Overcoming Activation-Induced Registration Errors in fMRI

Jeff J. Orchard[*a], Chen Greif[b], Gene H. Golub[c], Bruce Bjornson[d] and M. Stella Atkins[a]

[a]Computing Science, Simon Fraser University, Burnaby, BC, Canada V5A 1S6
[b]Dept. of Computer Science, Univ. of British Columbia, Vancouver, BC, Canada V6T 1Z4
[c]SCCM Program, Stanford University, Stanford, CA, USA 94305-9025
[d]Dept. of Pediatrics (Neurology), BC's Children's Hosp., Vancouver, BC, Canada V6H 3V4

ABSTRACT

It has been shown that the presence of a blood oxygen level dependent (BOLD) signal in high-field (3T and higher) fMRI datasets can cause stimulus-correlated registration errors, especially when using a least-squares registration method. These errors can result in systematic inaccuracies in activation detection. The authors have recently proposed a new method to solve both the registration and activation detection least-squares problems simultaneously. This paper gives an outline of the new method, and demonstrates its robustness on simulated fMRI datasets containing various combinations of motion and activation. In addition to a discussion of the merits of the method and details on how it can be efficiently implemented, it is shown that, compared to the standard approach, the new method consistently reduces false-positive activations by two thirds and reduces false-negative activations by one third.

Keywords: fMRI, least-squares, registration, BOLD, GLM, activation

1. INTRODUCTION

Patient motion can have a profound, negative impact on the accuracy of fMRI results. The misalignment of volumes can artificially increase the variation of a voxel's time-course, and prevent the detection of the blood oxygen level dependent (BOLD) signal.[1, 2] The resulting activation maps are often subject to false-negatives, the mis-classification of activated brain regions as inactive.

Patient motion has an opposite, and particularly insidious, effect when it is synchronized with the stimulus. This type of motion causes voxel intensity fluctuations that mimic those of activated voxels. As a result, unactivated voxels can take on the appearance of being activated.[1] A comprehensive study of the effects of stimulus-correlated motion on fMRI activation maps was conducted by Hajnal *et al.*[3] Seven of the eight *in vivo* datasets they acquired contained significant stimulus-correlated motion. They also showed that this motion, by itself, could produce the observed activation even without the BOLD signal. That study illustrates that motion cannot be assumed to be independent of the stimulation paradigm, and that stimulus-correlated motion has an overwhelming potential to produce false-positive activations.

A common approach to overcome these issues is to apply motion compensation algorithms to the dataset. The patient motion is estimated using an automatic motion detection program. These estimates can then be used to resample the dataset and align its volumes. The resulting activation maps are more accurate.[2] One of the most widely-used motion detection methods is least-squares registration. This method attempts to minimize the sum of the squared residuals between the two volumes being compared. It is computationally

---

efficient, and the sum of squares cost function has been shown to be optimal when the two volumes differ only by additive Gaussian noise.[4]

However, it has recently been shown[5] that the BOLD signal present in fMRI datasets acquired on high field scanners (3 tesla and higher) may violate the assumptions underlying the least-squares registration method. When two volumes are being compared, the presence of BOLD contrast in one image, but not the other, causes the corresponding regions to be mismatched. The impact of this phenomenon was published by Freire *et al.*[5] In his paper, he refers to these mismatched regions as "outliers" in the statistical fit. It is well-known that the sum of squares cost function is particularly sensitive to outliers. The resulting inaccurate motion estimates are then used in the motion "correction" process, removing some of the actual motion, but introducing erroneous motion. Since the false motion is caused by the brain activation, it is correlated to the stimulus. Thus, the resulting dataset has stimulus-correlated motion artifacts and is subject to the same problems as those with genuine uncorrected stimulus-correlated patient motion: the false detection of activation near high-contrast borders. Freire reported that, on datasets with voxel dimensions $3.75 \times 3.75 \times 4$ mm, motion errors as small as $0.05°$ and 0.05 mm resulted in the detection of false-positives. Freire's findings have also been supported by a theoretical study[6] in which the effect of the activation on motion estimates can be approximated linearly. The theoretical results match well with the observed registration errors from simulated fMRI datasets.

