Efficient Least Squares Multimodal Registration With a Globally Exhaustive Alignment Search

Jeff Orchard, *Member, IEEE*

***Abstract*—There are many image registration situations in which the initial misalignment of the two images is large. These registration problems, often involving comparison of the two images only within a region of interest (ROI), are difficult to solve. Most intensity-based registration methods perform local optimization of their cost function and often miss the global optimum when the initial misregistration is large. The registration of multimodal images makes the problem even more difficult since it limits the choice of available cost functions. We have developed an efficient method, capable of multimodal rigid-body registration within an ROI, that performs an exhaustive search over all integer translations, and a local search over rotations. The method uses the fast Fourier transform to efficiently compute the sum of squared differences cost function for all possible integer pixel shifts, and for each shift models the relationship between the intensities of the two images using linear regression. Test cases involving medical imaging, remote sensing and forensic science applications show that the method consistently brings the two images into close registration so that a local optimization method should have no trouble fine-tuning the solution.**

***Index Terms*—Image registration, least squares, multimodal registration, Fourier.**

I. Introduction

INTENSITY-BASED registration involves a cost function and an optimization strategy. The cost function is chosen so that its optimum value is achieved when the two images are properly registered (for an overview of image registration methods, see [1]). The optimization strategy is the method by which one seeks this optimal solution (by adjusting the registration parameters). In the design of automatic intensity-based registration algorithms, one of the major stumbling-blocks is the challenge of registering images when their initial misregistration is large. We loosely define a misregistration as *large* when a local optimization strategy (e.g., gradient descent) will fall toward a local optimum that is not the correct global optimum.

When only a subset of an image is appropriate for comparison, one can use a region of interest (ROI) to limit the extent over which the cost function is computed. For example, two images may have substantially different fields of view and overlap by only a small amount, or perhaps one image contains additional objects that are not present in the other image. A misregistration could be considered large if the ROI and its corresponding region in the other image do not initially overlap. If the images are taken from different modalities, the registration problem becomes even more difficult because the choice of appropriate cost functions is limited.

This kind of multimodal image registration is used in medical imaging [2]. To illustrate the problem, consider the scenario where a patient has a computed tomography (CT) scan of their head, then subsequently develops a brain tumor. Further diagnostic imaging might include the acquisition of a magnetic resonance (MR) image. Registering the CT to the MR image could help in surgical planning by determining the proximity of the tumor to bony landmarks. However, the presence of the tumor in the MRI can cause an automatic method to converge to an incorrect registration solution. If an ROI that excludes the tumor is selected from the MR image, then we start to run into the problem of large misregistrations (large relative to the ROI, at least). Other medical imaging scenarios might require one to register two images that have a vastly different field of view, suggesting the use of an ROI to constrain the comparison to a region of the overlap.

Another tricky registration scenario is mosaicking, stitching together overlapping images to render a scene with an extended field of view. Mosaicking is common in remote sensing [3] and is also used in the inspection of circuit boards [4]. Adjacent images overlap, but the exact amount of the overlap is not known. If the range of uncertainty is too large, we again run into the issue of a small ROI and a (relatively) large misregistration.

Applications of image processing in forensic science can also be plagued by large misregistration issues. It was recently proposed that image matching could be used to aid in the identification of human remains by finding the best match of a postmortem (after death) dental X-ray in a database of antemortem (before death) X-rays [5], [6]. One tooth (or a number of teeth) of a postmortem X-ray can be outlined to define the ROI. This image can then be registered to antemortem X-rays of candidate matches. The image that best matches the postmortem image within the ROI may indicate a positive identification. The registration cannot be performed over the entire image since the postmortem image might be missing teeth that could be present in the corresponding antemortem images, and the distance between the upper and lower teeth could be different for different X-ray sessions.

In all of the above scenarios, the initial misregistration could be large. Unfortunately, the vast majority of intensity-based image registration methods are based on the local optimization

Manuscript received September 13, 2006; revised June 25, 2007. This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and in part by the Canada Foundation for Innovation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hassan Foroosh.

J. Orchard is with the David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1 Canada (e-mail: jorchard@cs.uwaterloo.ca).

Digital Object Identifier 10.1109/TIP.2007.904956

of a cost function. The images must be nearly aligned, and typically a gradient-descent type method is employed to find a local minimum (or maximum, as the case may be). If the two images are initially misregistered by too much, the optimization scheme will fall into a local basin of attraction that does not contain the global minimum; the resulting solution does not give the correct registration. This issue becomes more pronounced as the ROI gets progressively smaller.

In this paper, we focus our attention on the difficult problem of globally optimal multimodal registration. We wish to produce an efficient method that performs an exhaustive search over at least part of the motion parameter space. Our goal is to bring the two images into close enough alignment that the fine-tuning of the registration can be achieved by a local optimization strategy.

A. Previous Work

Other researchers have solved pieces of the problem within this framework. Methods have been developed to perform multimodal registration on an ROI, but can only find the local optimum. Other methods have been developed to perform global registration, but only for image pairs that contain the same field of view, thus ruling out any use of an ROI. Finally, some methods perform global registration on an ROI, but are not appropriate for multimodal registration. The relevant methods in all three cases are outlined below. To our knowledge, however, no one has developed a global multimodal registration method that incorporates an ROI.

First, let us look at global registration methods. It is well known that global rigid-body registration can be achieved by decoupling the rotation and translation components in the frequency domain [7]–[10]. A rotation in the spatial (image) domain corresponds to the same rotation in the frequency domain. By looking at the magnitudes of the Fourier coefficients in the frequency domain, the optimal rotation can easily be found; when representing the frequency domain in polar coordinates, the problem amounts to finding the best shift along the θ axis. Once the rotation is corrected, the optimal translation can be determined using the phase correlation technique [11], [12]. This method finds the best shift by taking the Fourier transform of both images, dividing their Fourier coefficients on an element-by-element basis, and then computing the inverse Fourier transform on the result. If the two images are shifted versions of one another, this final image should contain a single impulse spike at the location corresponding to the best shift. Another approach to global rigid-body registration is to apply a log-polar coordinate system transformation [13]. If the centre of rotation is chosen the same in both images, then the rotation and scaling can be determined easily (since each manifests itself as a shift in the log-polar space). The problem is that there seems to be no easy way to choose the correct centre of rotation. To address the issue, Wolberg *et al.* [13] combine a multiresolution approach with an exhaustive search over all possible image centres to achieve good registration results. However, the theory underlying all of these methods assumes that the images are identical, except for rigid-body motion (although some also incorporate scaling [10], [13]). In particular, these methods (with the possible exception of [13]) do not work when registering to an ROI. Furthermore, these methods do not work for multimodal registration.

