

Image Registration for MRI

PETER J. KOSTELEC AND SENTHIL PERIASWAMY

ABSTRACT. To *register* two images means to align them so that common features overlap and differences — for example, a tumor that has grown — are readily apparent. Being able to easily spot differences between two images is obviously very important in applications. This paper is an introduction to image registration as applied to medical imaging. We first define image registration, breaking the problem down into its constituent component. We then discuss various techniques, reflecting different choices that can be made in developing an image registration technique. We conclude with a brief discussion.

1. Introduction

1.1. Background. To *register* two images means to *align* them, so that common features overlap and differences, should there be any, between the two are emphasized and readily visible to the naked eye. We refer to the process of aligning two images as *image registration*.

There are a host of clinical applications requiring image registration. For example, one would like to compare two Computed Tomography (CT) scans of a patient, taken say six months ago and yesterday, and identify differences between the two, e.g., the growth of a tumor during the intervening six months (Figure 1). One could also want to align Positron Emission Tomography (PET) data to an MR image, so as to help identify the anatomic location of certain mental activation [43]. And one may want to register lung surfaces in chest Computed Tomography (CT) scans for lung cancer screening [7]. While all of these identifications can be done in the radiologist's head, the possibility always exists that small, but critical, features could be missed. Also, beyond identification itself, the extent of alignment required could provide important quantitative information, e.g., how much a tumor's volume has changed.

Kostelec's work is supported in part by NSF BCS Award 9978116, AFOSR under award F49620-00-1-0280, and NIH grants PO1 CA80139. Periaswamy's work is supported in part by NSF Grants EIA-98-02068 and IIS-99-83806.

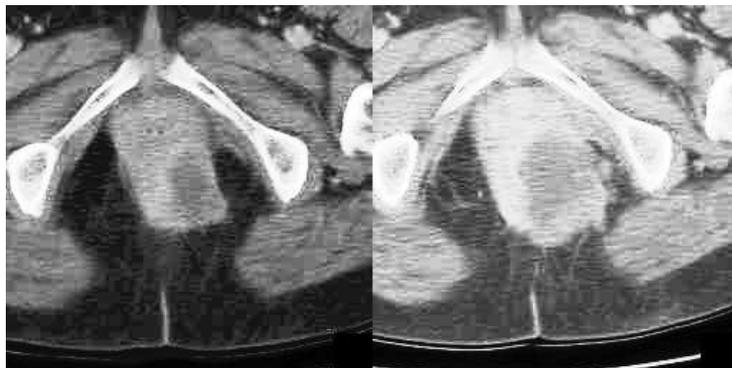


Figure 1. Two CT images showing a pelvic tumor’s growth over time. The grayscale has been adjusted so as to make the tumor, the darker gray area within the mass in the center of each image, more readily visible. In actuality, it is barely darker than the background tissue.

When registering images, we are determining a geometric transformation which aligns one image to fit another. For a number of reasons, simple image subtraction does not work. MR image volumes are acquired one slice at a time. When comparing a six month old MR volume with one acquired yesterday, chances are that the slices (or “imaging planes”) from the two volumes are not parallel. As a result, the perspectives would be different. By this, we mean the following. Consider a right cylindrical cone. A plane slicing through the cone, parallel to its base, forms a circle. If the slice is slightly off parallel, an ellipse results. In terms of human anatomy, a circular feature in the first slice appears as an ellipse in the second. In the case of mammography, tissue is compressed differently from one exam to the next. Other architectural distortions are possible. Since the body is an elastic structure, how it is oriented in gravity induces a variety of non-rigid deformations. These are just some of the reasons why simple image subtraction does not work.

For the neuroscientist doing research in functional Magnetic Resonance Imaging (fMRI), the ability to accurately align image volumes is of vital importance. Their results acutely depend on accurate registration. To provide a brief background, to “do” fMRI means to attempt to determine which parts of the brain are active in response to some given stimulus. For instance, the human subject, in the MR scanner, would be asked to perform some task, e.g., finger-tap at regular intervals, or attend to a particular instrument while listening to a piece of music [20], or count the number of occurrences of a particular color when shown a collection of colored squares [8]. As the subject performs the task, the researcher effectively takes 3-D MR movies of the subject’s brain. The goal is to identify those parts of the brain responsible for processing the information the

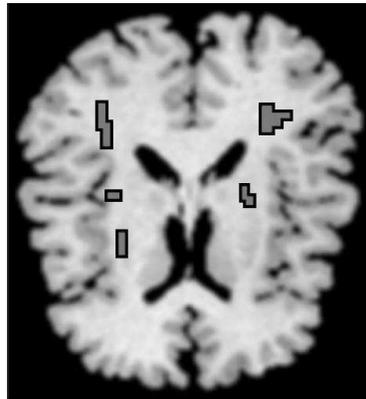


Figure 2. fMRI. By registering the frames in the MR “movie” and performing statistical analyses, the researcher can identify the active part(s) of the brain by finding those pixels whose intensities change most in response to the given stimulus. The active pixels are usually false-coloured in some fashion, to make them more obvious, similar to those shown in this figure.

stimulus provides. The researcher’s hope of accomplishing this is based on the Blood Oxygenation Level Dependent (BOLD) hypothesis (see [6]).

The BOLD hypothesis roughly states that the parts of the brain that process information, in response to some stimulus, need more oxygen than those parts which do not. Changes in the blood oxygen level manifest themselves as changes in the strength of the MR signal. This is what the researcher attempts to detect and measure. The challenge lies in the fact that the changes in signal strength are very small, on the order of only a few percent greater than background noise [5]. And to make matters worse, the subject, despite their noblest intentions, cannot help but move at least ever so slightly during the experiment. So, before useful analysis can begin, the signal strength must be maximized.

This is accomplished by task repetition, i.e., having the subjects repeat the task over and over again. Then all the image volumes are registered *within* each subject. Assuming gaussian noise, adding the registered images will strengthen the elusive signal. Statistical analyses are done within subject, and then combined across all subjects. This is the usual order of events [18].

1.2. What’s inside this paper. This will be a whirlwind, and by no means exhaustive, tour of image registration for MRI. We will briefly touch upon a few of the many and varied techniques used to register MR images. Note that the survey articles by Brown [11] and Van den Elsen [38] are excellent sources for more in-depth discussion of image registration, the problem and the techniques. Our purpose here, within this paper, is to whet the reader’s appetite, to stimulate her interest in this very important image processing challenge, a challenge which has a host of applications, both in medical imaging and beyond.

The paper is organized as follows. We first give some background and establish a theoretical framework that will provide a means of defining the critical components involved in image registration. This will enable us to identify those issues which need to be addressed when performing image registration. This will be followed by examples of various registration techniques, explained at varying depths. The methods presented are not meant to represent any sort of definitive list. We want to point out to the reader just some of the techniques which exist, so that they can appreciate how difficult the problem of image registration is, as well as how varied the solutions can be. We close with a brief discussion.

Acknowledgments. We thank Daniel Rockmore and Dennis Healy for inviting us to participate in the MSRI Summer Graduate Program in Modern Signal Processing, June 2001. We also thank Digger ‘The Boy’ Rockmore for helpful discussions, and for granting us the use of his image in this paper.

2. Theory

Suppose we have two brain MR images, taken of the same subject, but at different times, say, six months ago and yesterday. We need to align the six month old image, which we will call the *source image*, with the one acquired yesterday, the *target image*. (These terms will be used throughout this paper.) A tumor has been previously identified, and the radiologist would like to determine how much the tumor has grown during the six weeks. Instead of trying to “eyeball it,” the two images would enable an quantitative estimate of the growth rate. How do we proceed?

