when to stop, therefore assuming that the deepest point has been found. Unfortunately, there is no general guarantee at any stage that the deepest point has indeed been found.

In this analogy the process of letting out the chain is equivalent to calculating the similarity function. The position of the ship is equivalent, not to any position, but to a particular value of the transformation parameters (for instance, rotation angles and translation values). That is, the ship attempts to find the position where the depth is greatest, but the optimisation algorithm needs to find the transformation parameters that give the greatest similarity. This makes the problem much harder because for the ship analogy there were only two dimensions to search, but for transformations there are many dimensions (equal to the DOF) which makes the general scope of the search much, much larger.

For automatic registration, choosing an appropriate optimisation algorithm is crucial for two reasons. Firstly, this is the most time-consuming part of the registration process, and secondly, if the optimisation gives bad results then the overall registration will be bad, regardless of what similarity function and transformation model is chosen. Therefore, the algorithm should ideally be both fast and robust (i.e. unlikely to give bad results by finding local maxima).

Although there are many mathematical local optimisation algorithms available (like gradient ascent, Powell’s method, etc. — see [29] for more) there are few global optimisation algorithms. Of these, Simulated Annealing (also see [29]) is the most well known, and does provide a statistical guarantee of finding the global maximum, but has the disadvantage that it requires many, many evaluations of the similarity function (which takes a long time) before finding the desired solution. In many cases there is no standard optimisation algorithm that is suitable, and so a custom-made algorithm must be designed (e.g. see [17]) to satisfy both efficiency and robustness requirements, given the particular similarity function that is chosen.

For non-linear transformations, the problem is even more difficult as there are potentially millions of parameters to find, and any form of search in such a high dimensional parameter space is prohibitively slow. Therefore, most optimisation algorithms for non-linear registrations start by finding the best initial affine transformation and then assuming that this is ‘close’ to the desired solution so that all other parameter changes will be small and can be found using quick, local optimisation algorithms. In addition, they usually rely on using multi-scale and multi-resolution analysis techniques to speed the search, which is discussed next.

2.3.2 Multi-Scale and Multi-Resolution Analysis

When looking at brain images it is normally easy for a person to identify a rough alignment based on the gross features in the image, such as the outline of the skull or brain. Mathematically, a similar thing can be achieved by using a multi-scale approach. The basic idea is to initially blur and sub-sample the images so that the fine detail is lost; then only the gross features are used to get an initial alignment. This is usually repeated at several different scales (amounts of blurring and sub-sampling) so as to refine the fit using progressively finer and finer details.

There are two main advantages to using such a multi-scale/multi-resolution approach. One is to improve the robustness of the optimisation, by ensuring that the initial fit is ‘close’ to the desired solution (the global maximum), and the other is to speed up the optimisation process. This speed up is possible because the resampled image has a larger voxel size and hence contains fewer voxels. Consequently, the similarity function to be evaluated more quickly. For example, sub-sampling an image from $1 \times 1 \times 1 \text{mm}$ voxels to $4 \times 4 \times 4 \text{mm}$ voxels decreases the number of voxels in the image by a factor of 64, allowing the optimisation at this scale to be 64 times faster. Figure 7 shows an illustration of this sub-sampling scheme, with an initial image (1mm cubed voxels) and three typical sub-samplings (2mm , 4mm and 8mm cubed voxels respectively). Note that in order to ensure robustness the images must be blurred by an appropriate amount prior to sub-sampling.

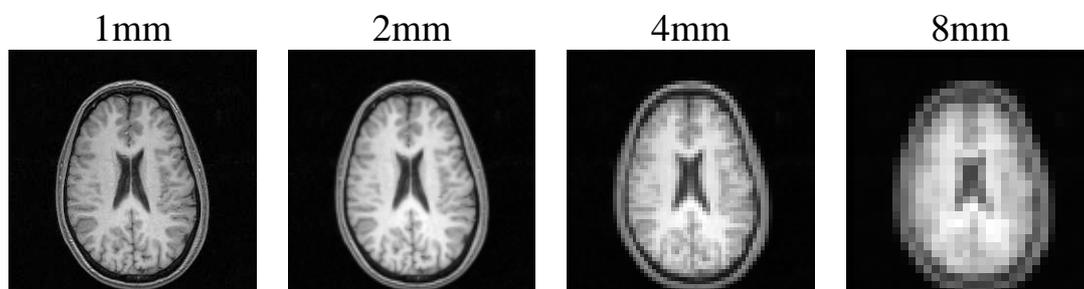


Figure 7: Example showing an image with $1 \times 1 \times 1 \text{mm}$ voxels, and three sub-samplings of this image with $2 \times 2 \times 2 \text{mm}$, $4 \times 4 \times 4 \text{mm}$ and $8 \times 8 \times 8 \text{mm}$ voxels.