A number of methods have been proposed to tackle multimodal image registration. The mutual information cost function is currently a popular choice [14], [15], its advantage being that it makes no assumption about the intensity correspondences between the two images. Unfortunately, there is no known way to efficiently perform an exhaustive search over parameter space, so all implementations rely heavily on local optimization techniques. A multiresolution and regular sampling method was developed to handle a wide range of medical image registration problems [16]. Later in this paper, however, we show that the method achieves little success when the initial misregistration is large compared to the size of the ROI. Some other registration methods attempt to model the relationship between the intensities in the two images as part of the registration process [17]–[19]. While these methods still require the images to be nearly registered, their approach is closer to the method we propose here.

The sum of squared differences (SSD) cost function, sometimes referred to as "least squares," has been used widely in image registration [9], [17]–[24]. With only a few exceptions [17]–[19], the SSD cost function is not generally used in multimodal image registration. The advantage of the SSD is that it is a correlation-based cost function, and can be evaluated efficiently for all possible integer translations using the fast Fourier transform (FFT) [25]. Using the FFT affords speedups (over direct evaluation of the cost function) of 50 to 500 times [21]. This approach has been implemented in a number of studies where correlation-based cost functions were used [21], [26]–[29].

In this paper, we present a method based on the SSD cost function that consistently returns the rigid-body transformation parameters that are close to the true values. This is done in the context of relatively small ROIs, and for both monomodal and multimodal image pairs. Our only limitations are that the method only considers integer shifts, and that the rotation between the images must be small (preferably less than 5°). In our opinion, these are not major stumbling blocks, since a local optimization method can subsequently be used to obtain subpixel accuracy, and the method can be rerun with several different starting angles, and the best solution chosen. Such a multistart strategy is also used in [16].

II. Theory

For notational brevity, we start by outlining the theory for aligning two 1-D signals of length N rather than images or volumes.

A. Exhaustive Alignment Search

The goal of alignment is to find the optimal shift for the signal f so that it spatially corresponds to the signal g. Moreover, if f and g have only a partial overlap, then the alignment should be over a subset of each of the images. In that case, we include a weighting function over g that assigns a value in the range [0, 1] to each sample in g. This weighting function (also referred to as an "alpha map" in [27]) essentially selects an ROI over which the comparison to f is to take place. We will refer to this

weighting function as w. Finding the optimal alignment is then a matter of finding the shift x to apply to f that minimizes

$$E(x; f, g, w) = \sum_{n=1}^{N} [f_{n-x} - g_n]^2 w_n. \tag{1}$$

Our strategy for optimizing the cost function in (1) is to exhaustively compute it for all integer shifts x, and simply choose the shift that yields the minimum value. There is a technical reason we compute the cost for integer values of x, as will be outlined below. However, since our goal is to find an approximate registration solution, it is enough to consider only integer shifts. Hence, for the remainder of this paper, we will only concern ouselves with integer values of x. As outlined in [21] and [27], expanding the brackets in (1) leads to

$$\sum_n f_{n-x}^2 w_n - 2\sum_n f_{n-x} g_n w_n + \sum_n g_n^2 w_n. \tag{2}$$

The last term in (2) is constant with respect to x, so is irrelevant to the optimization problem (although it can be computed cheaply). The summation in the second term can be reformulated into a convolution

$$\sum_n \bar{f}_{x-n}(g_n w_n) = (\bar{f} \star (gw))_x \tag{3}$$

where $\bar{f}$ is a reflected version of f (i.e., $\bar{f}_n = f_{-n}$). This convolution can be evaluated efficiently using the FFT. Let $\mathfrak{F}\{\cdot\}$ represent the Fourier transform. Then

$$(\bar{f} \star (gw))_x = \mathfrak{F}^{-1}\{\mathfrak{F}\{\bar{f}\}\, \mathfrak{F}\{gw\}\}_x. \tag{4}$$

That is, the convolution can be computed for all integer values of x by an element-wise multiplication of the Fourier coefficients of $\bar{f}$ and gw, followed by an inverse FFT. Note that $\mathfrak{F}\{\bar{f}\}$ is equivalent to $\overline{\mathfrak{F}\{f\}}$, the complex conjugate of $\mathfrak{F}\{f\}$. This method for computing the convolution implicitly assumes that f is periodic, so that as f shifts, any image content that gets moved out of the field of view on one side of the image enters the field of view on the other side of the image. A similar approach can be used to compute the first term in (2). The advantage of using the FFT in these computations is that it reduces the computational complexity of evaluating (1) for all integer shifts from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$ [25]. These convolution terms comprise the bulk of the processing. The FFT makes it feasible to find the global optimum for much larger signals (including images, volumes, etc.) than would be possible using direct evaluation of (1).

B. Linear Regression

The SSD cost function in (1) is not directly suitable for multimodal registration because it assumes that the two images have the same intensity mappings (i.e., that corresponding regions in the two images are represented by the same intensities). This is not the case for multimodal registration. For example, the bone in a CT image is bright, while bone is dark in an MR image.

Instead of using the SSD as shown in (1), we can simultaneously incorporate an arbitrary linear model in place of f, and easily compute the optimal linear coefficients using linear regression. To do this, we replace f with $\mathbf{h}^{\mathrm{T}}(f)\mathbf{s}$ in (1), where $\mathbf{h}^{\mathrm{T}}(f)$ is a $1 \times P$ row-vector of regressors that depend on f, and $\mathbf{s}$ is a $P \times 1$ column vector containing the corresponding coefficients. For example, if we want to perform the registration while considering all possible brightness and contrast adjusted versions of f (in other words, considering all images that arise from a linear remapping of the intensities in f), then we define

$$\mathbf{h}^{\mathrm{T}}(f) = [1\ f] \quad \text{and} \quad \mathbf{s} = \begin{bmatrix} s_0 \\ s_1 \end{bmatrix} \tag{5}$$

where s_0 is the brightness adjustment, and s_1 is the contrast adjustment. Then our cost function becomes

$$E(x, \mathbf{s}; \mathbf{h}, g, w) = \sum_n [\mathbf{h}^{\mathrm{T}}(f_{n-x})\mathbf{s} - g_n]^2 w_n. \tag{6}$$

Notice that now our cost function depends not only on the shift x, but also on the linear coefficients $\mathbf{s}$. To optimize this cost function, we follow the same process as before and expand the brackets in (6) and get

$$\sum_n \left[\mathbf{s}^{\mathrm{T}} \mathbf{h}_{n-x} \mathbf{h}_{n-x}^{\mathrm{T}} \mathbf{s} w_n - 2\mathbf{h}_{n-x}^{\mathrm{T}} \mathbf{s} g_n w_n + g_n^2 w_n\right] \tag{7}$$

where we have replaced $\mathbf{h}(f_{n-x})$ with $\mathbf{h}_{n-x}$ for notational simplicity. Factoring $\mathbf{s}$ out of the summations, and omitting the last term since it does not depend on x or $\mathbf{s}$, gives the equivalent cost function

