Simultaneous Registration and Activation Detection for fMRI

Jeff Orchard*, *Student Member, IEEE*, Chen Greif, Gene H. Golub, Bruce Bjornson, and M. Stella Atkins

Abstract—**Registration using the least-squares cost function is sensitive to the intensity fluctuations caused by the blood oxygen level dependent (BOLD) signal in functional MRI (fMRI) experiments, resulting in stimulus-correlated motion errors. These errors are severe enough to cause false-positive clusters in the activation maps of datasets acquired from 3T scanners. This paper presents a new approach to resolving the coupling between registration and activation. Instead of treating the two problems as individual steps in a sequence, they are combined into a single least-squares problem and are solved simultaneously. Robustness tests on a variety of simulated three–dimensional EPI datasets show that the stimulus-correlated motion errors are removed, resulting in a substantial decrease in false-positive and false-negative activation rates. The new method is also shown to decorrelate the motion estimates from the stimulus by testing it on different *in vivo* fMRI datasets acquired from two different 3T scanners.**

Index Terms—**Activation, fMRI, GLM, least-squares, registration.**

I. INTRODUCTION

IT has been shown that functional MRI (fMRI) experiments are highly sensitive to patient movement. Failure to correct gross subject motion can corrupt the resulting activation map, and lead to both false-positive and false-negative activations [1]. Rigid-body motion detection and correction, also known as registration, can greatly reduce the effect of motion and yield activation maps with improved accuracy [2]–[4]. But, the success of motion correction depends on the accuracy of the motion detection algorithm. If the motion estimates are flawed, the resulting resampled dataset will reflect these motion errors.

Among the most common techniques for motion detection of serially acquired MR images is the intensity-based least-squares algorithm [5]. The least-squares method is both fast and accurate for aligning volumes that differ only by a rigid-body transformation and Gaussian noise [6].

*J. Orchard was with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. He is now with the School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: jorchard@cs.uwaterloo.ca).

C. Greif is with the Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

G. H. Golub is with the Scientific Computing and Computational Mathematics Program, Stanford University, Stanford, CA 94305-9025 USA.

B. Bjornson is with the Department of Pediatrics (Neurology), University of British Columbia, BC's Children's Hospital, Vancouver, BC V6H 3V4, Canada

M. S. Atkins is with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

However, the least-squares cost function is known to be sensitive to outliers. Since the error measure is the sum of the squared residuals, outliers contribute a disproportionately large amount to the overall error. In studies conducted by Freire *et al.* [7], [8], motion estimates on fMRI datasets from a 3T scanner exhibited errors that were correlated with the amount of blood oxygen level dependent (BOLD) signal. They hypothesized that the voxels containing the BOLD signal act as outliers, and disrupt the least-squares registration.

It has been shown that the registration errors induced by activation are proportional to the magnitude of the BOLD signal [9]. Hence, with the increased prevalence of high-field MR scanners, we can expect this effect to take on greater significance. Higher field strengths result in increased BOLD signals, which in turn manifest themselves as larger registration errors.

Some registration algorithms, such as those that employ mutual information [6] or robust estimators [10], [11], are less sensitive to outliers than least-squares [7], [12]. Robust estimators remove the over-emphasis of outliers by utilizing an asymptotic error function that bounds the error contribution from each voxel. These methods, however, do not explicitly incorporate the time-varying activation present in fMRI data. As a result, it is not clear how these registration methods perform under all situations involving activation. It is likely that increased BOLD signal from scanners with higher field strengths will cause erroneous motion estimates. A method should be devised that fully incorporates, rather than down-plays, the activation signal.

One fMRI registration method that incorporates the stimulus function has been developed [13]. The method performs motion compensation in two registration steps. An initial registration step removes most of the patient motion, but is likely to leave some stimulus-correlated motion errors. Then, activation detection is used to construct a binary mask of voxels that exhibit a BOLD signal. This mask of activated voxels is dilated before a second pass of the registration algorithm is performed, involving only unmasked voxels. The author points out that as the size of the region of activation increases, the second registration pass must be carried out with fewer samples (voxels), thereby reducing its statistical significance. Furthermore, regions of false-positive activation in the activation map can cause inactive regions to be excluded from the second registration step.

The fundamental contribution of this paper is the formulation of two processing steps, registration and activation detection, into a single model that incorporates both motion and activation. This method is called the *simultaneous registration and activation* (SRA) method, and solves the least-squares registration and least-squares activation detection problems simultane-

ously. In Section II, we review the theoretical formulation of the individual problems, and outline a strategy for solving the combined problem. In Sections III and IV, we demonstrate the SRA method's feasibility and effectiveness on case studies involving simulated fMRI datasets and two *in vivo* fMRI datasets from different scanners. Finally, in Section V, we discuss the advantages and disadvantages of our approach, and suggest directions for further investigation.

II. THEORY

A. Registration

Determining the patient's head motion from a time series of volumes is often posed as a least-squares optimization problem. In particular, given two volume datasets, $\mathbf{f}$ and $\mathbf{g}$, the objective is to find the motion parameters $\mathbf{x}$ that define a p-degree-of-freedom rigid-body transformation, $T^{-1}(\mathbf{f}, \mathbf{x})$, such that $T^{-1}(\mathbf{f}, \mathbf{x}) = \mathbf{g}$. For three-dimensional (3-D) datasets, p is usually six. To solve this in the least-squares sense, we seek an $\mathbf{x}$ that minimizes the sum of squared residuals between $\mathbf{g}$ and $T^{-1}(\mathbf{f}, \mathbf{x})$

$$\min_{\mathbf{x}} \sum_i \left(\mathbf{g}_i - T^{-1}(\mathbf{f}, \mathbf{x})_i\right)^2 \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{g} - T^{-1}(\mathbf{f}, \mathbf{x})\|^2. \quad (1)$$

In (1), the delimiter $\| \cdot \|^2$ refers to the square of the Frobenius norm, while the subscripted ith specifies a single voxel in a volume. The variables $\mathbf{f}$ and $\mathbf{g}$ refer to column vectors of length m containing the intensity of all m voxels of their respective volumes (as a convention, volumes will be stored in columns because rows will later be used to represent voxel time series). In this context, we call $\mathbf{g}$ the "reference volume", and $\mathbf{f}$ the "floating volume".