3 Pre-processing, Artefacts and Registration Error

When registering two images the selection of the desired options (especially a similarity/cost function and transformation model) effectively define a model for what kind of differences can exist between well matched images. For instance, the use of the mean squared difference cost function and a rigid-body transformation model implies that the matched images will have the same intensities up to some amount of additive noise, and that the anatomical structures can be aligned with only a set of global rotations and translations. However, in practice, unwanted, complicated differences between images exist which cannot easily be dealt with by the available registration options.

The two most common types of unwanted differences are those due to physical/anatomical differences, and those due to scanning artefacts. Examples of the former category include new gross pathologies, such as lesions, and

changes in neighbouring structures which are of no interest, such as lung size, fatty tissue deposits or jaw position. Examples of the latter category include bias field effects (RF inhomogeneities), ghosting, motion distortions and distortions caused by B0 inhomogeneities.

Two common methods of dealing with these unwanted differences are by pre-processing the images and by using masking or weighting functions in the registration. Pre-processing strategies such as bias field removal, brain extraction, B0-unwarping and ghost reduction, to name a few, can be crucial for obtaining good registrations by removing the differences before registration is run. Alternatively, a masking or weighting function for the similarity/cost function [5] can be used to ignore (or downweight) areas of the image where differences are known, or expected, to exist. For instance, manually masking out lesions or calculating expected signal loss masks for B0 distortions are two cases where masks are useful. Note that this is different to masking the images themselves as that introduces false intensity changes on the edge of the mask, whereas incorporating masking within the similarity function itself just ignores the intensities in those voxels.

Errors in registration exist due to several reasons, including: unremoved artefacts/differences, failed optimisation (local maxima), inappropriate choice of options (similarity function, transformation, interpolation, etc) and noise. The last of these, noise, is always present and will induce some inaccuracy in the final registration. Therefore, all registration will be inaccurate to some degree, however many images have such high Signal to Noise Ratios (SNR) that the main reason for inaccuracy is not the noise but the other problems. Hence the typical accuracy with which a registration can be achieved depends on aspects of the registration algorithm (e.g. optimisation method), the user's selected options (e.g. similarity function) and the artefact levels present in the image. Consequently it is difficult to give any 'typical' error values, although in certain situations attempts to characterise the errors have been made (e.g. [39] for rigid-body registration). However, due to the difficulty in finding test data with known ground truth, it is more useful and practical to estimate the error for each individual application by careful manual assessment. This information about how accurate the registration is can be crucial for evaluating the reliability and accuracy of quantitative measures derived from the registered images.

4 Motion Correction and Other Applications

As mentioned in the introduction, there are many applications of registration and its use in the early (or late) stages of quantitative analysis is very common. This section will briefly touch on three common applications: motion correction, multiple acquisition averaging and detecting structural changes in longitudinal studies.

Motion correction is common in functional imaging, especially in neurological fMRI. In this case the motion that needs to be corrected is the bulk motion of the head (brain) which is a case of rigid-body motion. Consequently this can be solved by multiple application of intra-modal rigid-body registration to the fMRI series (by registering each image in the series to a chosen 'reference' image from the series) and this is what is implemented in most motion correction methods. However, this is still a difficult problem because of the following factors: (a) very high accuracy (approximately 0.5mm or better) is required to correct for induced intensity fluctuations of the size of the BOLD signal changes; (b) artefacts such as spin-history effects [12] and B0-distortion interacting with the motion [18, 1]; (c) changes in motion between slices [19, 3]; and (d) non-rigid motion in areas such as the brain stem. Consequently, motion correction methods are still an active area of research.