$$\begin{aligned} &\tilde{E}(x, \mathbf{s}; \mathbf{h}, g, w) \\ &\quad = \mathbf{s}^{\mathrm{T}} \left[\sum_n \mathbf{h}_{n-x} \mathbf{h}_{n-x}^{\mathrm{T}} w_n\right] \mathbf{s} - 2\left[\sum_n \mathbf{h}_{n-x}^{\mathrm{T}} g_n w_n\right] \mathbf{s} \\ &\quad = \mathbf{s}^{\mathrm{T}} \mathbf{A}_x \mathbf{s} - 2\mathbf{b}_x^{\mathrm{T}} \mathbf{s} \end{aligned} \tag{8}$$

where $\mathbf{A}_x$ is the $P \times P$ matrix $\sum_n \mathbf{h}_{n-x} \mathbf{h}_{n-x}^{\mathrm{T}} w_n$ and $\mathbf{b}_x^{\mathrm{T}}$ is the $1 \times P$ row vector $\sum_n \mathbf{h}_{n-x}^{\mathrm{T}} g_n w_n$. It is important to note that each different shift x produces a different $\mathbf{A}_x$ and $\mathbf{b}_x$. For example, the ijth element of the matrix $\mathbf{A}_x$ is

$$[\mathbf{A}_x]_{ij} = \sum_{n=1}^{N} \mathbf{h}_i(f_{n-x}) \mathbf{h}_j(f_{n-x}) w_n.$$

Finding the least-squares solution of the quadratic form in (8) can be achieved by solving the linear system

$$\mathbf{A}_x \mathbf{s} = \mathbf{b}_x \tag{9}$$

for $\mathbf{s}$ [30]. Our strategy is to compute each element of $\mathbf{A}_x$ and $\mathbf{b}_x$ for all shifts using the FFT method described above, and then use those values to solve (9) one candidate shift at a time. As a result, we exhaustively compute the optimal $\mathbf{s}$ for every shift. Using these optimal $\mathbf{s}$-values, we then evaluate the cost function for every shift. It is worth noting that the computational complexity of solving a $P \times P$ linear system is generally $\mathcal{O}(P^3)$, so to register two signals f and g of length N involves solving N such systems, resulting in a total operation count of $\mathcal{O}(P^3 N)$. Also, computing each entry of $\mathbf{A}_x$ and $\mathbf{b}_x$ for all possible shifts takes $\mathcal{O}(N \log N)$. The matrix $\mathbf{A}_x$ has P^2 entries, and $\mathbf{b}_x$ has P entries. Hence, computing $\mathbf{A}_x$ and $\mathbf{b}_x$ for all possible shifts takes $\mathcal{O}((P^2 + P) N \log N)$ operations.

Putting it all together, finding the globally optimal shift takes $\mathcal{O}(P^3N+(P^2+P)N\log N)$ floating-point operations (flops).

The method described above can trivially be extended to 2-D and higher. If f and g are both $M\times N$ images, the computational complexity becomes $\mathcal{O}(P^3MN + (P^2 + P)MN\log MN)$. However, this only finds the best shift (in the horizontal and vertical directions), and does not consider any rotations.

C. Rotation

We can incorporate small rotations using a linear approximation. If $R(f,\theta)$ represents the image f rotated by θ, then its linear approximation is

$$R(f,\theta) \approx R(f,0) + \frac{\partial f}{\partial \theta}\theta \;=\; f + \frac{\partial f}{\partial \theta}\theta$$

where $\partial f/\partial\theta$ is an image that contains the partial derivative of f with respect to rotation. This rotation component can be implemented using an additional regressor. Suppose we want to add rotation to the brightness/contrast adjusted model in (5). Replacing f with our approximation to $R(f,\theta)$, and multiplying $\mathbf{h}^{\mathrm{T}}(f)$ by $\mathbf{s}$ (a dot-product), we get

$$\begin{aligned} & s_0 + s_1 R(f,\theta) \\ &= s_0 + s_1\left(f + \frac{\partial f}{\partial \theta}\theta\right) && (10)\\ &= s_0 + s_1 f + s_2\frac{\partial f}{\partial \theta} && (11)\end{aligned}$$

where $s_2 = s_1\theta$. Hence, we can incorporate small rotations using the regressors $\mathbf{h}^{\mathrm{T}}(f) = [1\ f\ (\partial f/\partial\theta)]$ and computing θ by s_2/s_1. A locally optimal solution to the nonlinear problem can be computed by solving the linear problem in a fixed-point iteration framework.

It should be noted that this linear rotation component becomes more difficult to resolve if some of the regressors are nonlinear functions of f (for example, $\mathbf{h}_i(f) = f^2$ or $\mathbf{h}_i(f) = \cos(f)$). In such cases, the relationship between θ and the associated regression coefficients ($\mathbf{s}$-values) becomes nonlinear. Moreover, combinations of regressors can yield multiple values for θ; this is the case when using more than one nonconstant regression term, since each regressor will include its own linear rotation component. We see no built-in way to force all the estimates for θ to agree (see the second-last paragraph in Section III-A for an example).

III. Methods

A. Implementation

Based on the above theory, we implemented a program in C++ that takes a reference image g and weighting image w, as well as some regressors based on an image f, and finds the optimal shift, regression parameters, and small rotation that give the smallest cost function value. Algorithm 1 lists the pseudocode for efficiently evaluating the cost function using the regressors $[1\ f\ (\partial f/\partial\theta)]$.

The program was implemented on an Apple Macintosh Dual 2.5-GHz PowerPC G5 workstation with 8 GB of RAM, and running Mac OS X. A number of optimization measures were taken to speed up the code. Apple's AltiVec library of vectorized numerical routines is used for some simple image manipulations such as element-wise addition and multiplication of images. We use LAPACK [31] to solve the systems of linear equations in (9). We also use the FFTW library [32] to compute the FFTs. Finally, since our images are real-valued, we take advantage of the conjugate symmetry in the frequency domain by computing a half-sized FFT in one of the two dimensions, cutting the computational cost roughly in half. However, none of the code was optimized to run on a multiprocessor machine. In fact, one processor was disabled for the timing tests.