Alternatively, we can seek the transformation that minimizes the sum of squared residuals between $T(\mathbf{g}, \mathbf{x})$ and $\mathbf{f}$. A rigid-body transformation is nonlinear and cannot be represented using a finite number of linear transformations. We can, however, linearize the operator by ignoring the higher-order terms in the Taylor expansion

$$\begin{aligned} T(\mathbf{g}, \mathbf{x}) = & T(\mathbf{g}, 0) + \nabla T(\mathbf{g}, 0)\mathbf{x} + \mathcal{O}\left(\|\mathbf{x}\|^2\right) \\ = & \mathbf{g} + \mathbf{A}\mathbf{x} + \mathcal{O}\left(\|\mathbf{x}\|^2\right). \end{aligned} \quad (2)$$

It should be noted that this linearization is valid for small motions, and becomes less accurate as the motion gets larger. The matrix $\nabla T(\mathbf{g}, 0)$ (and, thus, $\mathbf{A}$) is an $m \times p$ matrix in which the ith row holds the gradient vector of the intensity of the ith voxel in $\mathbf{g}$ with respect to the p motion parameters. Equation (3) shows one row of $\mathbf{A}$ containing the six partial derivatives of $\mathbf{g}_i$ with respect to the three translations and three rotations of 3-D rigid-body motion

$$\mathbf{A} = \begin{bmatrix} \vdots \\ \frac{\partial \mathbf{g}_i}{\partial x} & \frac{\partial \mathbf{g}_i}{\partial y} & \frac{\partial \mathbf{g}_i}{\partial z} & \frac{\partial \mathbf{g}_i}{\partial \theta_x} & \frac{\partial \mathbf{g}_i}{\partial \theta_y} & \frac{\partial \mathbf{g}_i}{\partial \theta_z} \\ \vdots \end{bmatrix}. \quad (3)$$

The derivatives in $\mathbf{A}$ can be estimated using a finite difference approximation by applying small motions to $\mathbf{g}$.

The vector $\mathbf{x}$ in (2) holds the linear motion parameters. By substituting the approximation for T into $\mathbf{f} = T(\mathbf{g}, \mathbf{x})$, we get

$$\mathbf{f} \approx \mathbf{g} + \mathbf{A}\mathbf{x}. \quad (4)$$

The least-squares solution to (4) can be written analytically [14], [15] as

$$\mathbf{x} = \mathbf{A}^\dagger (\mathbf{f} - \mathbf{g}), \quad (5)$$

where the matrix $\mathbf{A}^\dagger$ is the pseudoinverse of $\mathbf{A}$, and equals $(\mathbf{A}'\mathbf{A})^{-1} \mathbf{A}'$ when $\mathbf{A}$ is of full rank.

After finding $\mathbf{x}$, the floating volume can be resampled using the inverse of T, and another motion estimate can be obtained from the new dataset. By applying this approximation iteratively, we generate a sequence of linear motion estimates, $\mathbf{x}^{(k)}$. These increments are accumulated into a current best guess set of motion parameters by simple addition. If this best guess is denoted $\Sigma\mathbf{x}^{(k)}$, then we can represent the kth resampled volume as a transformed version of the original floating volume

$$\mathbf{f}^{(k)} = T^{-1}\left(\mathbf{f}^{(0)}, \Sigma\mathbf{x}^{(k)}\right) = T^{-1}\left(\mathbf{f}, \Sigma\mathbf{x}^{(k)}\right). \quad (6)$$

Under certain mild conditions, the motion increments will converge as each iteration brings $T^{-1}\left(\mathbf{f}, \Sigma\mathbf{x}^{(k)}\right)$ closer to alignment with $\mathbf{g}$. This technique is called *fixed point iteration*.

This type of registration is often performed on a time series of n fMRI volumes, aligning each with the reference volume. The above notation is general enough to represent the registration of all volumes at once, achieved by defining the matrix $\mathbf{F}$ with n columns, each holding one of the volumes in the dataset. Since $\mathbf{F}$ now has n columns, the motion parameter matrix must have n columns, one set of motion parameters for each volume in the time series. Hence, we define $\mathbf{X}$ as just such a matrix. We must also define $\mathbf{G}$ as a matrix that holds a copy of the reference volume in each of its n columns. Thus, the full time-series registration problem is depicted by

$$\mathbf{F} = \mathbf{G} + \mathbf{A}\mathbf{X}. \quad (7)$$

If $T^{-1}\left(\mathbf{F}, \Sigma\mathbf{X}^{(k)}\right)$ represents the resampled dataset in which the n cumulative rigid-body transformations held in $\Sigma\mathbf{X}^{(k)}$ are applied to the columns of $\mathbf{F}$, then our nth-volume registration problem can be solved by iterating the system

$$\begin{cases} \mathbf{F}^{(k)} = T^{-1}\left(\mathbf{F}^{(0)}, \Sigma\mathbf{X}^{(k-1)}\right) \\ \mathbf{X}^{(k)} = \mathbf{A}^\dagger\left(\mathbf{F}^{(k)} - \mathbf{G}\right) \end{cases} \quad (8)$$

until the motion increments are sufficiently small.

B. Activation

Determining whether or not a voxel time series exhibits synchrony with the stimulus can also be posed as a least-squares problem using the general linear model (GLM). The idea behind the GLM is that a voxel time series is the sum of effects from different sources. For example, not only does each voxel have a constant, baseline intensity, but it may also have a component that increases over time, or a component that fluctuates with the stimulus. These components are called "regressors". The GLM

is based on the idea that some linear combination of the regressors produces a time series that is close to the voxel's actual time series. Activation detection consists of decomposing a voxel's time series by finding a scalar multiplier (fit coefficient) for each of these regressors.

Suppose we have s regressors, not including the voxel's baseline intensity (the baseline intensity can easily be accounted for by including a constant regressor, but at this point we choose to explicitly separate the baseline intensity). Conforming to the notation in Section II-A by keeping time series in rows, the GLM is

$$\mathbf{v} = \mathbf{u} + \mathbf{y}\mathbf{B} \tag{9}$$

where $\mathbf{v}$ is a row-vector containing the voxel's time series, $\mathbf{u}$ is a row-vector with n copies of the voxel's baseline value, and $\mathbf{y}$ is a row-vector of s parameters that correspond to the s regressors held in the rows of the matrix $\mathbf{B}$. It is more common to see this relationship in the form of its transpose, but we choose to keep time series in rows to maintain consistent notation with the registration least-squares formulation.

Using matrix pseudoinversion again, we can solve this least-squares problem analytically

$$\mathbf{y} = (\mathbf{v} - \mathbf{u})\mathbf{B}'\left(\mathbf{B}\mathbf{B}'\right)^{-1} = (\mathbf{v} - \mathbf{u})\mathbf{B}^{\dagger}. \tag{10}$$

Like in the registration case, we can combine the GLM of all the voxel time series into a single matrix equation

$$\mathbf{F} = \mathbf{G} + \mathbf{Y}\mathbf{B} \tag{11}$$

where $\mathbf{F}$ and $\mathbf{G}$ are the same as in Section II-A, $\mathbf{B}$ holds the regressors, and $\mathbf{Y}$ is an $m \times s$ matrix containing the regression fit coefficients for each voxel/regressor combination. Thus, the solution to the least-squares activation detection problem is

$$\mathbf{Y} = (\mathbf{F} - \mathbf{G})\mathbf{B}^{\dagger}. \tag{12}$$

Traditionally, fMRI processing proceeds by first solving the registration problem by itself, followed by solving the activation problem on the motion-"corrected" dataset [16], [17]. However, this strategy allows the BOLD signal to interfere with the registration algorithm. The resulting resampled dataset may depict less motion than the original, but may include stimulus-correlated motion induced by systematic registration errors. For the remainder of this paper, we will refer to this traditional, sequential solution as the *standard* method.