A common way of acquiring MR images with good SNR is to average several separate images together. However, in the same way that motion can occur between image acquisitions in a functional series, this motion can also occur between these separate acquisitions and needs to be accounted for. This correction can also be done using a series of pairwise image registrations, but the details of what sort of registration is required depends on what structures are involved and how they move. For example, in brain imaging it is normally assumed that rigid-body motion is sufficient (ignoring the small non-rigid motions at ventricle edges and in the brain stem). Alternatively, for imaging the spine it is often necessary to perform non-linear registration (possible with a rigid-body constraint for each vertebra) and it may also be necessary to acquire the images at a fixed point in the respiration cycle to limit the change in lung position and size. Therefore the details of the required registration settings are very application dependent, and careful thought must be put into choosing the correct registration options as well as in the appropriate acquisition and pre-processing measures.

Finally, longitudinal studies are commonly performed in order to detect subtle changes in anatomical shape or in quantitative parameters within tissue, which can act as markers for disease. Registration of the images prior to comparison is essential and errors in the registration can have a significant effect on the final outcome. Consequently it is important to minimise the errors as much as possible by a combination of all the available acquisition,

pre-processing and weighting steps outlined in the previous section. It is also useful to understand the effect that registration errors can have on the results, and ideally this can be taken into account in the analysis, or at least the common form of the erroneous results taken into account subjectively. For example, when there is a registration error, the largest changes in intensity occur at the boundary between structures (where the intensity change is largest) and so false positive results (reflecting large changes that are purely due to registration error) are likely to form near these boundaries. The distance from such edges that this kind of false positive is likely to exist depends on the registration accuracy, and so having an estimate of this is very useful, although post-processing stages such as spatial filtering (blurring) can alter this.

5 Summary

Most automatic (i.e. non-manual) registration methods require: (1) a transformation model, (2) a similarity function and (3) an optimisation algorithm. It is therefore necessary to select an appropriate transformation model, similarity function and optimisation method for each different application. Sometimes this involves choosing the right parameters within a given implementation, and sometimes this means choosing between different implementations depending on how many user-selectable options are available.

The transformation model must be selected to suit the problem at hand. If an exact match between two images (or the structures within these images) is required and the anatomy is different, then a high DOF transformation model (e.g. viscous-fluid) should be used. However, if the images have the same anatomy (e.g. brain of the same subject in the same session) then a rigid-body model should be used as this enforces the appropriate constraints. Note that, in general, the lower DOF models are less sensitive to image quality and hence more robust.

Similarity functions will determine what range of intensity changes are allowable. For instance, some similarity functions are only suitable for intra-modal registration while others are suitable for inter-modal registration as well (where arbitrary intensity relationships can hold). Note that any inter-modal similarity function should also be capable of working for intra-modal image pairs, but it may have more local maxima than the simpler intra-modal similarity functions and consequently produce less reliable or robust results.

The optimisation method is the final, and equally important, component of a registration tool. This is what principally determines the speed and robustness of the method. Usually each implementation of a registration tool will use one particular optimisation method, which is an integral part of the most inner workings of the algorithm, and not easily changed. Hence it is often only possible to change optimisation methods by choosing a different implementation.

Finally, it is equally if not more important to select appropriate acquisitions, pre-processing and masking options to minimise the effects of artefacts and non-modelled image changes such as gross anatomical differences between images (e.g. new lesions). Any changes in the images which cannot be modelled by the spatial transformation model and the similarity function assumptions needs to be dealt with prior to the registration. For example, bias field is a common and problematic artefact in MR images which introduces a intensity scaling factor that varies slowly across the image (although less slowly for surface coils and very high field scanners). Therefore using a bias field correction prior to registration is often essential to obtain good results, especially when large DOF transformations are being estimated, as the higher DOF transformations are more sensitive to effects like this. Similarly, masking or weighting the images to ignore areas containing pathologies (e.g. lesions) or structures of no interest (e.g. jaw in neuro-imaging) can considerably improve accuracy and robustness of registration results.

Any automatic registration method will contain some inaccuracy (due to noise, artefacts *etc.*). In addition, it is possible for gross mis-registrations to occur if, for instance, a local maximum is found rather than the global maximum. Such errors can be common in certain applications as no optimisation method, in practice, can guarantee to find the best solution every time. For these reasons, it is always advisable to check the registration results visually and obtain some subjective (or, ideally, objective) estimate of the registration accuracy. This estimate of error should then be taken into account in subsequent analyses of the image data.