Algorithm I: SSD with linear remapping and rotation

1: input: g, w, f
2: assign $\mathbf{h}_1 = 1$, $\mathbf{h}_2 = f$, and $\mathbf{h}_3 = (\partial f/\partial\theta)$
3: precompute $\overline{\mathfrak{F}\{\mathbf{h}_i\mathbf{h}_j\}}$ for $i,j = 1,2,3$
4: precompute $\overline{\mathfrak{F}\{\mathbf{h}_i\}}$ for $i = 1,2,3$
5: initialize the cumulative rotation, $\Theta = 0$
6: **repeat**
7: counter-rotate $g' = R(g,-\Theta)$ and $w' = R(w,-\Theta)$
8: find bounding box for nonzero part of w'
9: shift g' and w' equally so that the bounding box is in the upper-left corner
10: pad/crop the shifted g' and w' so that they are the same size as f
11: compute $\mathfrak{F}\{w'\}$ and $\mathfrak{F}\{g'w'\}$
12: compute $[\mathbf{A}_{x,y}]_{i,j} = \mathfrak{F}^{-1}\{\overline{\mathfrak{F}\{\mathbf{h}_i\mathbf{h}_j\}}\mathfrak{F}\{w'\}\}_{x,y}$, $i,j = 1,2,3$ for all integer shifts (x,y) simultaneously (Note that $\mathbf{A}_{x,y}$ is symmetric.)
13: compute $[\mathbf{b}_{x,y}]_i = \mathfrak{F}^{-1}\{\overline{\mathfrak{F}\{\mathbf{h}_i\}}\mathfrak{F}\{g'w'\}\}_{x,y}$, $i = 1,2,3$, for all (x,y)
14: **for** each valid shift (x,y) **do**
15: solve $\mathbf{A}_{x,y}\mathbf{s} = \mathbf{b}_{x,y}$ for $\mathbf{s}$
16: $C(x,y) = \mathbf{s}^{\mathrm{T}}\mathbf{A}_{x,y}^{\mathrm{T}}\mathbf{A}_{x,y}\mathbf{s} - 2\mathbf{b}_{x,y}^{\mathrm{T}}\mathbf{s}$
17: **if** $C(x,y)$ is the lowest so far **then**
18: record $(x^*,y^*) = (x,y)$ and $\mathbf{s}^* = \mathbf{s}$
19: **end if**
20: **end for**
21: adjust (x^*,y^*) to compensate for the initial shift of g' and w' (see step 9)
22: compute $\theta = \mathbf{s}_2^*/\mathbf{s}_1^*$
23: increment $\Theta = \Theta + \theta$
24: **until** $|\theta| < 0.5°$, or completed more than 6 iterations
25: the solution is (x^*,y^*), Θ

Our implementation skips candidate translations if the ROI is not entirely inside the image f. The FFT-based convolutions intrinsically assume that the images are periodic, and afford the ability to consider wrapped versions of f (where f is shifted such that it only partially overlaps with the ROI). The portion of f inside the ROI for such a shift would actually contain spatially-disperate regions of f on each side of the wrapping seam. We chose to exclude these cases from our set of valid shifts be-

cause they were not appropriate to the medical, forensic and remote-sensing applications cited in this paper. While we have no choice but to compute $\mathbf{A}_{x,y}$ and $\mathbf{b}_{x,y}$ for all shifts (valid or otherwise), we can easily avoid solving (9) for invalid shifts.

During preliminary tests of the method, we found that a two-pronged approach worked best, combining the results from two different SSD applications. One approach uses the gradient magnitude of the images, while the other uses a piecewise-linear function to model the relationship between the intensities of the two images. Define $f_\bigtriangledown = \| \bigtriangledown f\|$ as the magnitude of the gradient of f, (similarly define $g_\bigtriangledown$). Then our combined SSD cost function is the weighted sum of the two SSD cost functions

$$C(x,y,\mathbf{s},\theta) = \frac{1}{\gamma}\tilde{E}(x,y,\theta,\mathbf{s};\mathbf{h}_{\mathrm{GM}}(f_\bigtriangledown),g_\bigtriangledown,w) + \frac{1}{\tau}\tilde{E}(x,y,\theta,\mathbf{s};\mathbf{h}_{\mathrm{PL}}(f),g,w) \quad (12)$$

where (x,y) represents the shift vector, γ and τ are relative weighting coefficients, $\mathbf{h}_{\mathrm{GM}}(f_\bigtriangledown)$ is an array of two regressors made up of $f_\bigtriangledown$ and its rotational derivative, $\partial f_\bigtriangledown/\partial\theta$, and $\mathbf{h}_{\mathrm{PL}}(f)$ is an array of nine regressors

$$\mathbf{h}_{\mathrm{PL}}(f) = \left[1^{(1)}\left|f^{(1)}\right|\frac{\partial f^{(1)}}{\partial\theta}\;\middle|\;1^{(2)}\middle|f^{(2)}\middle|\frac{\partial f^{(2)}}{\partial\theta}\right|\cdots 1^{(3)}\left|f^{(3)}\right|\frac{\partial f^{(3)}}{\partial\theta}\right] \quad (13)$$

representing a piecewise-linear function with three pieces (see Fig. 7). The symbol $f^{(i)}$ is a masked version of f where all but the pixels with intensities in the range of the ith piece are set to zero

$$f_n^{(i)} = \begin{cases} f_n & \text{if, } 85(i-1) \le f_n < 85i \\ 0, & \text{otherwise} \end{cases} \quad \text{for } i = 1,2,3. \quad (14)$$

That is, each piece is implemented by setting all pixels outside the corresponding intensity range to zero. The symbol $1^{(i)}$ represents a "unit" regressor of ones, masked by the same mask used on $f^{(i)}$. The three pieces were chosen by splitting up the intensity range of f into three equal parts. More inspired methods for determining the break points might include studying the histogram of g, or using prior knowledge if it is available. However, these options have not been examined in this paper.

Once the two SSD components are evaluated for all shifts (x,y) (and their corresponding optimal $\mathbf{s}$-values are stored for each shift), they are combined using a weighted sum so that the minimum value of each SSD is scaled to one (clearly, this assumes that the minimum before scaling is nonzero). Thus, in the combined cost function, γ is the minimum computed value of the gradient magnitude SSD, τ is the minimum computed value of the piecewise-linear SSD, and the resulting lowest possible value of the combined cost function is two. If the minima of the two SSD components occur at different shifts, then the minimum of the combined SSD will be greater than two. The minimum of the combined SSD indicates the optimal shift, which in turn dictates the optimal values for the two sets of regression coefficients (the $\mathbf{s}$-values). Notice that once values are assigned to these regression coefficients, they can yield four different estimates for θ (one for each of the regressors $\partial f^{(1)}/\partial\theta, \partial f^{(2)}/\partial\theta, \partial f^{(3)}/\partial\theta$, and $\partial f_\bigtriangledown/\partial\theta$). In our implementation, we use the median value of these four candidate values as our estimate for θ.

As mentioned earlier, we use a linear approximation for small rotations, and solve the linear problem within a fixed-point iteration scheme. In particular, we continue iterating until the rotation increment, θ, is less than 0.5°, to a maximum of six iterations. Also, when solving the system in (9), there are situations when the system is very ill-conditioned. For example, for a given candidate shift, if the region of f that falls within the ROI is nearly constant, then the set of regressors is nearly linearly dependent. Since these cases typically do not correspond to a correct registration, we do not solve the system (9) if the condition number of $\mathbf{A}_{x,y}$ is above a chosen threshold (10^{15}).