C. Simultaneous Registration and Activation

We propose a new approach, called the SRA algorithm, that avoids the interference of the BOLD signal in the least-squares registration algorithm. Observe that (7) and (11) are similar in nature. By combining them, a model to account for both motion and activation effects can be stated in a single matrix equation

$$
\begin{aligned}
\{\texttt{observed dataset}\} =& \{\texttt{reference dataset}\} \\
& + \{\texttt{motion}\} \\
& + \{\texttt{activation}\} \\
\mathbf{F} =& \mathbf{G} + \mathbf{A}\mathbf{X} + \mathbf{Y}\mathbf{B}.
\end{aligned} \tag{13}
$$

Recall that $\mathbf{F}$ holds a full volume in each column. For a dataset of volumes with dimensions $64 \times 64 \times 30$, the matrices $\mathbf{F}$, $\mathbf{G}$, $\mathbf{A}$ and $\mathbf{Y}$ have 122 880 rows each. Hence, columns represent the spatial component of the dataset. On the other hand, rows represent the temporal component of the dataset. The matrices $\mathbf{F}$, $\mathbf{G}$, $\mathbf{X}$ and $\mathbf{B}$ all hold time series in their rows. For a dataset with 80 volumes, they each have 80 columns. If we are using six-degree-of-freedom motion, and two stimulus regressors, then our matrix dimensions are

$$
\begin{aligned}
\mathbf{F} =& \mathbf{G} + \mathbf{A}\mathbf{X} + \mathbf{Y}\mathbf{B} \\
[122\,880 \times 80] =& [122\,880 \times 80] \\
& + [122\,880 \times 6][6 \times 80] \\
& + [122\,880 \times 2][2 \times 80].
\end{aligned}
$$

Since $\mathbf{G} + \mathbf{A}\mathbf{X}$ is only an approximation to the rigid-body transformation, (13) is also an approximation, and must be solved iteratively just like in the original registration problem. However, the advantage here is that the effects of both motion and activation can be accounted for. Hence, we can solve the least-squares problem for $\mathbf{X}$ and $\mathbf{Y}$ simultaneously. The solution gives the rigid-body motion increment ($\mathbf{X}$) and activation ($\mathbf{Y}$) that minimizes the sum of squared residuals between $\mathbf{F}$ and $\mathbf{G}$, accounting for both motion and BOLD contrast in parallel.

Our new problem formulation involves an iterative process, requiring the solution of the least-squares problem

$$\min_{(\mathbf{X}, \mathbf{Y})} \|\mathbf{A}\mathbf{X} + \mathbf{Y}\mathbf{B} - \mathbf{C}\| \tag{14}$$

for each iteration. Here, $\|\cdot\|$ again refers to the Frobenius norm, and we have replaced $(\mathbf{F} - \mathbf{G})$ with $\mathbf{C}$.

D. General Solution

The problem stated in (14) does not have a unique solution. To illustrate this fact, assume that $(\mathbf{X}, \mathbf{Y})$ is a solution, and consider the perturbation $(\mathbf{X} + \delta_{\mathbf{X}}, \mathbf{Y} + \delta_{\mathbf{Y}})$. Substituting this into (14) gives us

$$
\begin{aligned}
\min_{(\mathbf{X}, \mathbf{Y})} & \|\mathbf{A}(\mathbf{X} + \delta_{\mathbf{X}}) + (\mathbf{Y} + \delta_{\mathbf{Y}})\mathbf{B} - \mathbf{C}\| \\
& = \min_{(\mathbf{X}, \mathbf{Y})} \|\mathbf{A}\mathbf{X} + \mathbf{Y}\mathbf{B} - \mathbf{C} + (\mathbf{A}\delta_{\mathbf{X}} + \delta_{\mathbf{Y}}\mathbf{B})\|.
\end{aligned}
$$

Thus, if $(\mathbf{A}\delta_{\mathbf{X}} + \delta_{\mathbf{Y}}\mathbf{B})$ is zero, then $(\mathbf{X} + \delta_{\mathbf{X}}, \mathbf{Y} + \delta_{\mathbf{Y}})$ is also a solution. It can be shown that $\delta_{\mathbf{X}}$ and $\delta_{\mathbf{Y}}$ must have the form $\delta_{\mathbf{X}} = \alpha\mathbf{B}$ and $\delta_{\mathbf{Y}} = -\mathbf{A}\alpha$, where α is an arbitrary matrix with dimensions $p \times s$ (p is the number of degrees of freedom in the motion, and s is the number of stimulus regressors). Thus, the problem in (14) has an infinite number of solutions, and the solution space is the ps-dimensional space

$$\{(\mathbf{X} + \alpha\mathbf{B}, \mathbf{Y} - \mathbf{A}\alpha)|\forall\alpha\} \tag{15}$$

where $(\mathbf{X}, \mathbf{Y})$ is any particular solution.

When an optimization problem does not have a unique solution, or is ill-posed, a common strategy used to steer the problem toward a sensible, unique solution is regularization. This can be

accomplished by reformulating (14) to include an additional parameter-dependent term, as follows:

$$\min_{(\mathbf{X},\mathbf{Y})} [\|\mathbf{AX} + \mathbf{YB} - \mathbf{C}\| + \gamma\Phi(\mathbf{X},\mathbf{Y})]. \qquad (16)$$

The regularization term, $\gamma\Phi(\mathbf{X},\mathbf{Y})$, is chosen to alter the cost function in a way that steers the optimization toward a solution which achieves a compromise between minimizing the original cost function and minimizing the function Φ. The relative weighting of the original cost function and the regularization term is controlled by the regularization parameter, γ. One of the challenges for this type of formulation is choosing an appropriate value for γ (see [18] for details). Once a value is chosen, there are numerical methods (depending on the form of Φ) that solve such optimization problems [19]–[21].

In our case, however, we already know that (15) is a parameterization of the solution space of (14). Solving the constrained optimization problem can then be accomplished with a two-step process. First, we find a particular solution, $(\mathbf{X},\mathbf{Y})$, to use as the anchor of our parameterized general solution. Then, we use an additional constraint to find the α that gives us our desired, unique solution. Once we have found α, we can calculate our constrained solution using (15).

E. Particular Solution

In order to obtain the general solution stated in (15), we must first find a particular solution to (14). To do this, we can use QR decomposition (see [15, section 5.2]) to decouple $\mathbf{X}$ and $\mathbf{Y}$. We start by performing a full QR decomposition on $\mathbf{B}'$, which gives us

$$\mathbf{B}' = [\mathbf{Q}_1\ \mathbf{Q}_2] \begin{bmatrix} \mathbf{R} \\ 0 \end{bmatrix} \qquad (17)$$

where $[\mathbf{Q}_1\ \mathbf{Q}_2]$ and $[\mathbf{R}\ 0]'$ are block matrices such that $\mathbf{Q}_1$ is $n \times s$, $\mathbf{Q}_2$ is $n \times (n-s)$, $\mathbf{R}$ is $s \times s$ and is upper-triangular, and the remaining $(n-s)$ rows of $[\mathbf{R}\ 0]'$ contain all zeros. Our problem then becomes

$$\min_{(\mathbf{X},\mathbf{Y})} \left\| \mathbf{AX} + \mathbf{Y}[\mathbf{R}'\ 0] \begin{bmatrix} \mathbf{Q}_1' \\ \mathbf{Q}_2' \end{bmatrix} - \mathbf{C} \right\|. \qquad (18)$$

Since the matrix $[\mathbf{Q}_1\ \mathbf{Q}_2]$ is orthogonal, we can multiply by it (or its transpose) without changing the norm [15]. Multiplying on the right by the orthogonal matrix $[\mathbf{Q}_1\ \mathbf{Q}_2]$ gives us the equivalent problem

$$\min_{(\mathbf{X},\mathbf{Y})} \|\mathbf{AX}[\mathbf{Q}_1\ \mathbf{Q}_2] + \mathbf{Y}[\mathbf{R}'\ 0] - \mathbf{C}[\mathbf{Q}_1\ \mathbf{Q}_2]\|. \qquad (19)$$

Notice that in (19), the elements of $\mathbf{Y}$ are only present in the first s columns of the product $\mathbf{Y}[\mathbf{R}'\ 0]$.