Do we assume that a simple rigid motion will suffice? Determining the correct rotation and translation parameters is, as we will see later, a relatively quick and straightforward process. However, if non-linear deformations have occurred within the brain (which, as described in Sec. 1.1, is likely for any number of reasons), applying a rigid motion model in this situation will not produce an optimal alignment. So probably some sort of non-rigid or elastic model would be more appropriate.

Are we looking to perform a global alignment, or a local one? That is, will the same transformation, e.g., affine, rigid body, be applied to the entire image, or should we instead employ a local model of sorts, where different parts of the image/volume are moved in different, though smoothly connected, ways?

Should the method we use depend on active participation by the radiologist, to help “prime” or “guide” the method so that accurate alignment is achieved? Or do we instead want the technique to be completely automated and free of human intervention?

Wow, that’s a lot of questions we have to think about, and answer, too. How do we begin? To tackle the alignment problem, we had first better *organize* it.

2.1. The four components. The multitude of challenges inherent in performing image registration can be better addressed by distilling the problem into four distinctive components [11].

I. The feature space. Before registering two images, we must decide exactly *what* it is that will be registered. The type of algorithm developed depends critically on the *features* chosen. And when you think about it, there are a lot of features from which to choose. Will we work with the raw pixel intensities themselves? Or perhaps the edges and contours of the images? If we have volumetric data, perhaps we should use the surface the volume defines, as in a 3-D brain scan? We could have the user identify features common to both images, with the intent to aligning those landmarks. Then again, if we wish to align images of different modalities, say MRI with PET, then perhaps statistical properties of the images that would be optimal for our purpose. So you see, the *feature space* we choose will really drive the algorithm we develop.

II. The search space. When one says, “I want to align these two images,” what is one really saying? That is, what is the rigorous form of the sentence? The two images can be considered samples of two (unknown), compact, real-valued functions, $f(\mathbf{x})$, $g(\mathbf{x})$, defined on \mathbb{R}^n (where n is 2 or 3). To align the images means we wish to find a transformation $T(\mathbf{x})$ such that $f(\mathbf{x}) = g(T(\mathbf{x}))$ for all \mathbf{x} . Fine. So what kind of transformation are we willing to consider? This is the *Search Space* we need to define.

For example, you can consider the simple rigid body transformations, rotation plus translation. Or, if you would like to account for differences in scale, you may instead decide to search for the best affine transformation. But both of these transformations are global in some respect, and you may want to do something more localized or *elastic*, and transform different parts of the image by differing amounts, e.g., to account for non-uniform deformations. Your decision here will very much influence the nature of the registration algorithm.

III. The search strategy. Suppose we have chosen our Search Space. We select a transformation $T_0(\mathbf{x})$ and try it. Based on the results of $T_0(\mathbf{x})$, how should we choose the next transformation, $T_1(\mathbf{x})$, to try? There are any number of ways: Linear Programming techniques; a relaxation method; some sort of energy minimization.

IV. The similarity metric. This ties in with the Search Strategy. When comparing the new transformation with the old, we need to quantify the differences between the geometrically transformed source image with the target image. That is, we need to measure how well $f(\mathbf{x})$ compares with $g(T(\mathbf{x}))$. Using mean-squared error might be the suitable choice. Or perhaps correlation is the key. Our choice will depend on many factors, such as whether or not the two images are of the same modality.

So once these choices are made, our search for an optimal transformation, one that aligns the source image with the target, continues until we find one that makes us happy.

3. A Potpourri of Methods

Given the content in Section 2, the reader can well believe that there are a multitude of registration methods possible, each resulting from a particular choice of feature and search spaces, search strategy, and similarity metric. But always bear in mind that there is no single *right* registration algorithm. Each technique has its own strengths and weaknesses. It all depends on what *you* want.

Very broadly speaking, registration techniques may be divided into two categories, *rigid* and *nonrigid*. Some examples of Rigid registration techniques include: Principal Axes [2], Correlation-based methods [12], Cubic B-Splines [37], and Procrustes [19; 34]. For Non-Rigid techniques, there are Spline Warps [9], Viscous Fluid Models [13], and Optic Flow Fields [30].

The survey articles [11; 38] mentioned previously go into some of these techniques in greater depth. Now, to begin our “If it’s Tuesday, this must be Belgium” tour of MR image registration techniques.

3.1. Principal Axes. We begin with the Principal Axes algorithm (e.g., see [2]). To summarize its properties, based on the classification scheme of Section 2.1, the feature space the algorithm acts upon effectively consists of the features of the images, such as edges, corners, and the like. The search space consists of global translations and rotations. The search strategy is not so much a “search,” as we are finding the closed formed solution based on the eigenvalue decomposition of a certain covariance matrix. The similarity metric is the variance of the projection of the feature’s location vector onto the principal axis.

The algorithm is based on the straightforward and powerful observation that the head is shaped like an ellipse/ellipsoid (depending on the dimension). For purposes of image registration, the critical features of an ellipse are its center of mass, and principal orientations, i.e., major and minor axes. Using these properties, one can derive a straightforward alignment algorithm which can automatically and quickly determine a rotation + translation that aligns the source image to the target.

Let I denote the 2-D array representing an image, with pixel intensity $I(x, y)$ at location (x, y) . The center of mass, or centroid, is

$$\hat{x} = \frac{\sum_{x,y} x I(x, y)}{\sum_{x,y} I(x, y)} \quad \hat{y} = \frac{\sum_{x,y} y I(x, y)}{\sum_{x,y} I(x, y)}.$$

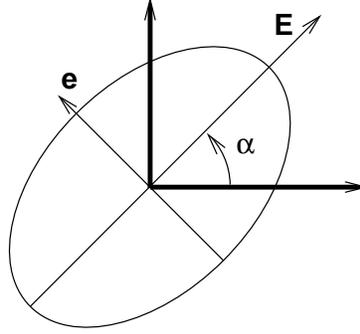


Figure 3. Principal axes. The eigenvectors \mathbf{E} and \mathbf{e} , corresponding to the largest and smallest eigenvalues, respectively, indicate the directions of the major and minor axes, respectively.

With the centroid in hand, we form the covariance matrix

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix},$$

where

$$c_{11} = \sum_{x,y} (x - \hat{x})^2 I(x, y),$$

$$c_{22} = \sum_{x,y} (y - \hat{y})^2 I(x, y),$$

$$c_{12} = \sum_{x,y} (x - \hat{x})(y - \hat{y}) I(x, y),$$

$$c_{21} = c_{12}.$$

The eigenvectors of \mathbf{C} corresponding to the largest and smallest eigenvalues indicate the direction of the major and minor axes of the ellipse, respectively. See Figure 3.

The principal axes algorithm may be described as follows. First, calculate the centroid, and eigenvectors of the source and target images via an eigenvalue decomposition of the covariance matrices. Next, align the centers of mass via a translation. Next, for each image determine the angle α (Figure 3) the maximal eigenvector forms with the horizontal axis, and rotate the test image about its center by the difference in angles. The images are now aligned.

Figure 4 shows the procedure in action. In this example, the target image is a rotated version of the source image, with a small block missing. Subtracting the target from the aligned source renders the missing data quite apparent.