B. Experiments

The method was tested in three different image processing genres: medical imaging, remote sensing, and forensic science. Each genre consisted of two distinct pairs of images (giving a total of six image pairs over all three genres). For each pair of images, one image was selected as f, and the other as g. For each image g, two different ROIs were chosen, a large one and a small one. The large ROI covered approximately 40% of the object in the overlapping portion of the images, while the small ROI covered approximately 10%. In all, there were 12 different registration scenarios. Each scenario was registered with the SSD method described above, as well as with two cost functions implemented by FLIRT [33], part of the FSL family of programs developed at Oxford [34]. We used FLIRT's implementation of the correlation ratio (CR) and normalized mutual information (NMI) cost functions. For the NMI trials, the number of bins was set to 64 (all images were 8-bit). Also, a custom schedule file was created to mimic FLIRT's default 3-D registration strategy in a 2-D rigid-body registration context. The schedule file was catered to small rotations by sampling the cost function for angles 10° or smaller. All methods used linear interpolation for resampling.

The gold-standard rigid-body transform for each registration scenario was estimated by manually selecting five corresponding points in the image pair and computing the optimal (least-squares) transform to align them.

To assess the success of a registration method, we computed the average pixel displacement (from its true registered position) for pixels within the ROI. A displacement of zero means that the registration was perfect, while a very large average displacement indicates that the registration failed.

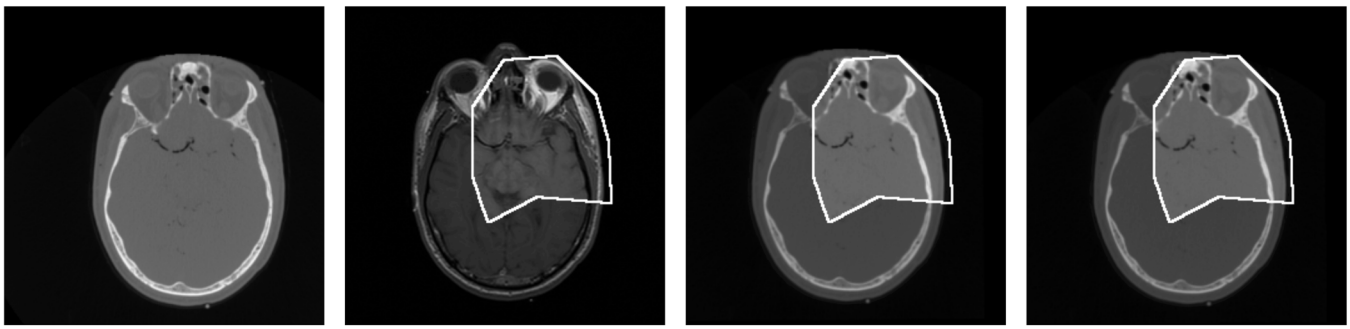
IV. Results

Table I shows the average pixel displacement for the three registration methods on the 12 registration scenarios. These results are reported in measurement units of pixels.

The SSD method took about 160 s to run all 12 cases, thus averaging about 13 seconds per registration. Running FLIRT's CR method on all 12 cases took 348 s, averaging about 29 s per

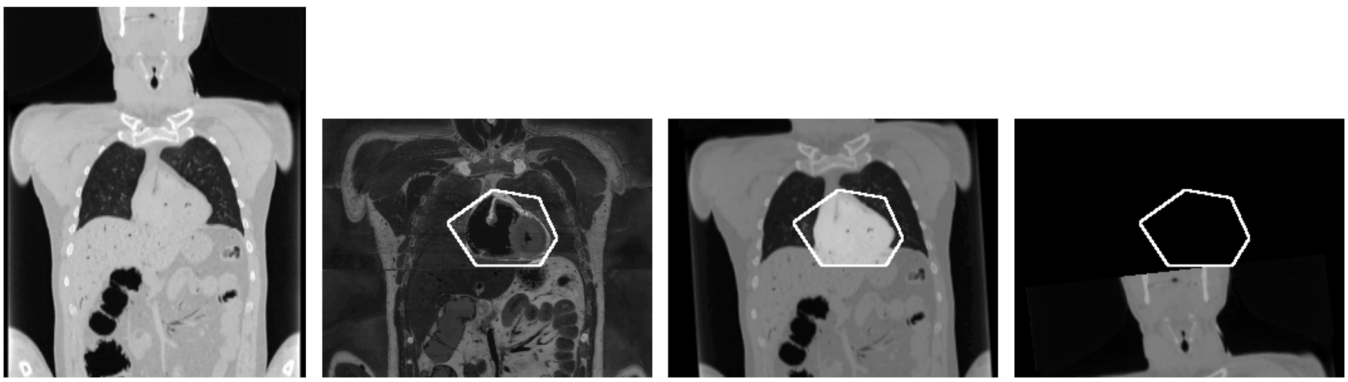
TABLE I
AVERAGE PIXEL DISPLACEMENT (PIXELS) FOR EACH OF THE 12 SCENARIOS

	SSD		FLIRT (Corr. Ratio)		FLIRT (Norm. MI)	
Image Pair	Small ROI	Large ROI	Small ROI	Large ROI	Small ROI	Large ROI
Head	4.1	3.3	209.9	18.0	148.1	6.4
Torso	9.8	2.4	184.4	244.4	171.5	4.8
Intermap	1.7	3.0	77.7	44.1	31.6	54.8
Landsat	3.8	2.6	222.1	412.5	337.2	400.2
Dental	17.1	11.8	151.0	228.5	303.3	302.8
Fingerprint	10.9	14.4	108.0	253.1	384.8	243.4



(a) (b) (c) (d)

Fig. 1. Results for registering the CT image to the T1-weighted MR image (large ROI) in the "Head" image pair. Both the CT and T1 images are 256 × 256. The gold-standard parameters to register the CT to the T1 are $(\theta, x, y) = (1.8°, -17.3, 2.8)$, where the initial placement aligned their upper-left corners. The images are from the Visible Human Project (National Library of Medicine), (a) CT, (b) TI-MRI with ROI, (c) Moved CT (SSD), and (d) Moved CT (FLIRT NMI).



(a) (b) (c) (d)

Fig. 2. Results for registering the CT image to the graytone photo (small ROI) in the "Torso" image pair. The CT image is 240 × 295 and the graytone photo is 262 × 206. The gold-standard parameters to register the CT to the photo are $(\theta, x, y) = (-0.3°, 14.4, 66.4)$, where the initial placement aligned their upper-left corners. The images are from the Visible Human Project (National Library of Medicine), (a) CT, (b) Photo with ROI, (c) Moved CT (SSD), and (d) Moved CT (FLIRT NMI).

case.[1] Finally, FLIRT's NMI method took at total of 404 seconds, averaging about 34 s per registration. These timing results exclude the time it takes to save the registered image since the option was disabled for these timing tests. In ten out of the 12 cases, the SSD method converged in just one iteration (i.e., the estimated rotation was less than 0.5°). The remaining two cases each converged after four iterations.