Multiplying a row-vector on the right by $[\mathbf{Q}_1\ \mathbf{Q}_2]$ is essentially a coordinate transformation to a reference-frame for which the columns of $[\mathbf{Q}_1\ \mathbf{Q}_2]$ form an orthogonal basis. The matrix $\mathbf{Q}_1$ has s columns spanning the row-space of $\mathbf{B}$, while $\mathbf{Q}_2$ has $(n-s)$ columns spanning the orthogonal space. Thus, the matrix $\mathbf{XQ}_1$ corresponds to the components of $\mathbf{X}$ that are in the row-space of $\mathbf{B}$. Since our goal is to find *any* particular solution to (14), these components can be set to zero without loss

of generality because our general solution already includes the row-space of $\mathbf{B}$. Hence, ignoring the term $\mathbf{AX}[\mathbf{Q}_1\ 0]$ in (19) simply results in an $\mathbf{X}$ solution that is orthogonal to the rows of $\mathbf{B}$. In doing so, we also completely decouple $\mathbf{X}$ and $\mathbf{Y}$, breaking the single least-squares problem into two separate problems

$$\min_{\mathbf{X}} \|\mathbf{AX}[0\ \mathbf{Q}_2] - \mathbf{C}[0\ \mathbf{Q}_2]\| \qquad (20)$$

$$+ \min_{\mathbf{Y}} \|\mathbf{Y}[\mathbf{R}'\ 0] - \mathbf{C}[\mathbf{Q}_1\ 0]\|. \qquad (21)$$

Assuming $\mathbf{R}$ is nonsingular (which must be the case if all the regressors in $\mathbf{B}$ are linearly independent) and that $\mathbf{A}$ has full column rank, the solution is

$$\mathbf{X} = \mathbf{A}^\dagger \mathbf{C}\mathbf{Q}_2\mathbf{Q}_2' \qquad (22)$$

$$\mathbf{Y} = \mathbf{C}\mathbf{Q}_1(\mathbf{R}')^{-1}. \qquad (23)$$

It should be noted that an equivalent approach is to take the QR decomposition of the matrix $\mathbf{A}$. However, $\mathbf{A}$ is much larger than $\mathbf{B}$. For a $64 \times 64 \times 30$ voxel volume, $\mathbf{A}$ has 122 880 rows, one row for each voxel. The resulting $\mathbf{Q}$ matrix would contain over 15 million elements, and require over 57 GB of RAM to store. The matrix $\mathbf{B}'$, on the other hand, has one row for every time-step, which is typically less than 200.

III. METHODS

In the above theoretical derivation, (22) and (23) yield a particular solution of (14). That particular solution is used in (15) to get the general solution. The task remains to determine which solution in the multidimensional solution space is appropriate for our needs. We need a mathematical expression that distinguishes one solution from all the others in the solution space. That is, what additional constraint can we impose on our general solution $(\mathbf{X} + \alpha\mathbf{B}, \mathbf{Y} - \mathbf{A}\alpha)$ to obtain a solution that is meaningful in the context of fMRI? For this, we focus on the activation maps.

The very nature of an activation map is that it contains regions of high intensity where the brain shows correlated activity (the active regions). The rest of the activation map is low-intensity "background" (the inactive regions). This property, in which many of the elements are close to zero, is called *sparsity*. We can expect the true activation map to exhibit some sparsity.

Because we know that the general solution contains activation maps of the form $(\mathbf{Y} - \mathbf{A}\alpha)$, any particular solution we obtain is composed of the true, desired activation mixed with the columns of $\mathbf{A}$. Moreover, other equivalent solutions can be derived from the particular solution by adding linear combinations of the columns of $\mathbf{A}$.

Recall that each column of $\mathbf{A}$ holds the partial derivative of $\mathbf{g}$ with respect to one of the motion parameters, and is generally nonzero near brain edges. Thus, adding linear combinations of the columns of $\mathbf{A}$ to the true activation map will cause edge artifacts to appear, thus reducing its sparsity. To isolate the true activation, we seek an α that removes these $\mathbf{A}$-components from our particular solution, causing $(\mathbf{Y} - \mathbf{A}\alpha)$ to be more sparse (i.e., to have more voxels close to zero). It is critical to understand that the sparsity of the true activation map itself is not as important as the degree to which these unwanted $\mathbf{A}$-components *decrease*

the sparsity. Hence, removing them yields an activation map that is *more* sparse. One constraint found to enforce sparsity is

$$\min_\alpha \sum_i \arctan\left(c|\mathbf{Y}_i - [\mathbf{A}\alpha]_i|\right) \qquad (24)$$

where the sum is over all voxels, and c is a chosen constant. This constraint is similar in nature to the Geman-McClure robust estimator [11]. Another constraint that has been found to favor sparse images [22] is

$$\min_\alpha \|\mathbf{Y} - \mathbf{A}\alpha\|_1 \qquad (25)$$

where $\|\cdot\|_1$ represents the ℓ_1 norm (sum of the absolute values of the elements). Many other constraints might also work, but the remainder of this paper is a case study on the effectiveness of the $\arctan$ constraint in (24).

There is a danger in choosing a particular frame as the reference volume, $\mathbf{g}$. If, for example, the chosen volume happens to be somewhat darker than the others, then the model will not be able to accurately represent the dataset, forcing all the frames except for the first to be treated as outliers. This issue becomes more important as the noise in the dataset increases.

We can instead allow the model to calculate its own optimal baseline volume by simply appending a row of ones to the matrix $\mathbf{B}$ and a corresponding column to $\mathbf{Y}$. That column of $\mathbf{Y}$ will hold the reference volume and is calculated in (23) just like the other columns of $\mathbf{Y}$. In this context, (13) could be written using block-matrices

$$\mathbf{F} = \mathbf{A}\mathbf{X} + \begin{bmatrix} \mathbf{Y}\mathbf{g} \end{bmatrix} \begin{bmatrix} \mathbf{B} \\ 1 \end{bmatrix} \qquad (26)$$

where $\mathbf{Y}$ and $\mathbf{B}$ are as before, the 1 refers to a row of ones, and $\mathbf{g}$ is an undetermined baseline volume. By making the appropriate substitutions for $\mathbf{B}$, $\mathbf{Y}$ and $\mathbf{C}$, the same solution process can be used as with the fixed reference volume case.