Appendix A: Mathematical Definitions

This appendix contains mathematical definitions of some commonly used similarity/cost functions for both intra-modal and inter-modal registration. Let the two images be I_A and I_B with the voxel intensity at a location j (in I_A) denoted by $I_{A,j}$. Furthermore, let N be the total number of voxels in an image.

A.1 Intra-modal

$$\begin{aligned} \text{Mean Absolute Difference} &= \frac{1}{N} \sum_j |I_{A_j} - I_{B_j}| \\ \text{Mean Square Difference} &= \frac{1}{N} \sum_j (I_{A_j} - I_{B_j})^2 \\ \text{Normalised Correlation} &= \frac{\frac{1}{N} \sum_j (I_{A_j} - \overline{I_A})(I_{B_j} - \overline{I_B})}{\sqrt{\frac{1}{N} \sum_j (I_{A_j} - \overline{I_A})^2} \sqrt{\frac{1}{N} \sum_j (I_{B_j} - \overline{I_B})^2}} \end{aligned}$$

where $\overline{I_A} = \frac{1}{N} \sum_j I_{A_j}$ and $\overline{I_B} = \frac{1}{N} \sum_j I_{B_j}$. Note that the MAD and MSD are cost functions, NC is a similarity function.

A.2 Inter-modal

$$\begin{aligned} \text{Correlation Ratio} &= 1 - \sum_j \frac{n_j \sigma_j^2}{N \sigma^2} \\ \text{Mutual Information} &= H(I_A) + H(I_B) - H(I_A, I_B) \\ \text{Normalised Mutual Information} &= \frac{H(I_A, I_B)}{H(I_A) + H(I_B)} \end{aligned}$$

where σ_j^2 is the variance, μ_j is the mean value, and n_j is the number of voxels in area j ; and σ^2 is the total variance across the whole image. Joint entropy is given by $H(I_A, I_B) = - \sum_{ij} \frac{n_{ij}}{N} \log \left(\frac{n_{ij}}{N} \right)$ where n_{ij} is the number of voxels that were assigned to the bin pair (i, j) . The marginal entropies, $H(I_A)$ and $H(I_B)$, are defined similarly, but using the individual image histograms rather than the joint histogram. Note that NMI is a cost function, while CR and MI are similarity functions.

References

- [1] J.L.R. Andersson, C Hutton, J Ashburner, R Turner, and K Friston. Modelling geometric distortions in EPI time series. *NeuroImage*, 13(5):903–919, 2001.
- [2] M.A. Audette, F.P. Ferrie, and T.M. Peters. An algorithmic overview of surface registration techniques for medical imaging. *Medical Image Analysis*, 4(3):201–217, 2000.
- [3] P.M. Bannister, J.M. Brady, and M. Jenkinson. Tiger: A new model for spatio-temporal realignment of fmri data. In *Workshop on Computer Vision Approaches to Medical Image Analysis*, Prague, May 2004. 8th European Conference on Computer Vision - ECCV.
- [4] F.L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2(6):567–585, 1989.
- [5] M. Brett, A.P. Leff, C. Rorden, and J. Ashburner. Spatial normalization of brain images with focal lesions using cost function masking. *NeuroImage*, 14(2):486–500, 2001.
- [6] M. Bro-Nielsen. Modelling elasticity in solids using active cubes application to simulated operations. In *Computer vision, virtual reality, and robotics in medicine: Lecture notes in computer science*, volume 905, pages 535–541, Berlin, 1995. Springer-Verlag.
- [7] M. Bro-Nielsen and C. Gramkow. Fast fluid registration of medical images. In *Visualization in Biomedical Computing*, pages 267–276. Springer-Verlag, 1996.
- [8] G.E. Christensen, S.C. Joshi, and M.I. Miller. Volumetric transformation of brain anatomy. *IEEE Trans. on Medical Imaging*, 16(6):864–877, 1997.
- [9] W. Crum, L. Griffin, D. Hill, and D. Hawkes. Zen and the art of medical image registration: correspondence, homology, and quality. *NeuroImage*, 20(3):1425–1437, 2003.
- [10] W. Crum, T. Hartkens, and D. Hill. Non-rigid image registration: Theory and practice. *British Journal of Radiology: Special Supplement on Medical Image Analysis*, 2005.
- [11] K.J. Friston, J. Ashburner, C.D. Frith, J.-B. Poline, J.D. Heather, and R.S.J. Frackowiak. Spatial registration and normalization of images. *Human Brain Mapping*, 2:165–189, 1995.
- [12] K.J. Friston, S.R. Williams, R. Howard, R.S.J. Frackowiak, and R. Turner. Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, 35:346–355, 1996.
- [13] A. Hagemann, K. Rohr, H.S. Stiehl, U. Spetzger, and J.M. Gilsbach. Biomechanical modeling of the human head for physically based, nonrigid image registration. *IEEE Trans. on Medical Imaging*, 18(10):875–884, 1999.
- [14] J.V. Hajnal, D.L.G. Hill, and D.J. Hawkes, editors. *Medical Image Registration*. CRC Press, Boca Raton, 2001.
- [15] D.L.G. Hill, P.G. Batchelor, M. Holden, and D.J. Hawkes. Medical image registration. *Phys. Med. Biol.*, 46:R1–R45, 2001.
- [16] B.F. Hutton, M. Braun, L. Thurfjell, and D.Y.H. Lau. Image registration: an essential tool for nuclear medicine. *European Journal of Nuclear Medicine and Molecular Imaging*, 29(4):559–577, 2002.
- [17] M. Jenkinson and S.M. Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156, June 2001.