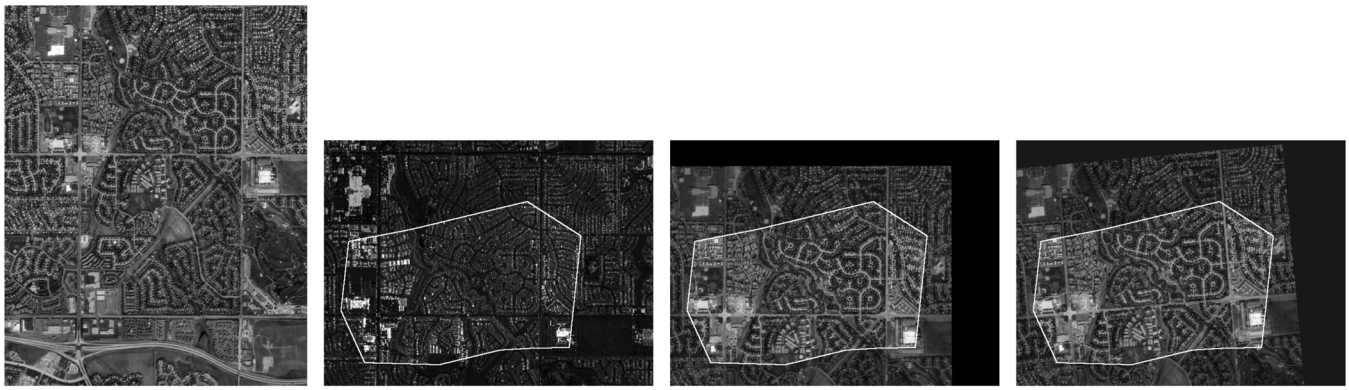
Figs. 1–6 show some example registration results using the three methods on six of the 12 registration scenarios (choosing one of the two ROIs for each of the six image pairs). The images depicting an ROI [labeled (b)–(d)] are dimmed outside the ROI to make the ROI more visible. Starting with a large ROI, the size alternates from one figure to the next. For each figure, the result for either FLIRT's correlation ratio or normalized mutual information is shown, whichever yielded the lower average pixel displacement.

Fig. 7 shows a joint-intensity scatter plot for the pixels inside the large ROI for the "Torso" image pair after being registered correctly by the SSD method. Superimposed over the scatter plot is the optimal piecewise-linear regressor.

V. DISCUSSION AND CONCLUSIONS

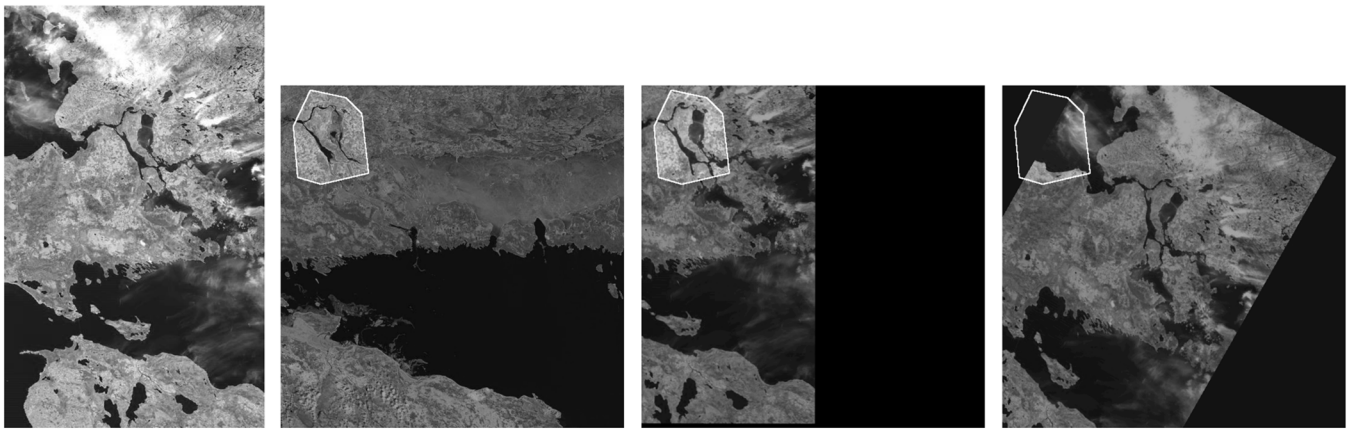
The results clearly suggest that the SSD method proposed here is comparable to, if not superior to, the FLIRT registration method when applied to registration scenarios where the initial misregistration is large. The method works for both mono- and multimodal registration.

[1]FLIRT is not necessarily optimized for our computing environment.



(a) (b) (c) (d)

Fig. 3. Results for registering the graytone photo to the height map (large ROI) in the "Intermap" image pair. The graytone photo is 602 × 756, and the height map image is 655 × 488. The gold-standard parameters to register the photo to the height map are $(\theta, x, y) = (0^\circ, -42.8, -53.2)$, where the initial placement aligned their upper-left corners. The images are courtesy of Intermap Technologies, Inc. (Englewood, CO), (a) Photo, (b) Hight map with ROI, (c) Moved Photo (SSD), and (d) Moved Photo (FLIRT CR).



(a) (b) (c) (d)

Fig. 4. Results for registering image 1 to image 2 (small ROI) in the "Landsat" image pair. Image 1 is 430 × 698, and image 2 is 567 × 566. The gold-standard parameters to register image 1 to image 2 are $(\theta, x, y) = (0^\circ, -141.0, 141.6)$, where the initial placement aligned their upper-left corners. The images are courtesy of the U.S. Geological Survey Landsat Project, (a) Image 1, (b) Image 2 with ROI, (c) Moved Image 1 (SSD), and (d) Moved Image 1 (FLIRT CR).

The proposed SSD method is very efficient, and even outperforms the FLIRT methods despite the fact that FLIRT is designed with speed in mind. The SSD method generally completed in less than half the time it took FLIRT. However, our implementation of the SSD method takes advantage of several optimizations, whereas it is not clear how optimized FLIRT is for our computing platform. Also, part of FLIRT's runtime is spent fine-tuning its registration solution, something that our SSD method does not do. However, the amount of time spent on this fine-tuning is likely not large, since FLIRT uses "a simple but fast local optimization method" [16].

Even though we outline here a method for the approximate registration of 2-D images, the method scales very naturally to data of higher dimension, such as 3-D, 4-D, etc.

The images used in these experiments are orientation-aligned. That is, we assume that the rotation is small (e.g., less than 5°). Our method computes the locally best rotation for all possible translations. In this sense, out method is global in terms of translation, but local in terms of rotation. For 2-D rigid-body registration, this leaves only one parameter that is not exhaustively searched. The method is fast enough that it can be computed starting from a number of different candidate rotations, choosing the best result as the final solution. This "multistart" idea is used in other registration methods, including FLIRT [16]. It should be noted that such a search becomes much more difficult in higher dimensions because the number of rotational parameters increases.

Currently, our SSD method does not incorporate scale changes. The scaling transform can be linearly approximated (as rotation is in this study), and incorporated using additional regressors. In our two-pronged approach, this would give us three regressors for the gradient magnitude SSD computation (rather than two), and 12 for the piecewise-linear portion (rather than nine).