However, adding a row of ones to $\mathbf{B}$ increases the dimension of our solution space (see (15)), and we need six additional constraints to get a unique solution. That is, the row of ones in $\mathbf{B}$ allows every row of $\mathbf{X}$ to contain a different constant offset. We cannot use (24) or (25) to constrain this new freedom because the corresponding column of $\mathbf{Y}$ is not an activation map and does not need to be sparse. Instead, notice that we can arbitrarily choose one frame as stationary since all the motion parameters are relative to each other. Simply subtracting the stationary frame's motion parameters from all frames shifts all the volumes in the same way. Hence, we set the last column in α to the negative of the stationary frame's motion parameters. This approach is valid only because we assumed the motion increments are small, allowing us to treat the rigid-body transformation as a linear operator.

In our implementation, we let the model calculate the optimal baseline volume at each iteration, and used the Nelder/Mead simplex method [23] (also known as the Downhill simplex method) to find the other columns of α. Once the desired α was established, the constrained motion parameters and activation maps were found using (15).

A. Simulated Data

Testing of the algorithm included analyses on simulated fMRI time-series datasets. The datasets were created using an initial $64 \times 64 \times 30$ echo-planar imaging (EPI) volume with in-plane resolution of 3.75 mm and slice thickness of 4 mm. An activation mask was manually drawn, covering a large portion of the occipital region, as well as some small regions in the parietal lobes. A stimulus function was created using a slightly smoothed "box-car" function with a maximum value of 0.05, nonzero for frames 5–15, 25–35, 45–55, and 65–75. Both motion-free and motion-corrupted datasets were generated by performing the following steps.

1) Duplicate the original volume 80 times.
2) Add the appropriate amount of activation to the volumes. Activation was added to voxels under the activation mask by multiplying their intensities by $(1 + b)$, where b is the relative signal change due to BOLD contrast, specified by the stimulus function.
3) Apply rigid-body motion (motion-corrupted datasets only). The transformations were applied using AFNI's `3drotate` program [24], selecting the option for Fourier interpolation. The rotations were implemented as a composition of four shears [25], [26].
4) Add Gaussian-distributed, additive noise with a standard deviation of 2.5% of the average brain voxel intensity.
5) Apply spatial Gaussian smoothing to each volume (full-width at half-maximun = 5 mm).

Our SRA method was implemented in C++, using routines from *Numerical Recipes in C* [27] for QR decomposition and the Nelder/Mead simplex method. The algorithm involves an iterative process, just like in the standard least-squares motion detection algorithm. During each iteration, the general solution to (14) is found and constrained using (24). From the resulting motion parameters, the cumulative transformations were incremented, and the dataset was resampled and used in the next iteration.

The SRA method and the standard approach were each tested on simulated datasets that were produced with varying motion and activation characteristics. Ten datasets were generated for each of the following four scenarios:

Scenario 1: The datasets contain activation and random motion.
Scenario 2: The datasets contain activation and true stimulus-correlated random motion. Motion profiles were generated using a random mixture of the stimulus function and random motion.
Scenario 3: The datasets contain activation, but no motion.
Scenario 4: The datasets contain no activation, but contain stimulus-correlated random motion.

For each of the 40 trials, both motion detection algorithms (standard least-squares, and SRA) were used to estimate the motion. These motion estimates were used to resample the datasets before an activation map was generated. Calculation of the activation maps was done using the GLM, and solved in the least-squares sense (see (12) in Section II-B) yielding a linear fit coefficient for each voxel. We also calculated the correlation coefficient of each voxel's time-course with the stimulus function.
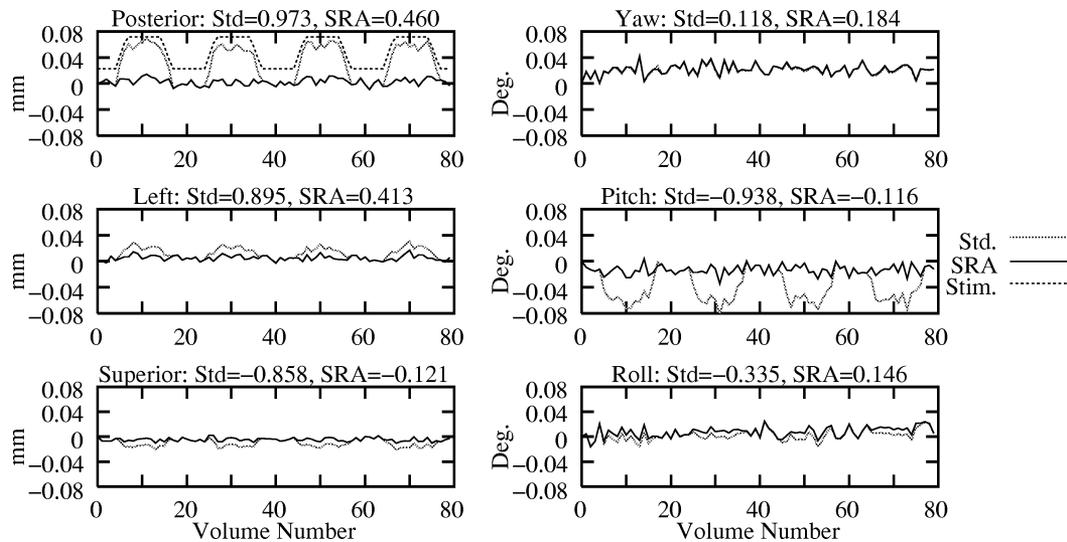
Fig. 1.   Errors in the motion parameters for a motion-corrupted dataset (scenario 1). The standard method ("Std.") exhibits errors that are correlated to the stimulus function, while the SRA method does not. The stimulus function is shown in the posterior translation plot, and the corresponding correlation coefficients are listed above each graph.

Finally, a binary activation mask was produced by thresholding, taking only those voxels with a correlation coefficient bigger (in magnitude) than 0.363, and a fit coefficient bigger (in magnitude) than 40 (approximately 15% of the maximum fit coefficient). A true activation mask was also produced in the same way, but from a dataset that was neither motion corrupted or motion compensated. The true mask contained a total of 3267 active voxels.

The activation masks were compared to the true activation mask yielding false-positive and false-negative counts for each trial. These counts were then averaged over the ten trials in each of the four scenarios.

B. In Vivo Data

The SRA method was also tested on two *in vivo* fMRI datasets from different patients and different scanners. Dataset I is an EPI scan of 180 volumes (192 in total, but the first 12 were discarded) acquired at two-second intervals on a 3T Bruker scanner. Each volume consists of 30 contiguous slices (4 mm thick), each having 64×64 voxels (3.75×3.75 mm). The subject was presented with two alternating visual stimuli, each lasting for nine frames. Dataset II is an EPI scan containing 60 volumes acquired at three-second intervals on a Philips 3T scanner. Each volume consists of 24 contiguous slices of 4-mm thickness. The slices have dimensions 128×128 (1.719×1.719 mm). The stimulus alternated between a visual and motor task, each lasting for ten frames.