While the principal axes algorithm is easy to implement, it does have the shortcoming that it is sensitive to missing data. As an exaggerated example, suppose the target MR image covers the entire head, while the source MR image has only the top half, say from the eyes on up. In this case, the anatomical

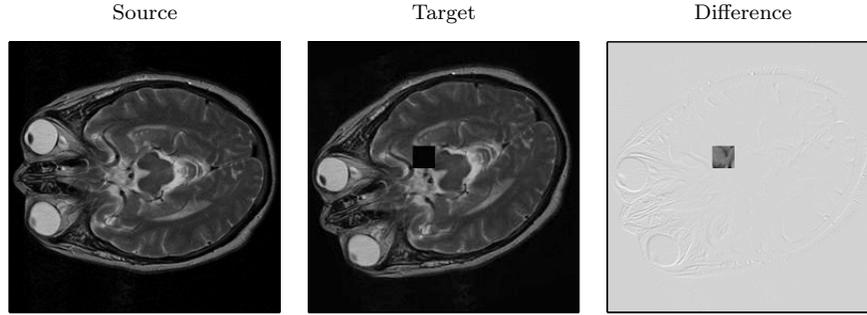


Figure 4. Principal axes: aligning axial images. The difference between the aligned source and target images is easily apparent in the far right panel.

feature located at the centroid of the source image will differ from the anatomical feature located at the centroid of the target. However, be that as it may, one can certainly use the algorithm to provide a coarse approximation to “truth.” That is, one may use rotation + translation parameters as “seed” values for more accurate methods.

3.2. Fourier-based correlation. Fourier-based Correlation is another method for performing rigid alignment of images. The feature space it uses consists of all the pixels in the image, and its search space covers all global translations and rotations. (It can also be used to find local translations and rotations [31].) As the name implies, the search strategy are the *closed form* Fourier-based methods, and the similarity metric is correlation, and its variants, e.g., phase only correlation [12]. As with Principal Axes, it is an automatic procedure by which two images may be rigidly aligned. Furthermore, it is an efficient algorithm, courtesy of the FFT [12].

The algorithm may be described as follows. Let $f(x, y)$ and $g(x, y)$ denote the source and target images, respectively. Uppercase letters will denote the function’s Fourier transform (FT):

$$f(x, y) \xleftrightarrow{\text{FT}} F(\omega_x, \omega_y), \quad g(x, y) \xleftrightarrow{\text{FT}} G(\omega_x, \omega_y).$$

To clarify, (x, y) denote coordinates in the spatial domain, and (ω_x, ω_y) denote coordinates in the frequency domain. Suppose the source and target are related by a translation (a, b) and rotation θ :

$$f(x, y) = g((x \cos \theta + y \sin \theta) - a, (-x \sin \theta + y \cos \theta) - b).$$

Then, using properties of the Fourier transform, we have

$$F(\omega_x, \omega_y) = e^{-i(a\omega_x + b\omega_y)} G(\omega_x \cos \theta + \omega_y \sin \theta, -\omega_x \sin \theta + \omega_y \cos \theta).$$

By taking norms and obtaining the power spectrum, all evidence of translation by (a, b) has disappeared:

$$|F(\omega_x, \omega_y)|^2 = |G(\omega_x \cos \theta + \omega_y \sin \theta, -\omega_x \sin \theta + \omega_y \cos \theta)|^2.$$

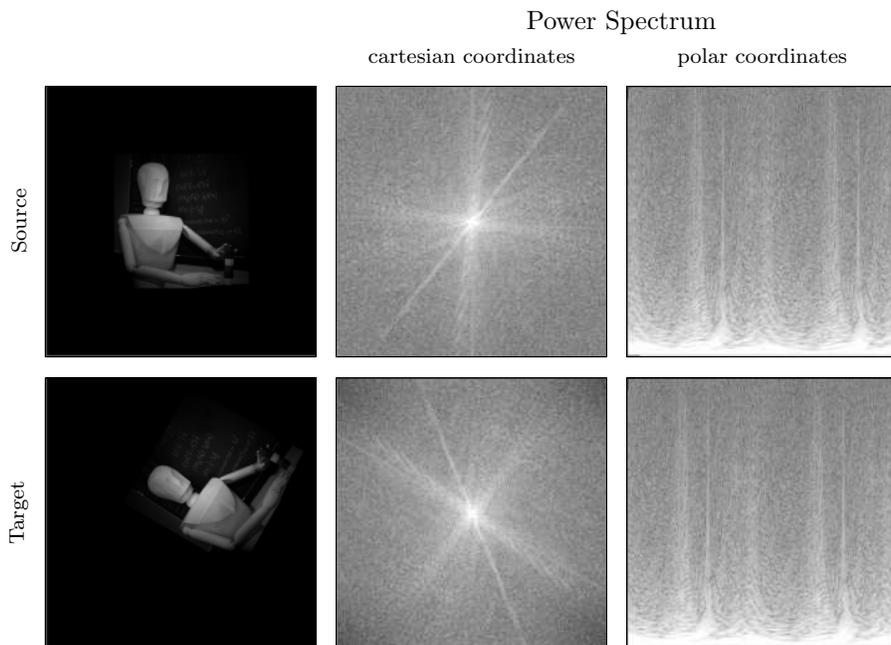


Figure 5. By considering the power spectra, translations vanish. Furthermore, in polar coordinates, rotations become translations.

Note that rotating $g(x, y)$ by θ in the spatial domain is equivalent to rotating $|G(\omega_x, \omega_y)|^2$ by that amount in the frequency domain. By switching to polar coordinates (setting $x = r \cos \psi$, $y = r \sin \psi$), we have

$$|F(r, \psi)|^2 = |G(r, \psi - \theta)|^2$$

and hence rotation in the cartesian plane becomes translation in the polar plane. See Figure 5.

We are now in a position to give an outline for the Fourier-based correlation method of image registration:

1. Take the discrete Fourier transform of the source image $f(x)$ and target image $g(x)$.
2. Next, send the power spectra to polar coordinates land:

$$|F(r, \psi)|^2 = |G(r, \psi - \theta)|^2.$$

3. Use your favourite correlation technique to determine the rotation angle. (Note that this is strictly a *translation* problem.) And then rotate the source image (which is in the spatial domain) by that amount.
4. Use your favourite correlation to now determine the translation amount in the spatial domain, between the (so far) only-rotated source image, and the target image.

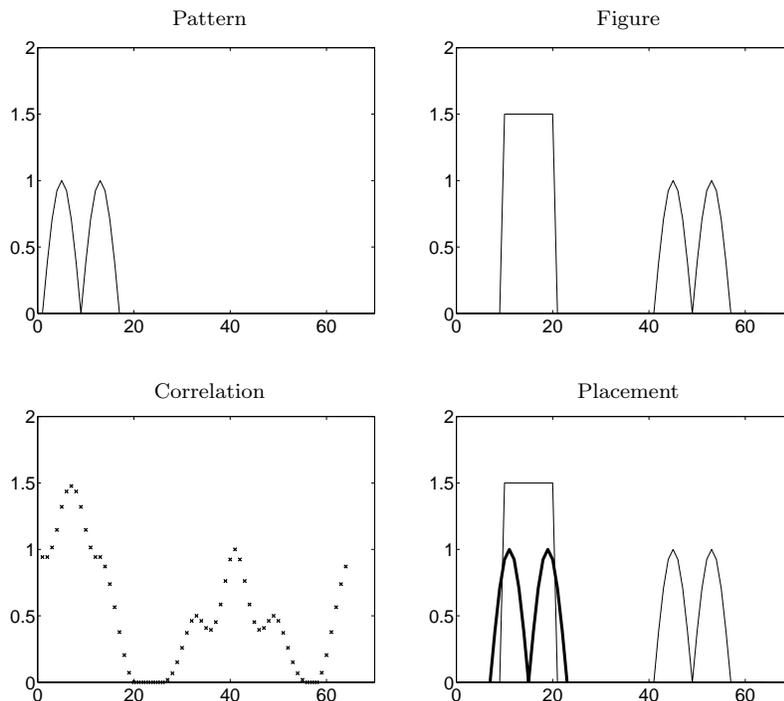


Figure 6. We seek the pattern, shown on the top left, in the signal shown on the top right. In the lower left, we plot the correlation values. The location of the maximum value should indicate the location of the pattern within the signal, but as we see in the lower right figure, placing the pattern, drawn in a thick line, at this “maximum” location is incorrect.