Higher dimensional polynomial regressors can also be used in the linear regression, although the cross-terms that result from the linear rotation term start to pile up, and the method becomes computationally cumbersome. This is one reason we opted for a piecewise-linear regression model (the other reason being the ability to model discontinuities). Ultimately, the method outlined here affords a great deal of flexibility. It remains to be seen where the computational trade-offs overwhelm the benefits.

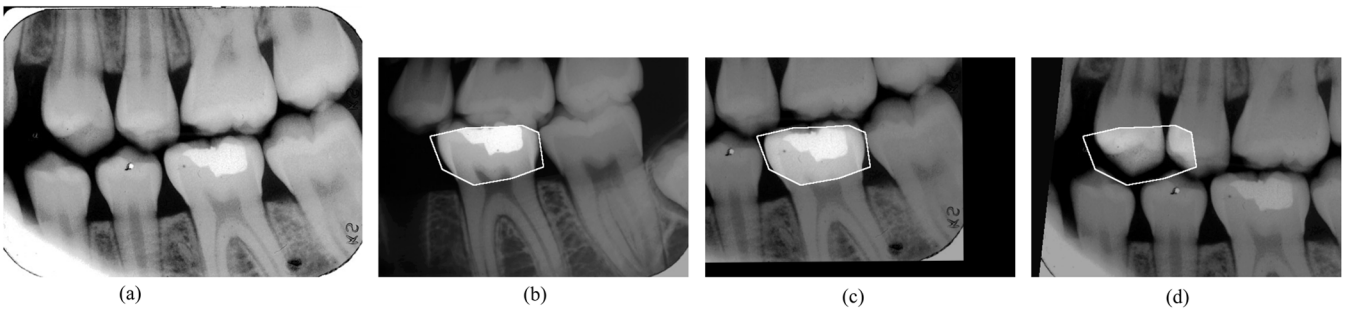


Fig. 5. Results for registering the antemortem image to the postmortem image (large ROI) in the "Dental" image pair. The antemortem image is 596 × 450, and the postmortem image is 515 × 365. The gold-standard parameters to register image 1 to image 2 are $(\theta, x, y) = (-3.6°, -166.5, 125.5)$, where the initial placement aligned their upper-left corners. The images are courtesy of Dr. David Sweet (Bureau of Legal Dentistry, University of British Columbia), (a) Antemortem, (b) Postmaster with ROI, (c) Moved Antemortem (SSD), and (d) Moved Antemortem (FLIRT CR).

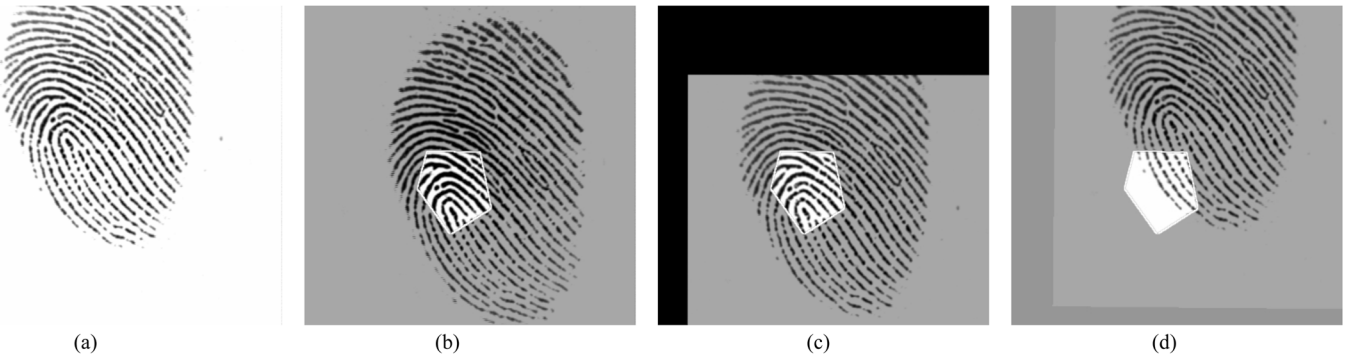


Fig. 6. Results for aligning print 1 to print 2 (small ROI) in the "Fingerprint" image pair. Both images are 388 × 374. The gold-standard parameters to register image 1 to image 2 are $(\theta, x, y) = (-2.1°, 32.6, -83.2)$, where the initial placement aligned their upper-left corners. The images are taken from the training dataset of the FVC2002 competition [35], (a) Print 1, (b) Print 2 with ROI, (c) Moved Print 1 (SSD), and (d) Moved Print 1 (FLIRT CR).

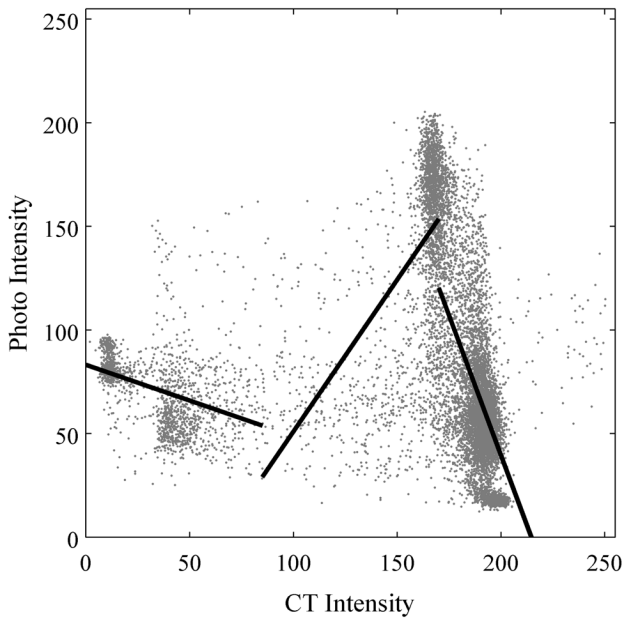


Fig. 7. Intensity scatter plot for the SSD registration solution for the "Torso" image pair using the large ROI. The three lines depict the optimal piecewise-linear function relating the intensities in the CT image to the intensities in the graytone photo.

For the most part, the piecewise-linear SSD worked fine by itself. However, it failed on a few of the 12 scenarios in this study. The same is true for the gradient magnitude SSD, although it failed more often than the piecewise-linear method. Each SSD approach by itself does not always find the minimum cost value at the true registration. However, the combination of the two SSD approaches has thus far proven to be very effective at combatting this phenomenon, probably because it is unlikely that both SSD approaches will assign a low cost to the same erroneous registration. In a sense, the approaches complement each other, and as long as they both consistently assign a relatively low cost to the true registration, the minimum of the combined SSD will likely occur at the true solution.

The piecewise-linear function used in the linear regression was generic, consisting of three equal-width linear pieces. Improvement to our SSD registration method might include using prior knowledge or image histograms to guide the design of the linear pieces to strategically place the discontinuities.