Both registration methods were applied to both datasets, using the first frame as the stationary frame. In an effort to remove any authentic subject motion, the motion parameter curves were detrended by subtracting their low-frequency components. These low-frequency components were computed using a one-stimulus-period windowed averaging (18 frames for dataset I, and 20 frames for dataset II). The detrended motion parameter curves were then compared to the stimulus function by way of correlation coefficient.

IV. RESULTS

A. Simulated Data

Fig. 1 shows the errors in the six motion parameters (which are stored in the rows of $\Sigma \mathbf{X}$) when the two registration methods are applied to a simulated dataset. The method labeled "Std" refers to the standard, sequential algorithm, while "SRA" refers to our simultaneous registration and activation algorithm. The errors in the standard registration method show significant correlation to the stimulus. The SRA method does not exhibit stimulus-correlated errors.

The motion estimates were used to resample the datasets before activation masks were produced. The activation masks corresponding to the dataset in Fig. 1 are shown in Fig. 2. In the mask resulting from the standard approach, regions of false-positive activation are evident on the anterior edge of the brain. The mask resulting from the SRA method looks very much like the true activation, with virtually no false-positive regions of activation.

The performance of the two methods over the various scenarios outlined in Section III-A is summarized in the false-positive and false-negative activation rates shown in Fig. 3. Consistently, the standard method resulted in approximately 50% (between 42% and 72%) more false-positives and approximately 40% (between 31% and 49%) more false-negatives than the SRA method on the datasets containing activation.

B. In Vivo Data

Improved separation of motion and activation is also evident in results obtained for *in vivo* dataset I. Fig. 4 plots the detrended motion profiles for dataset I for both the standard least-squares method and the SRA method. A marked reduction in the correlation coefficients (also shown in the figure) demonstrates the ability of the SRA method to avoid stimulus-correlated motion errors. In particular, the pitch angle estimates returned by the standard approach are visibly correlated to the stimulus (correlation coefficient is 0.735), while they are much less so in the

(a) True Activation   (b) Standard Method   (c) SRA Method
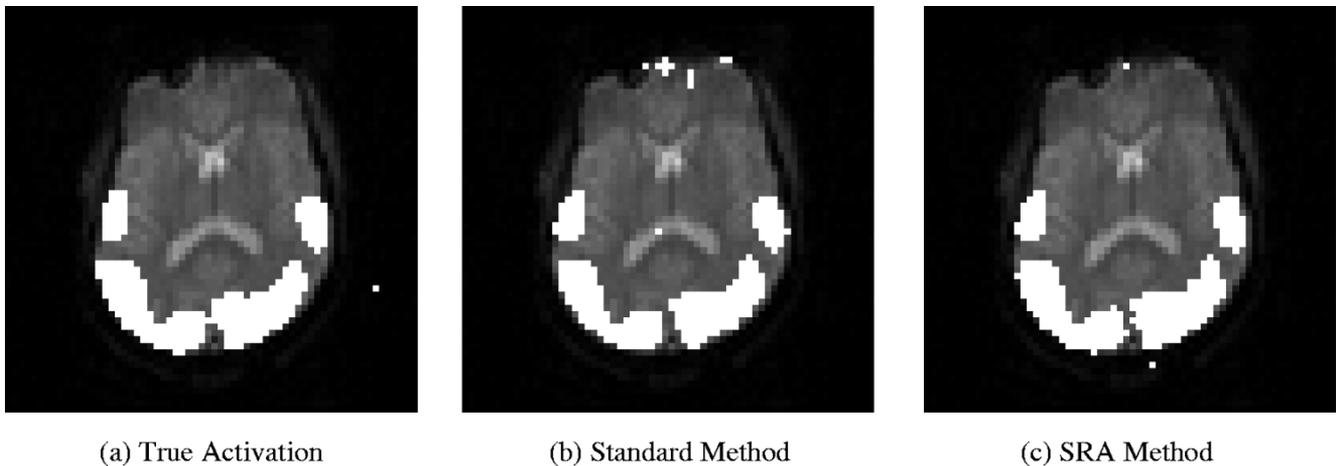
Fig. 2.   Activation maps for motion-corrupted dataset (scenario 1). The highlighted voxels exhibit high correlation to the stimulus, determined by a Student's T-test with $p = 0.001$. The activation map in (a) shows the actual, uncorrupted activation, while (b) shows the activation resulting from motion compensation based on the estimates from the standard method, and (c) shows the results after motion compensation based on the motion estimates returned by SRA.
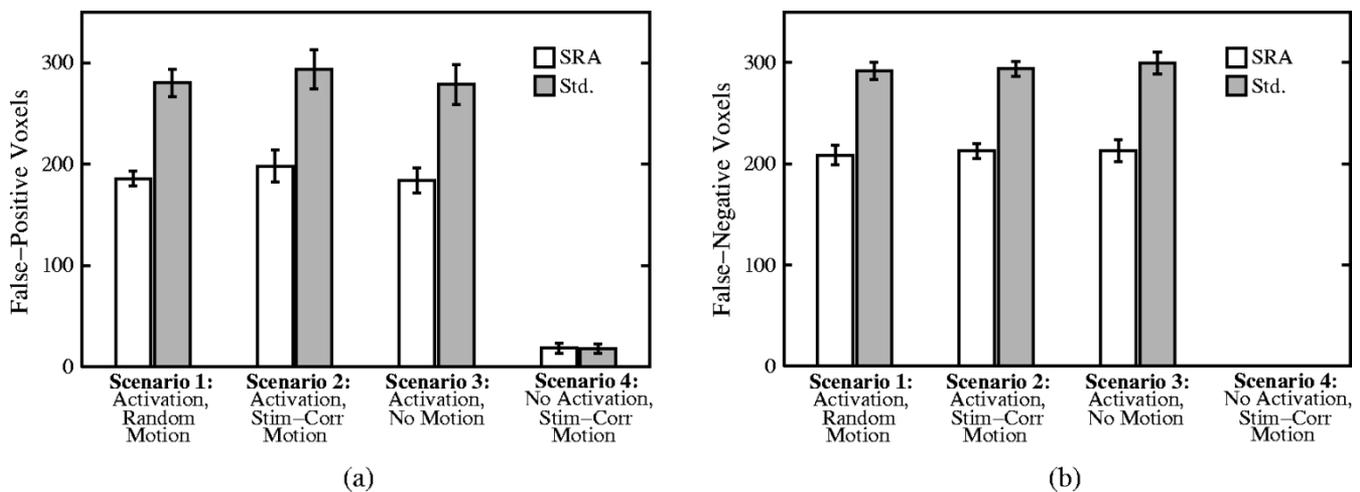
Fig. 3.   False-positive (a) and false-negative (b) activation rates for simulated datasets. Note that there are no false-negative activations for scenario 4 because these datasets have no activation.

estimates returned by the SRA method (correlation coefficient is 0.170). The remaining observable oscillatory fluctuations, especially in the posterior translation, suggest that the motion is authentic because they are not consistently in phase with the stimulus function.