Given how easy and direct the algorithm is, it would come as a surprise if there were *not* any caveats associated with it.

In practice, the source and target images are probably not exactly identical. This could easily result in multiple peaks, which means that the maximum peak may not be the correct one. This phenomenon is illustrated in Figure 6. Therefore, when using correlation to determine the proper rotation and translation parameters, several potential sets of parameters, e.g., corresponding to the 4 largest correlation peaks, need to be tried. The best (in some sense, e.g., least-squares) is the value you choose. Secondly, the images certainly should be of the same modality. Registering an MR with a PET image probably won’t work at all!

But on the bright side, along with computation efficiency, one can apply the technique to subregions of images and “glue” the results together. For example, one can divide the images into quarters, determine rotation and translation parameters for each, all independent of each other, and then smoothly apply these four sets of parameters, to encompass a complete (and non-rigid) registration of the source to target image [31]. Also, as with Principal Axes, Fourier-based

correlation may be used to achieve coarse registrations, as starting points for fancier methods.

3.3. Procrustes algorithm. The Procrustes Algorithm [19; 34] is an image registration algorithm that depends on the active participation of the user. It does have as its inspiration a rather colourful character from Greek mythology. Especially for this reason, we feel compelled to briefly mention it.

It is a “one size fits all” algorithm: one image is compelled to fit another. The name is most appropriate for this algorithm. Procrustes is a character from Greek mythology. He was an innkeeper who guaranteed all his beds were the correct length for his guests. “The top of your head will be at precisely the top edge of the bed. Similarly the soles of your feet will be at the bottom edge.” And for his (unfortunate) guests of varying heights, they were. Procrustes would employ some rather gruesome measures to make his claim true. Ouch.

As already mentioned, the algorithm depends on human intervention. Quite simply, the user identifies common features or *landmarks* in the images (so this is the feature space) and, by rigid rotation and translation (the search space), forces a registration that respects these landmarks. In a perfect world, to determine the proper rotation and translation parameters, three pairs of landmarks would suffice. The rotation parameters place the images in the same orientation, the translation parameters, well, translate the images into alignment.

But we do not inhabit a perfect world. The slightest variation in distance between any homologous pair represents an error in landmark identification which cannot be reconciled with rigid body motions. And so we need to compromise. (Procrustes would have difficulty understanding this. While his enthusiasm for achieving a perfect fit is admirable, it could result in some uncomfortable side effects for the patients.) Lacking a perfect match, the similarity metric employed is instead the mean squared distance between homologous landmarks when computing the six rigid body parameters. The search strategy is to minimize via least-squares.

The good news is that this can be accomplished efficiently. A closed form solution exists, in fact. However, the not so good news is that it depends on the accurate identification of landmarks. If you say that the anatomical feature at Point A_1 in source image A really corresponds with the anatomical feature at Point B_1 in the target image B , you had better be right. And being right takes time, especially since the slightest deviation is a source of error.

3.4. AIR: automated image registration. AIR is a sophisticated and powerful image registration algorithm. Developed by Woods et al [41; 42; 43], the feature space it uses consists of all the pixels in the image, and the search space consists of up to fifth-order polynomials in spatial coordinates x , y (and z , if 3-D), involving as many as 168 parameters. The goal is to define a single, global transformation. We outline some of AIR’s characteristics:

- AIR is a fully automated algorithm.
- Unlike the algorithms so far discussed, AIR can be used in *multi-modal* situations.
- AIR does not depend on landmark identification.
- AIR uses overall *similarity* between images.
- AIR is iterative.

It is a robust and versatile algorithm. The fact that AIR software is publicly available [1] has only added to its widespread use.

AIR is based on the following assumption. If two images, acquired the same way (i.e., same modality) are perfectly aligned, then the *ratio* of one image to another, on a pixel by pixel basis, ought to be fairly uniform across voxels. If registration is not spot on correct, then there would be a substantial degree of nonuniformity in ratios. Ergo, to register the two images, compute the standard deviation of the ratio, and *minimize it*. This error function is called the “ratio of image uniformity”, or RIU. The algorithm’s search strategy is based on gradient descent, and the similarity metric is actually a normalized version of the RIU between the two volumes. An iterative procedure is used to minimize the normalized RIU in which the registration parameters (three rotation and three translation terms) with the largest partial derivative is adjusted in each iteration [41].

Since we are dealing with ratios and not pixel intensities themselves, it is this idea of using the ratios to register images which provides us with the flexibility to align images of different modalities.

Suppose we are in the situation where we want to align an MR to a PET image. On the face of it, the ratios will not be uniform across the images. Different tissue types will have different ratios. However, and this is key, *within* a given tissue type, the ratio ought to be fairly uniform when the images are registered. Therefore, what you want to do is maximize the uniformity within the tissue type, where the tissue-typing is based on the MRI voxel intensity. This requires two modifications of the original algorithm [43]. First, one has to manually edit the scalp, skull and meninges from the MR image since these features are not present in the PET image. The second modification consists of first performing a histogram matching. Denote the two images to be histogram matched as $f_1(\cdot)$ and $f_2(\cdot)$, and $c_2(\cdot)$ as the sampled cumulative distribution function of image $f_2(\cdot)$. The histogram of $f_2(\cdot)$ is made to match that of $f_1(\cdot)$ by mapping each pixel $f_1(x, y)$ to $c_2(f_1(x, y))$, between the MR and PET images (with 256 bins), followed by a segmentation of the images according to the 256 bin values. Each of the segmented MR and PET images (with corresponding bin values) are then registered separately.

In terms of implementation, both the within-modality and cross-modality versions of the algorithm, the registration is performed on sub-sampled images, in decreasing order of sub-sampling.

There are a number of things to keep in mind. AIR's global approach implies the transformation will be consistent throughout the entire image volume. However, this does introduce the possibility of obtaining an unstable transformation, especially near the image boundaries. And small and/or local perturbations may result in disproportionate changes in the global transformation. And the AIR algorithm is also computationally intensive. It is not easy, after all, to minimize the standard deviation of the ratios. However, the algorithm does perform well with noisy data [36].

3.5. Mutual information based techniques. Mutual Information [39] is an error metric (or similarity metric) used in image registration based on ideas from Information Theory. Mutual Information uses the pixel intensities themselves. The strategy is this: minimize the information content of the *difference image*, i.e., the content of target-source.