The full range of possibilities of the weighting function (alpha mask) w have not been explored in this paper. Here, we essentially use w as a binary mask. With no additional computational cost, more complicated weighting functions can be used, involving two or more regions of interest, as well as fractional weights (between 0 and 1).

By no means does this paper constitute an exhaustive study of the benefits and pitfalls of the SSD method or the FLIRT registration method. Indeed, there are plenty of registration scenarios for which the proposed SSD method will fail. In these cases, we expect most other methods will also fail. This study simply

stands as a demonstration of the potential of the SSD method. Future work includes examining what other degrees of freedom we can build into the spatial transformation, and still realize the efficiency and robustness.

ACKNOWLEDGMENT

The author would like to thank the National Library of Medicine, Intermap Technologies, Inc., the United States Geological Survey, the Fingerprint Verification Competition, and Dr. D. Sweet for the use of their images in this study.

REFERENCES

[1] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[2] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2-D-3-D medical image registration," *IEEE Trans. Med. Imag.*, vol. 17, no. 4, pp. 586–595, Aug. 1998.

[3] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, 1992.

[4] R. Piché, J. Macki, Ed., "Stitching IC images," in *Proc. PIMS Industrial Problem Solving Workshop*, Jun. 2002, pp. 1–25 [Online]. Available: http://www.pims.math.ca/publications/proceedings/ipsw6.pdf, Pacific Institute for Mathematical Sciences

[5] M. Omanovic, "Matching of dental X-rays for human forensic identification," M.S. thesis, Univ. Waterloo, Waterloo, ON, Canada, Sep. 2006.

[6] H. Chen and A. K. Jain, "Dental biometrics: Alignments and matching of dental radiographs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1319–1326, Aug. 2005.

[7] E. De Castro and C. Morandi, "Registration of translated and rotated images using finite fourier transforms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 5, pp. 700–703, May 1987.

[8] Y. Keller, A. Averbuch, and M. Israeli, "Pseudopolar-based estimation of large translations, rotations, and scalings in images," *IEEE Trans. Image Process.*, vol. 14, no. 1, pp. 12–22, Jan. 2005.

[9] L. C. Maas, B. D. Frederick, and P. F. Renshaw, "Decoupled automated rotational and traslational registration for fMRI time series data: The DART registration algorithm," *Magn. Res. Med.*, vol. 37, pp. 131–139, 1997.

[10] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.

[11] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proc. Int. Conf. Cybernetics and Society*, 1975, pp. 163–165.

[12] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 188–200, Mar. 2002.

[13] G. Wolberg and S. Zokai, "Robust image registration using log-polar transform," in *Proc. Int. Conf. Image Processing*, 2000, vol. 1, pp. 493–496.

[14] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, Y. Bizais, C. Barillot, and R. Di Paola, Eds., "Automated multi-modality image registration based on information theory," in *Proc. Inf. Proc. Med. Imag.*, 1995, pp. 263–274.

[15] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–51, 1996.

[16] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for robust and accurate linear registration and motion correction of brain images," *NeuroImage*, vol. 17, no. 2, pp. 825–841, 2002.

[17] K. J. Friston, J. Ashburner, C. D. Frith, J.-B. Poline, J. Heather, and R. S. J. Frackowiak, "Spatial registration and normalization of images," *Human Brain Map.*, vol. 33, pp. 165–189, 1995.

[18] A. Guimond, A. Roche, N. Ayache, and J. Meunier, "Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 58–69, Jan. 2001.

[19] S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations," *IEEE Trans. Med. Imag.*, vol. 22, no. 7, pp. 865–874, Jul. 2003.

[20] S. Grootoonk, C. Hutton, J. Ashburner, A. M. Howseman, O. Josephs, G. Rees, K. J. Friston, and R. Turner, "Characterization and correction of interpolation effects in the realignment of fMRI time series," *NeuroImage*, vol. 11, pp. 49–57, 2000.

[21] J. Orchard, M. Kamel and A. Campilho, Eds., "Efficient global weighted least-squares translation registration in the frequency domain," in *Proc. ICIAR Lecture Notes in Computer Science*, Sep. 2005, vol. 3656, pp. 116–124.

[22] S. Thesen, O. Heid, E. Mueller, and L. R. Schad, "Prospective acquisition correction for head motion with image-based tracking for real-time fMRI," *Magn. Res. Med.*, vol. 44, pp. 457–465, 2000.

[23] Y. Wang, R. Grimm, J. Felmlee, S. Riederer, and R. Ehman, "Algorithms for extracting motion information from navigator echoes," *Magn. Res. Med.*, vol. 36, pp. 117–123, 1996.

[24] 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. Comput. Assist. Tomogr.*, vol. 22, pp. 139–152, 1998.

[25] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Math. Comput.*, vol. 19, no. 90, pp. 297–301, 1965.

[26] A. E. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using fast fourier transform techniques," *IEEE Trans. Geosci. Electron.*, vol. GE-8, no. 4, pp. 353–368, Oct. 1970.

[27] A. J. Fitch, A. Kadyrov, W. J. Christmas, and J. Kittler, "Fast robust correlation," *IEEE Trans. Image Process.*, vol. 14, no. 8, pp. 1063–1073, Aug. 2005.

[28] J. Lewis, Fast normalized cross-correlation [Online]. Available: http://www.idiom.com/zilla/Work/nvisionInterface/nip.pdf [Online]. Available:

[29] J. G. Pipe, "Motion correction with propeller MRI: Application to head motion and free-breathing cardiac imaging," *Magn. Res. Med.*, vol. 42, pp. 963–969, 1999.

[30] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[31] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed. Philadelphia, PA: SIAM, 1999.

[32] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3," *Proc. IEEE*, vol. 93, no. 2, pp. 216–231, Feb. 2005.

[33] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, 2001.

[34] S. Smith, M. Jenkinson, M. Woolrich, C. Beckmann, T. Behrens, H. Johansen-Berg, P. Bannister, M. D. Luca, I. Drobnjak, D. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. Brady, and P. Matthews, "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23, no. S1, pp. 208–219, 2004.

[35] FVC2002, second international fingerprint verification competition. [Online]. Available: http://bias.csr.unibo.it/fvc2002 [Online]. Available:

Jeff Orchard (M'03) received the B.Math. degree in applied mathematics from the University of Waterloo, Waterloo, ON, Canada, in 1994, the M.Sc. degree in applied mathematics from the University of British Columbia, BC, Canada, in 1996, and the Ph.D. degree in computing science from Simon Fraser University, Canada, in 2003.

Since 2003, he has been an Assistant Professor in the David R. Cheriton School of Computer Science, University of Waterloo. His research interests revolve around applying mathematics and computation to visual data. He has worked on projects in image registration, motion compensation for medical imaging, functional MRI, medical image reconstruction, and image mosaicking. At the University of Waterloo, he is affiliated with the Scientific Computing Research Group, the Waterloo Institute for Health Informatics Research, and the Centre for Computational Mathematics in Industry and Commerce. In 2005, he organized a workshop called the "Grand Mathematical Challenges in Medical Image Processing."