The detrended motion estimates for *in vivo* dataset II are shown in Fig. 5. As with dataset I, the correlation coefficients are lower for the SRA method than for the standard least-squares registration method. Again, the subject motion seems to contain some stimulus related motion, although it is not strictly stimulus-correlated motion.

V. DISCUSSION AND CONCLUSION

Registration using the standard least-squares algorithm exhibits errors that are correlated to the BOLD signal in fMRI experiments. As MR scanners of higher field strengths become more commonly used for fMRI, the predominance of these stimulus-correlated mis-registrations can be expected to increase. Using a stimulus-contaminated motion estimate to resample a dataset has the potential to introduce voxels with the appearance of activation, resulting in false-positives. However, the simultaneous registration and activation algorithm accounts for the intensity changes due to the BOLD signal, and minimizes the sum of squares cost function for both the registration and activation detection least-squares problems in parallel. Considerable decreases in stimulus-correlated motion errors resulted in a dramatic decrease in the number of false-positive and false-negative activations.

The SRA algorithm is very robust, and can handle datasets that contain various combinations of stimulus-independent motion, authentic stimulus-correlated motion, and activation. Patient motion that is in-step with the stimulus does not confound the SRA algorithm's ability to decouple motion effects from activation effects. The result is accurate registration despite any actual confluence between patient motion and the stimulus.

Preliminary results on *in vivo* datasets indicate that the SRA method can reduce the correlation between the stimulus function and the motion estimates. While these results are promising, further tests on *in vivo* datasets with known motion will help to validate the SRA method for clinical use.
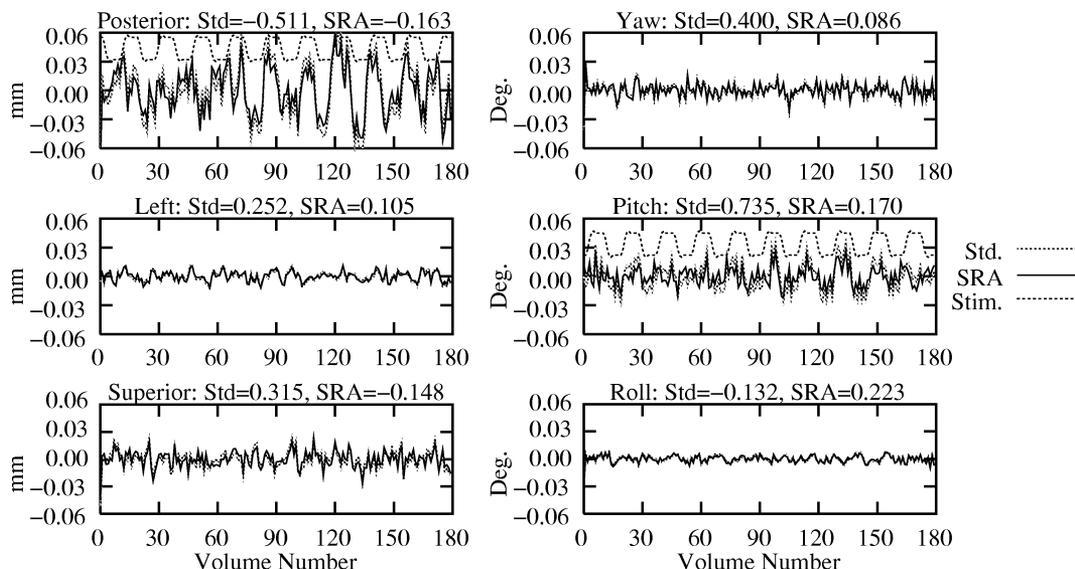
Fig. 4.   Detrended motion parameters for *in vivo* dataset I. The corresponding correlation coefficients are listed above each graph. For the purpose of comparison, the stimulus function is also represented in the posterior translation graph, as well as in the pitch angle graph.
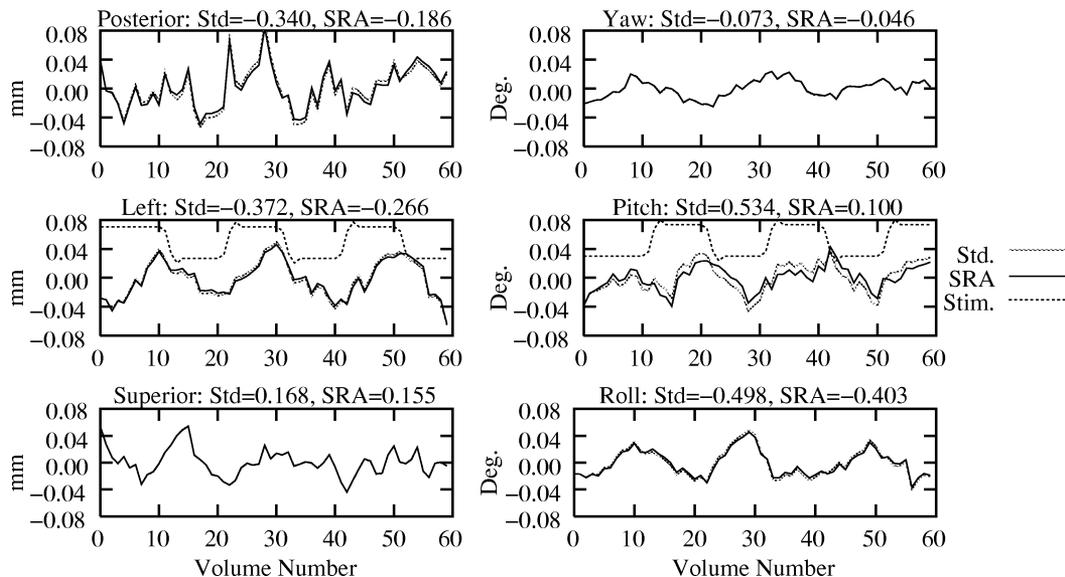
Fig. 5.   Detrended motion parameters for *in vivo* dataset II. The corresponding correlation coefficients are listed above each graph. For the purpose of comparison, the stimulus function is also represented in the left translation graph, as well as in the pitch angle graph.

One of the issues facing the solution of the combined problem is the fact that the least-squares solution is not unique. An additional constraint is needed to arrive at a single solution. As a case study for this paper, a particular constraint was used to maximize the sparsity of the activation map. However, there remains an abundance of different constraints that can be used instead.

In its current implementation, the SRA method is designed to work only on datasets that have small movements (up to 5-mm translation, and 5° rotation). This is not a serious limitation. A different registration method could be used as a first-pass registration step to bring the volumes to near-alignment. Then, the SRA method could be used to fine-tune the motion parameters and remove the activation bias. This first-pass step could also be built into the SRA program.

The simultaneous method is generalizable to scaled least-squares registration, in which a seventh parameter is added to the motion detection cost function. This parameter incorporates global intensity scaling [28] and can be used to adjust for global intensity changes, such as the slow, linear trend often observed in EPI datasets. It should be noted that including this option and the option to find the optimal baseline volume creates an additional cross-term between the registration and activation detection problems. The influence of this coupling on the final solution is an area of future investigation.