Consider Figure 7. The particular example is a bit of a cheat, but it illustrates the point. In the top row we have two axial images. They are the source

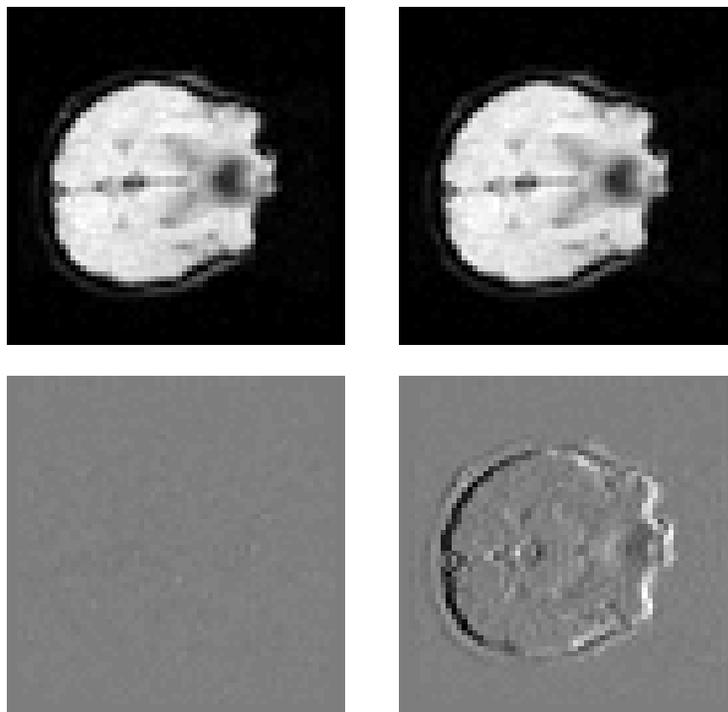


Figure 7. The philosophy behind Mutual Information. The source is the top left image, and the target is the top right. The difference image between the aligned source and target (lower left) looks nearly completely blank. Some structure might be vaguely visible, but not nearly as much as the difference image resulting translating the aligned source by 1 pixel (lower right).

and target images. The image on the lower left is the difference between the aligned source and target. Since the pixel intensities of the source and target are nearly identical, the difference image is basically blank. Now suppose we take the aligned source and translate it by one pixel. In the resulting difference image, the boundary of the skull is quite obvious. Whereas in the first difference image one has to “hunt” for features (and fail to find any), in the second we do not. Features stand out. So, in a sense, the second difference image has more information than that first: we see a shape. Mutual Information wants that difference image to have as little information as possible.

To go a little further, let us begin with the question: how well does one image explain, or “predict”, another? We use a joint probability distribution. Let $p(a, b)$ denote the probability that a pixel value a in the source and b in the target occurs, for all a and b . We estimate the joint probability distribution by making a joint histogram of pixel values. When two images are in alignment, the corresponding anatomical area overlap, and hence there are lots of high values. In misalignment, anatomical areas are mixed up, e.g., brain over skin, and this results in a somewhat more dispersed joint histogram. See Figures 8 and 9.

What we want to do is make the “crispiest” joint probability distribution possible. Let $I(A, B)$ denote the *Mutual Information* of two images A and B . This can be defined in terms of the entropies (i.e., “How dispersed is the joint probability distribution?”) $H(A)$, $H(B)$ and $H(A, B)$:

$$I(A, B) = H(A) + H(B) - H(A, B) = \sum_{x \in A, y \in B} p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right).$$

Therefore, to *maximize* their mutual information $I(A, B)$, to get image A to tell us as much as possible about B , we need to *minimize* the entropy $H(A, B)$. The reader is encouraged to read the seminal paper by Viola et al. [39] for further information regarding exactly how the entropy $H(A, B)$ is minimized. In brief, [39] use a stochastic analog of the gradient descent technique to maximize $I(A, B)$, after first approximating the derivatives of the mutual information error measure. In order to obtain these derivatives, the probability density functions are approximated by a sum of Gaussians using the Parzen-window method [16] (after this approximation, the derivatives can be obtained analytically). The geometric distortion model used is global affine. In general, the various implementations differ in the minimization technique. For example, Collignon et al. [14] use Powell’s method for the minimization.

In the final analysis, we find that Mutual Information is quite good in multi-modal situations. However, it is computationally very expensive, as well as being sensitive to the how the interpolation is done, e.g., the minimum found may not be the correct/optimal one.

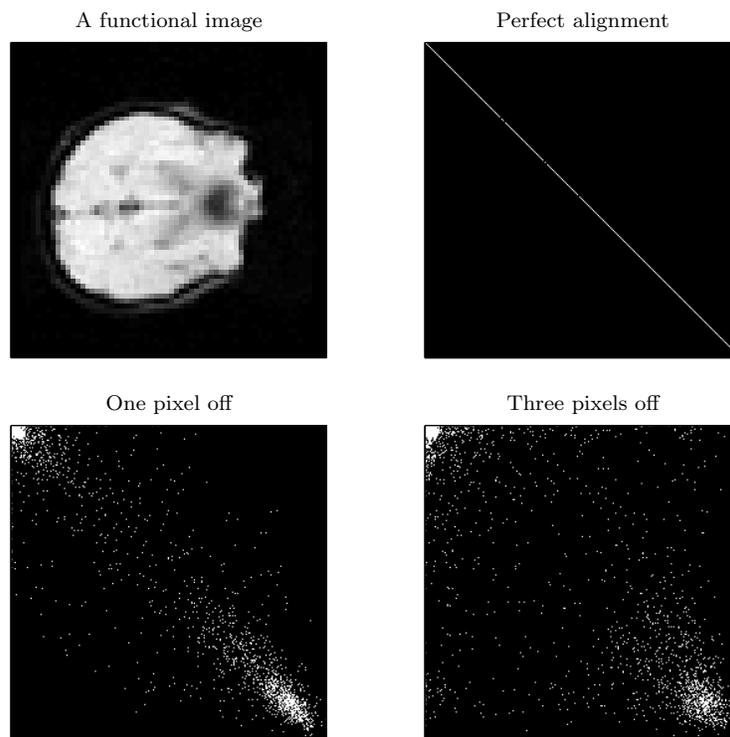


Figure 8. Joint histograms of identical source and target images. No registration is necessary to align them. The resulting joint histogram is a diagonal line. Translating by 1 pixel significantly disperses the diagonal (lower left), and by 3 pixels, further still (lower right).

3.6. Optic flow fields. This registration technique [30] borrows tools from differential flow estimation. The underlying philosophical principle of the algorithm is that we want to *flow* from the source to the target. Think of an air bubble that is rising to the surface of a lake. The bubble's surface smoothly bends and flexes this way and that as it floats upward. The source and target images are two snapshots taken of the rising bubble. Starting from the two snapshots, the algorithm determines the deformations that occur when going from source to target. The source image is the bubble at $t = 0$, and the target image is the bubble at $t = 1$. What happened between 0 and 1 ?

The highlights of this technique are:

- The technique based on differential flow estimation.
- Idea: Want to *flow* from the source image to reference image.
- The procedure is fully automated.
- Uses an affine model.
- Allows for intensity variations between the source and target images.

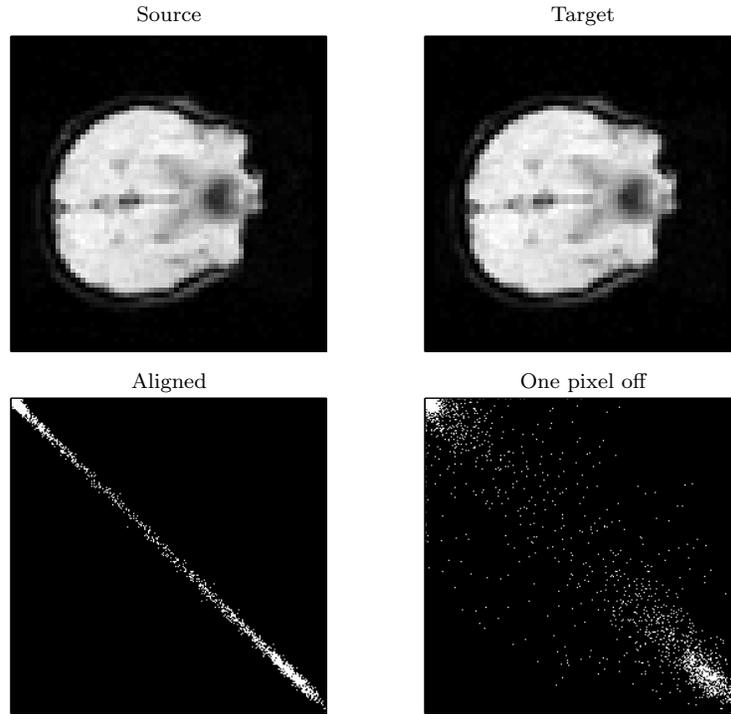


Figure 9. Joint histograms of different source and target images. While not strictly a diagonal line, the joint histogram of the aligned source and target images is relatively narrow (lower left). Translating by one pixel significantly disperses the diagonal (lower right).