- [18] P. Jezzard and R.S. Balaban. Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine*, 34:65–73, 1995.
- [19] B. Kim, J.L. Boes, P.H. Bland, T.L. Chenevert, and C.R. Meyer. Motion correction in fMRI via registration of individual slices into an anatomical volume. *Magnetic Resonance in Medicine*, 41:964–972, 1999.
- [20] Hava Lester and Simon R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32:129–149, 1999.
- [21] J.A. Little, D.L.G. Hill, and D.J. Hawkes. Deformations incorporating rigid structures. *Computer Vision and Image Understanding*, 66(2):223–232, 1997.
- [22] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximisation of mutual information. *IEEE Trans. on Medical Imaging*, 16(2):187–198, 1997.
- [23] J.B.A. Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.
- [24] T. Makela, P. Clarysse, O. Sipila, N. Pauna, Q.C. Pham, T. Katila, and I.E. Magnin. A review of cardiac image registration methods. *IEEE Trans. on Medical Imaging*, 21(9):1011–1021, 2002.
- [25] E.H.W. Meijering, W.J. Niessen, and M.A. Viergever. Retrospective motion correction in digital subtraction angiography: a review. *IEEE Trans. on Medical Imaging*, 18(1):2–21, 1999.
- [26] Jan Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, Institute of Mathematics, University of Lbeck, 2004.
- [27] C. Pierpaoli, P. Jezzard, P.J. Basser, A. Barnett, and G. Di Chiro. Diffusion tensor mr imaging of the human brain. *Radiology*, 201(3):637–48, 1996.
- [28] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever. Mutual information based registration of medical images: a survey. *IEEE Trans. on Medical Imaging*, 22(8):986–1004, 2003.
- [29] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C*. Cambridge University Press, second edition, 1995.
- [30] A. Roche, G. Malandain, X. Pennec, and N. Ayache. Multimodal image registration by maximization of the correlation ratio. Technical Report 3378, INRIA Sophia-Antipolis, 1998.
- [31] J.G. Rosenman, E.P. Miller, G. Tracton, and T.J. Cullip. Image registration: An essential part of radiation therapy treatment planning. *International Journal of radiation Oncology Biology Physics*, 40(1):197–205, 1998.
- [32] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. on Medical Imaging*, 18(8):712–721, 1999.
- [33] D. Shen and C. Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. on Medical Imaging*, 21(11):1421–39, 2002.
- [34] C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86, 1999.
- [35] J. Talairach and P. Tournoux. *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme Medical Publisher Inc., New York, 1988.
- [36] J.-P. Thirion. Image matching as a diffusion process: an analogy with Maxwell’s demons. *Medical Image Analysis*, 2(3):243–260, 1998.
- [37] A.W. Toga and P. Thompson. *Brain Warping*, chapter 1. Academic Press, 1999.
- [38] P. Viola and W.M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.
- [39] J. West et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21(4):554–566, 1997.
- [40] R.P. Woods, S.R. Cherry, and J.C. Mazziotta. Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*, 16(4):620–633, 1992.
- [41] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.