Furthermore, the method can also incorporate experimental paradigms with multiple stimulus regressors. Although this paper only includes results obtained using a single stimulus regressor, our implementation has been successfully run on experiments with two stimulus regressors, and can be run with many more. Nonstimulus regressors can also be added to the analysis (eg. physiological variation). However, in order to incorporate *any* regressor, an appropriate additional constraint

must be devised (since sparsity might not pertain). Investigation into these extensions is in progress.

The combined registration/activation model itself is amenable to solution using other cost functions. In this paper, we found the least-squares solution. But one could find the motion and activation that minimizes some other measure of the difference between the right-hand side and left-hand side of (13). The feasibility of using other cost functions needs to be investigated.

In addition to being robust and reliable, the SRA algorithm is very efficient. Even though it involves the consideration of all voxels from all time steps simultaneously, it is computationally inexpensive and its memory requirements are modest. The processing time for both the standard and SRA methods are linear in the number of volumes in the dataset. All processing for this paper was done on a 1.2-GHz PC-based computer with 512 Mb of memory. In our implementation of the SRA algorithm, processing of the 180 volumes from dataset I converged in less than 10 min after achieving motion increments of less than 0.001 mm and 0.001°. Considering the superior results of the SRA method, the stringent convergence criteria, and the fact that the SRA code can be further optimized and parallelized, the overall performance of the method is impressive and promising.

Finally, we are developing an alternative method to solve the combined problem. This method involves the iterative application of two separate programs: one that does least-squares registration, and one that does least-squares activation detection.

REFERENCES

[1] J. V. Hajnal, R. Myers, A. Oatridge, J. E. Schwieso, I. R. Young, and G. M. Bydder, "Artifacts due to stimulus correlated motion in functional imaging of the brain," *Magn. Reson. Med.*, vol. 31, pp. 283–291, 1994.

[2] K. Mathiak and S. Posse, "Evaluation of motion and realignment for functional magnetic resonance imaging in real time," *Magn. Reson. Med.*, vol. 45, pp. 167–171, 2001.

[3] V. L. Morgan, D. R. Pickens, S. L. Hartman, and R. R. Price, "Comparison of functional MRI image realignment tools using a computer-generated phantom," *Magn. Reson. Med.*, vol. 46, pp. 510–514, 2001.

[4] N. F. Ramsey, J. S. van den Brink, M. M. C. van Muiswindle, P. J. M. Folkers, and C. T. W. Moonen, "Phase navigator correction in 3D fMRI improves detection of brain activation: Qualitative assessment with a graded motor activation procedure," *NeuroImage*, vol. 8, pp. 240–248, 1998.

[5] J. V. Hajnal, N. Saeed, E. J. Soar, A. Oatridge, I. R. Yound, and G. M. Bydder, "A registration and interpolation procedure for subvoxel matching of serially acquired mr images," *J. Comput. Assist. Tomogr.*, vol. 19, no. 2, pp. 289–296, 1995.

[6] D. L. G. Hill and D. J. Hawkes, "Across-modality registration using intensity-based cost functions," in *Handbook of Medical Imaging: Processing and Analysis*, I. Bankman, Ed. New York: Academic, 2000, ch. 34, pp. 537–553.

[7] L. Freire and J.-F Mangin, "Motion correction algorithms of the brain mapping community create spurious functional activations," in *Proc. Information Processing in Medial Imaging Conf.*, Davis, CA, June 2001, pp. 246–258.

[8] ——, "Motion correction algorithms may create spurious brain activations in the absence of subject motion," *NeuroImage*, vol. 14, pp. 709–722, 2001.

[9] J. Orchard and M. S. Atkins, "Theoretical analysis of the effect of fMRI brain activation on motion correction," in Proc. 10th Annu. Meeting ISMRM, Honolulu, HI, May 2002.

[10] O. Nestares and D. J. Heeger, "Robust multiresolution alignment of MRI brain volumes," *Magn. Reson. Med.*, vol. 43, pp. 705–715, 2000.

[11] C. Nikou, F. Heitz, J.-P. Armspach, I.-J. Namer, and D. Grucker, "Registration of MR/MR and MR/SPECT brain images by fast stochastic optimization of robust voxel similarity measures," *NeuroImage*, vol. 8, pp. 30–43, 1998.

[12] L. Freire, A. Roche, and J.-F. Mangin, "What is the best similarity measure for motion correction in fMRI time series?," *IEEE Trans. Med. Imag.*, vol. 21, pp. 470–484, May 2002.

[13] L. Freire and J.-F. Mangin, "Two-stage alignment of fMRI time series using the experiment profile to discard activation-related bias," in *Lecture Notes in Computer Science*, vol. 2489, Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI'02), T. Dohi and R. Kikinis, Eds. Berlin, Germany, Sept 2002, pp. 663–670.

[14] A. Bjorck, *Numerical Methods for Least Squares Problems*. Philadelphia, PA: SIAM (Soc. Ind. Appl. Math.), 1996.

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

[16] E. Bullmore, M. Brammer, S. C. R. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham, "Statistical methods of estimation and inference for functional MR image analysis," *Magn. Reson. Med.*, vol. 5, pp. 261–277, 1996.

[17] N. Lange, "Statistical procedures for functional MRI," in *Functional MRI*, C. W. T. Moonen and P. A. Bandettini, Eds. Berlin, Germany: Springer-Verlag, 1999, pp. 301–335.

[18] P. C. Hansen, *Rank-Defficient and Discrete Ill-Posed Problems*. Philadelphia, PA: SIAM (Soc. Ind. Appl. Math.), 1998.

[19] S. Alliney and S. Ruzisky, "An algorithm for the minimization of mixed l_1 and l_2 norms with applications to bayesian estimation," *IEEE Trans. Med. Imag.*, vol. 42, pp. 618–627, Mar. 1994.

[20] A. K. Katsaggelos, J. Biemond, R. W. Schafer, and R. M. Merserau, "A regularized iterative image restoration algorithm," *IEEE Trans. Med. Imag.*, vol. 39, pp. 914–929, Apr. 1991.

[21] Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image restoration," *IEEE Trans. Med. Imag.*, vol. 5, pp. 987–995, June 1994.

[22] J. Nocedal and S. Wright, *Numerical Optimization* Springer, 1999.

[23] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.

[24] R. W. Cox, "AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages," *Comput. Biomed. Res.*, vol. 29, pp. 162–173, 1996.

[25] R. W. Cox and A. Jesmanowicz, "Real-time 3D image registration for functional MRI," *Magn. Reson. Med.*, vol. 42, pp. 1014–1018, 1999.

[26] W. F. Eddy, M. Fitzgerald, and D. C. Noll, "Improved image registration using fourier interpolation," *Magn. Reson. Med.*, vol. 36, no. 6, pp. 923–931, 1996.

[27] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art Of Scientific Computing*, 2 ed. Cambridge, U.K.: Cambridge Univ. Press, 1988.

[28] N. M. Alpert, D. Berdichevsky, Z. Levin, E. D. Morris, and A. J. Fischman, "Improved methods for image registration," *NeuroImage*, vol. 3, pp. 10–18, 1996.