Full details and results of the algorithm may be found in [30]. Since the model is very straightforward, we will delve a little deeper into this algorithm than we have so far with the previous algorithms discussed. It can be considered as an example of how, beginning with basic principles, a registration technique is born.

Our starting point is the general form of a 2-D affine transformation:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_5 \\ m_6 \end{bmatrix}$$

where x, y denote spatial coordinates in the source image and x_1, y_1 denote spatial coordinates in the target. Depending on the values m_1, m_2, m_3 and m_4 , certain well known geometric transformations can result (see Figure 10).

Now recall our description at the beginning of this section, that of a bubble rising through the water. We took two snapshots, one at $t = 0$, and one at $t = 1$, of the *same bubble*. Hence it is reasonable to have a single function, with temporal variable t , represent the bubble at time t .

With this in mind, let $f(x, y, t)$, $f(\hat{x}, \hat{y}, t - 1)$ represent the source and target images, respectively. To further simplify the model, at least for the moment, we

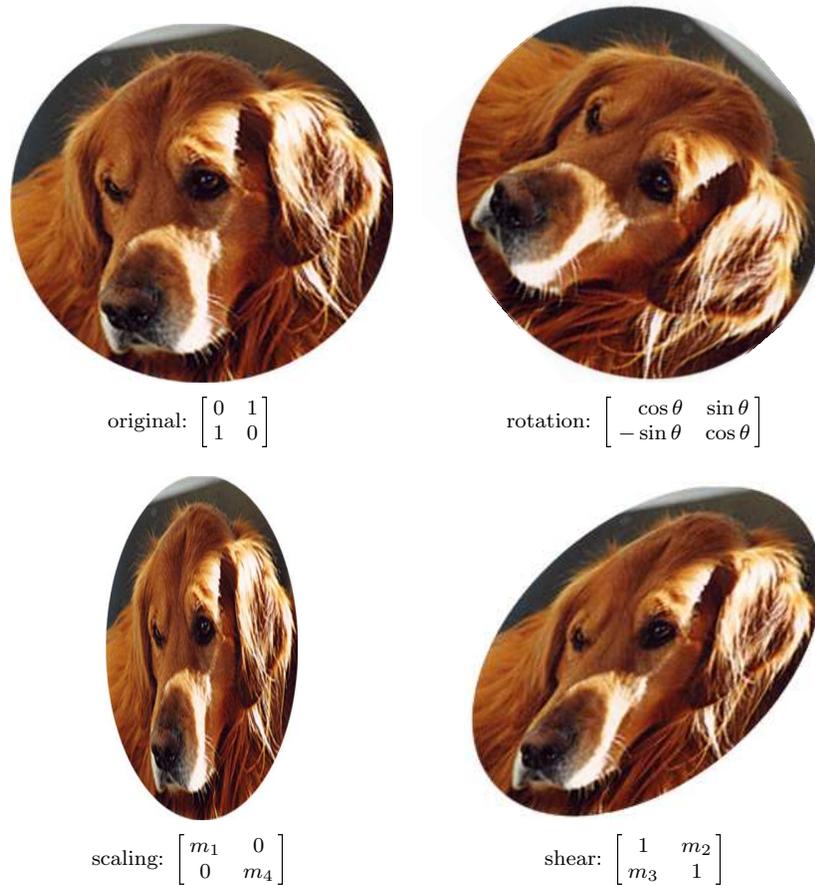


Figure 10. A smattering of linear transformations.

will make the “Brightness-Constancy” assumption: identical anatomical features in both images will have the same pixel intensity. That is, we are not allowing for the possibility that, say, the left eye in the MR source image to be brighter or darker than the left eye in the MR target image. Before tackling more difficult issues later, we want to ensure that only an affine transformation, and nothing else, is required to mold the source into the target.

Using the notation we have just introduced (which we will slightly abuse now), we have the situation:

$$f(x, y, t) = f(m_1x + m_2y + m_5, m_3x + m_4y + m_6, t - 1) \quad (3-1)$$

We use a least squares approach to estimate the parameters $\vec{m} = (m_1 \dots m_6)^T$ in (3-1). Now the function we *really* want to minimize is:

$$E(\vec{m}) = \sum_{x,y \in \Omega} (f(x, y, t) - f(m_1x + m_2y + m_5, m_3x + m_4y + m_6, t - 1))^2 \quad (3-2)$$

where Ω denotes the region of interest. However, the fact that $E(\vec{m})$ is not linear means that minimizing will be tricky. So we take an easy way out and instead take its truncated, first-order Taylor series expansion. Letting

$$\begin{aligned} k &= f_t + xf_x + yf_y, \\ \vec{c} &= (xf_x \quad yf_x \quad xf_y \quad yf_y \quad f_x \quad f_y)^T, \end{aligned} \quad (3-3)$$

where the subscripts denote partial derivatives, we eventually arrive at this much more reasonable error function:

$$E(\vec{m}) = \sum_{x,y \in \Omega} (k - \vec{c}^T \vec{m})^2. \quad (3-4)$$

To minimize (3-4), we differentiate with respect to \vec{m} :

$$\frac{dE}{d\vec{m}} = \sum_{\Omega} -2\vec{c}(k - \vec{c}^T \vec{m}),$$

set equal to 0, and solve for the model parameters to obtain:

$$\vec{m} = \left(\sum_{\Omega} \vec{c}\vec{c}^T \right)^{-1} \left(\sum_{\Omega} \vec{c}k \right). \quad (3-5)$$

And lo! we have determined \vec{m} . However, there is a caveat. We are assuming that the 6×6 matrix $(\sum_{\Omega} \vec{c}\vec{c}^T)$ in (3-5) is, in fact, invertible. We can usually guarantee this by making sure that the spatial region Ω is large enough to have sufficient image content, e.g., we would want some “interesting” features in Ω like edges, and not simply a “bland” area. The parameters \vec{m} are for the region Ω .

In terms of actual implementation, the parameters \vec{m} are estimated locally, for different spatial neighborhoods. By applying this algorithm in a multi-scale fashion, it is possible to capture large motions. (See [30] for details.) This is illustrated in Figure 11, in the case where the target image is a synthetically warped version of the source image.

Editorial. As an aside, we mention that doing an experiment such as this, registering an image with a warped version of itself is not altogether silly. If an algorithm being developed fails in an ideal test case such as this, chances are very good that it will fail for genuinely different images. However, to make a “fair” ideal test, the method of warping the image should be independent of the registration method. For example, if the registration algorithm is to determine an affine transform, do not warp the image using an affine transform. Use some other method, e.g., apply Bookstein’s thin-plate splines [9].

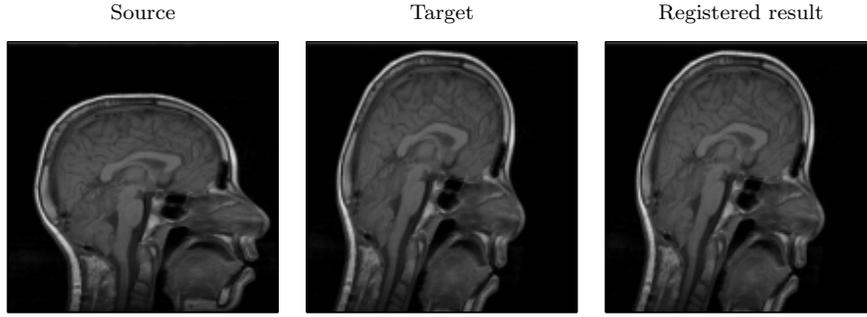


Figure 11. Flowing from source to target: An “ideal” experiment.

The optic flow model can next be modified to account for differences of contrast and brightness between the two images with the addition of two new parameters, m_7 for contrast, and m_8 for brightness. The new version of (3-1) is

$$m_7 f(x, y, t) + m_8 = f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t - 1). \quad (3-6)$$

We are also assuming that, in addition to the affine parameters, the brightness and contrast parameters are constant within small spatial neighborhoods.

Minimizing the least squares error as before, using a first-order Taylor series expansion, gives a solution identical in form to (3-5) except that this time

$$\begin{aligned} k &= f_t - f + x f_x + y f_y, \\ \vec{c} &= (x f_x \quad y f_x \quad x f_y \quad y f_y \quad f_x \quad f_y \quad -f \quad -1)^T; \end{aligned} \quad (3-7)$$

compare equations (3-3).

Now, we have been working under the assumption that the affine and contrast/brightness parameters are constant within some small spatial neighborhood. This introduces two conflicting conditions.

Recall $(\sum_{\Omega} \vec{c} \vec{c}^T)$. This matrix needs to have an inverse. As was mentioned earlier, this can be arranged by considering a large enough region Ω , i.e., a region with sufficient image content. However, the larger the area, the *less* likely it is that the brightness constancy assumption holds. Think about it: image content can be edges, and edges can have very different intensities, when compared with surrounding tissue.

Fortunately, the model can be modified one more time. Instead of a single error function (3-4), we can instead consider the sum of *two* errors:

$$E(\vec{m}) = E_b(\vec{m}) + E_s(\vec{m}) \quad (3-8)$$

where

$$E_b(\vec{m}) = (k - \vec{c}^T \vec{m})^2$$

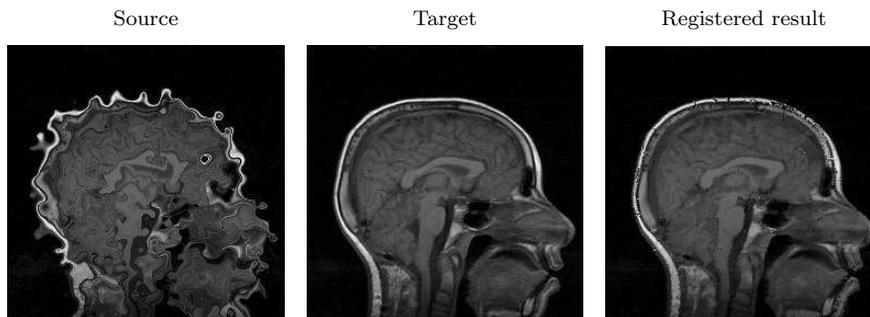


Figure 12. Registering an excessively distorted source image to a target image.

with k and \vec{c} defined as in (3-7) and (??), and

$$E_s(\vec{m}) = \sum_{i=1}^8 \lambda_i \left(\left(\frac{\partial m_i}{\partial x} \right)^2 + \left(\frac{\partial m_i}{\partial y} \right)^2 \right),$$

where λ_i is a positive constant, set by the user, that weights the smoothness constraint imposed on m_i .

As before, one works with Taylor series expansions of (3-8), but things become a little more complicated. Complete details of how to work with (3-8), as well with generalizations to 3-D, may be found in [30]. Some results are shown in Figures 12-13.

4. Conclusion

We have presented a whirlwind introduction to image registration for MRI. After providing a theoretical framework by which the problem is defined, we presented, in no particular order, a number of different algorithms. We then provided a more detailed discussion of an algorithm based on the idea of optic flow fields.

Our intent in this paper was to illustrate how the problem of image registration can have a wide variety of very dissimilar solutions. And there exist many more techniques than those presented here. For example, image features that some of these methods depend upon include surfaces [28; 15; 17], edges [27; 21], and contours [26; 35]. There are also methods based on B-splines [37; 22; 33], thin-plate splines [9; 10], and low-frequency discrete cosine basis functions [3; 4].

There are many survey articles the reader may wish to read, to learn more about medical image registration, In addition to those cited earlier ([11; 38]), we also call attention to [25; 24; 23; 40]. The simple existence of so many techniques provides more than sufficient support for the thesis that there are many paths to the One Truth: perfect image alignment.

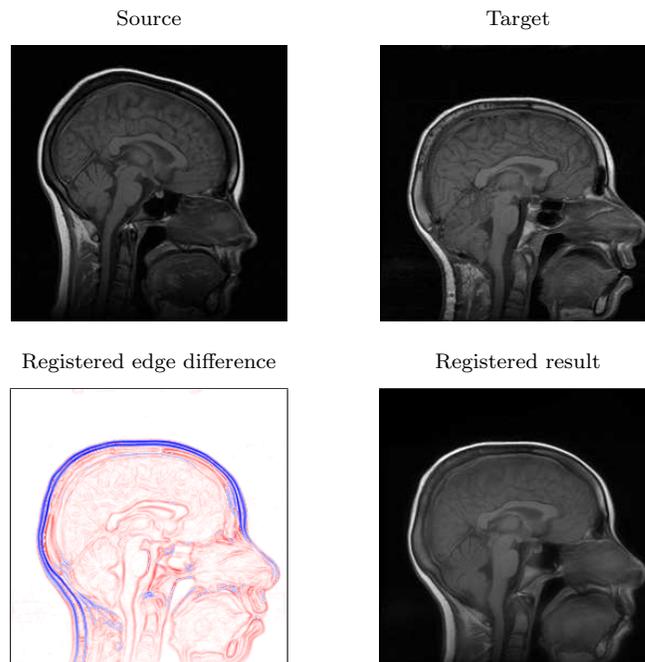


Figure 13. Registering two different clinical images. The lower left image shows how the edges of the registered source compare with the target's edges. The lower right image shows the registered source itself, after it has undergone both geometric *and* intensity-correction transformations.

References

- [1] The homepage for the AIR (“Automated Image Registration”) software package is <http://bishopw.loni.ucla.edu/AIR5/>.
- [2] N. Alpert, J. Bradshaw, D. Kennedy, and J. Correia, The principal axes transformation: a method for image registration, *J. Nuclear Medicine* **31** (1990), 1717–1722.
- [3] J. Ashburner and K. J. Friston, Multimodal image coregistration and partitioning: a unified framework, *NeuroImage* **6:3** (1997), 209–217.
- [4] J. Ashburner, P. Neelin, D. L. Collins, A. C. Evans and K. J. Friston, Incorporating prior knowledge into image registration, *NeuroImage* **6** (1997), 344–352.
- [5] Peter A. Bandettini, Eric C. Wong, R. Scott Hinks, Ronald S. Tikofsky, and James S. Hyde, Time course EPI of human brain function during task activation, *Magnetic Resonance in Medicine* **25** (1992), 390–397.
- [6] P. A. Bandettini, E. C. Wong, J. R. Binder, S. M. Rao, A. Jesmanowicz, E. Aaron, T. Lowry, H. Forster, R. S. Hinks, J. S. Hyde, Functional MRI using the BOLD approach: Dynamics and data analysis techniques, pp. 335–349 in *Perfusion and Diffusion: Magnetic Resonance Imaging*, edited by D. LeBihan and B. Rosen, Raven Press, New York, 1995.
- [7] M. Betke, H. Hong, and J. P. Ko, Automatic 3D registration of lung surfaces in computed tomography scans, pp. 725–733 *Fourth International Conference on Medical*

- Image Computing and Computer-Assisted Intervention*, Utrecht, The Netherlands, October 2001.
- [8] Amanda Bischoff-Grethe, Shawnette M. Proper, Hui Mao, Karen A. Daniels, and Gregory S. Berns, Conscious and unconscious processing of nonverbal predictability in Wernicke's Area, *Journal of Neuroscience* **20**:5 (March 2000), 1975–1981.
 - [9] F. L. Bookstein, Principal Warps: Thin-plate splines and the decomposition of deformations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**:6 (June 1989), 567–585.
 - [10] F. L. Bookstein, Thin-plate splines and the atlas problem for biomedical images, *Information Processing in Medical Imaging*, July 1991, 326–342.
 - [11] Leslie G. Brown, A Survey of Image Registration Techniques, *ACM Computing Surveys*, **24**:4 (December 1992), 325–376.
 - [12] E. D. Castro and C. Morandi, Registration of translated and rotated images using finite Fourier transforms, *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-9** (1987), 700–703.
 - [13] G.E. Christensen, R.D. Rabbit, and M.I. Miller, A deformable neuroanatomy textbook based on viscous fluid mechanics, pp. 211–216 in *Proceedings of the 1993 Conference on Information Sciences and Systems*, Johns Hopkins University, March 1995.
 - [14] A. Collignon and F. Maes and D. Delaere and D. Vandermeulen, P. Suetens and G. Marchal, Automated multimodality image registration using information theory, pp. 263–274 in *Information Processing in Medical Imaging*, edited by Y. Bizais, C. Barillot, and R. Di Paolo, Kluwer, Dordrecht, 1995.
 - [15] A. M. Dale, B. Fischl, and M. I. Sereno, Cortical surface-based analysis, I: Segmentation and surface reconstruction, *NeuroImage* **9**:2 (Feb 1999), 179–194.
 - [16] R. O. Duda and P.E. Hart, *Pattern classification and scene analysis*, Wiley, New York, 1973.
 - [17] B. Fischl, M. I. Sereno, and A. M. Dale, Cortical surface-based analysis, II: Inflation, flattening, and a surface-based coordinate system, *NeuroImage* **9**:2 (Feb. 1999), 195–207.
 - [18] R. S. J. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, and J. C. Mazziotta, *Human Brain Function*, Academic Press, San Diego, 1997.
 - [19] J. R. Hurley and R. B. Cattell, The PROCUSTES program: Producing direct rotation to test a hypothesized factor structure, *Behav. Sci.* **7** (1962), 258–262.
 - [20] P. Janata, B. Tillman, and J. J. Bharucha, Listening to polyphonic music recruits domain-general attention and working memory circuits, *Cognitive, Affective, and Behavioral Neuroscience*, **2**:2 (2002), 121–140.
 - [21] W. S. Kerwin and C. Yuan, Active edge maps for medical image registration, pp. 516–526 in *Proceedings of SPIE – The International Society for Optical Engineering*, July 2001.
 - [22] P. J. Kostelec, J. B. Weaver, and D. M. Healy, Jr., Multiresolution Elastic Image Registration, *Medical Physics*, **25**:9 (1998), 1593–1604.
 - [23] H. Lester and S. Arridge, A Survey of Hierarchical Non-linear Medical Imaging Registration, *Pattern Recognition* **32**:1 (1999), 129–149.

- [24] J. B. Antoine Maintz and Max A. Viergever, A Survey of Medical Image Registration, *Medical Image Analysis* **2**:1 (1998), 1–36.
- [25] C. R. Maurer Jr. and J. M. Fitzpatrick, A Review of Medical Image Registration, chapter in *Interactive Image-Guided Neurosurgery*, American Association of Neurological Surgeons, Park Ridge, IL, 1993.
- [26] G. Medioni and R. Nevatia, Matching images using linear features, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**:6 (Nov. 1984), 675–685.
- [27] M. L. Nack, Rectification and registration of digital images and the effect of cloud detection, pp. 12–23 in *Machine Processing of Remotely Sensed Data*, West Lafayette, IN, June 1977.
- [28] C. A. Pelizzari, G. T. Y. Chen, D. R. Spelbring, R. R. Weichselbaum, and C. T. Chen, Accurate three-dimensional registration of CT, PET and/or MR images of the brain, *J. Computer Assisted Tomography*, **13**:1 (1989), 20–26.
- [29] Senthil Periaswamy, www.cs.dartmouth.edu/~sp.
- [30] Senthil Periaswamy and Hany Farid, Elastic registration in the presence of intensity variations, to appear in *IEEE Transactions in Medical Imaging*.
- [31] S. Periaswamy, J. B. Weaver, D. M. Healy, Jr., and P. J. Kostelec, Automated multiscale elastic image registration using correlation, pp. 828–838 in *Proceedings of the SPIE – The International Society for Optical Engineering* **3661**, 1999.
- [32] William K. Pratt, *Digital signal processing*, Wiley, New York, 1991.
- [33] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, Non-rigid registration using free-form deformations: Application to breast MR images, *IEEE Trans. Medical Imaging*, **18**:8 (August 1999), 712–721.
- [34] P. H. Schonemann, A generalized solution of the orthogonal Procrustes problem, *Psychometrika*, **31**:1 (1966), 1–10.
- [35] Wen-Shiang V. Shih, Wei-Chung Lin, and Chin-Tu Chen, Contour-model-guided nonlinear deformation model for intersubject image registration, pp. 611–620 in *Proceedings of SPIE - The International Society for Optical Engineering* **3034**, April 1997.
- [36] Arthur W. Toga and John C. Mazziotta (eds.), *Brain mapping: the methods*, Academic Press, San Diego, 1996.
- [37] M. Unser, A. Aldroubi and C. Gerfen, A multiresolution image registration procedure using spline pyramids, pp. 160–170 *Proceedings of the SPIE – Mathematical Imaging: Wavelets and Applications in Signal and Image Processing* **2034**, 1993.
- [38] P. A. Van den Elsen, E. J. D. Pol, M. A. Viergever, Medical Image Matching - a review with classification, *IEEE Engineering in Medicine and Biology* **12**:1 (1993), 26–39,
- [39] P. Viola and W. M. Wells, III, Alignment by maximization of mutual information, pp. 16–23 in *International Conf. on Computer Vision*, IEEE Computer Society Press, 1995.
- [40] J. West, J. Fitzpatrick, M. Wang, B. Dawant, C. Maurer, R. Kessler, and R. Maciunas, Comparison and evaluation of retrospective intermodality image registration techniques, pp. 332–347 *Proceedings of the SPIE - The International Society for Optical Engineering*, Newport Beach, CA., 1996.

- [41] R. P. Woods, S. R. Cherry, and J. C. Mazziotta, Rapid automated algorithm for alignment and reslicing PET images, *J. Computer Assisted Tomography* **16** (1992), 620–633.
- [42] R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, and J. C. Mazziotta, Automated image registration, I: General methods and intrasubject, intramodality validation, *J. Computer Assisted Tomography* **22** (1998), 141–154.
- [43] R. P. Woods, J. C. Mazziotta, and S. R. Cherry, MRI-PET registration with automated algorithm, *J. Comp. Assisted Tomography*, **17**:4 (1993) 536–546.

PETER J. KOSTELEK
DEPARTMENT OF MATHEMATICS
DARTMOUTH COLLEGE
HANOVER, NH 03755
UNITED STATES
geelong@cs.dartmouth.edu

SENTHIL PERIASWAMY
DEPARTMENT OF COMPUTER SCIENCE
DARTMOUTH COLLEGE
HANOVER, NH 03755
UNITED STATES
sp@cs.dartmouth.edu