Some methods have been proposed to deal with residual motion artifacts after a first pass of registration.[1, 7, 8] In general, their approach is to remove the motion-related component from each voxel time-course. However, when patient motion is authentically stimulus-correlated, these methods have difficulty distinguishing the true BOLD signal from motion artifacts, and run the risk of discounting genuine activation.

With all the interdependencies between registration and activation, it seems that these two problems must be considered coupled, and neither should be solved before the other. Hence, a simultaneous solution is required. Such a solution method has been proposed by the authors,[9] and is called the *Simultaneous Registration and Activation* (SRA) method.

In this paper, we compare the SRA method to the standard least-squares approach in which least-squares registration is done first, followed by least-squares activation detection. Both methods are tested on a variety of different simulated datasets involving mixtures of motion and activation. We motivate and provide algorithmic details of our approach, and demonstrate its impressive performance on a comprehensive set of experiments.

2. THEORY

The registration and activation detection problems in fMRI can be combined into a single problem in which motion parameters and activation are solved for simultaneously. In this section, we outline the models for subject motion and activation separately, before combining them. This combined solution is achieved by formulating the models using matrix notation, involving all voxels and all time steps. A dataset with m volumes, each having n voxels, can be stored in a single $m \times n$ matrix, $\mathbf{F}$, as shown in Fig. 1. Each of the m volumes in the time series is stored in a column of $\mathbf{F}$, and each row holds the time series for a single voxel.

2.1. Registration

A 3D rigid-body transformation can be specified using six parameters: three rotations (about each of the three principal axes), and three translations (along each of the three principal axes). The problem of registration amounts to finding, for each volume in $\mathbf{F}$, the six motion parameters that align it with a reference volume. If we let $\mathbf{G}$ be an $m \times n$ matrix holding n copies of the reference volume, we seek to minimize the difference between $\mathbf{F}$ and transformed versions of the reference volume, $\mathbf{T}(\mathbf{G}, \mathbf{X})$, where $\mathbf{X}$ is a $6 \times n$ matrix, each column
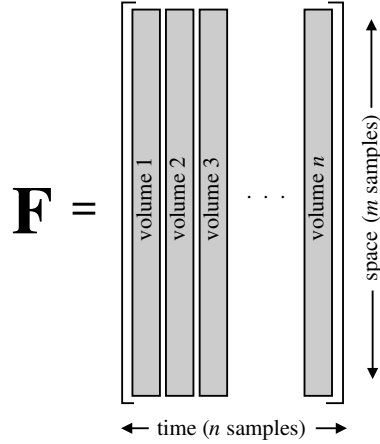
Figure 1. Storage of the time-series dataset in one matrix. Each of the n volumes is stored in a column of length m.

of which holds the 6 motion parameters for one of the n volumes:

$$\mathbf{X} = \begin{bmatrix} x^{(1)} & \cdots & x^{(n)} \\ y^{(1)} & \cdots & y^{(n)} \\ z^{(1)} & \cdots & z^{(n)} \\ \theta_x^{(1)} & \cdots & \theta_x^{(n)} \\ \theta_y^{(1)} & \cdots & \theta_y^{(n)} \\ \theta_z^{(1)} & \cdots & \theta_z^{(n)} \end{bmatrix} . \tag{1}$$

The rigid-body transformation acts non-linearly on the volume. That is, it cannot be represented as a linear operation on the intensities of a volume (not to be confused with the fact that it *is* a linear transformation on the coordinates). By taking only the linear terms of the Taylor expansion of $\mathbf{T}$ with respect to $\mathbf{X}$, we get

$$\begin{aligned} \mathbf{T}(\mathbf{G}, \mathbf{X}) &= \mathbf{T}(\mathbf{G}, 0) + \nabla \mathbf{T}(\mathbf{G}, 0)\mathbf{X} \\ &= \mathbf{G} + \nabla \mathbf{T}(\mathbf{G}, 0)\mathbf{X} . \end{aligned} \tag{2}$$

Thus, registration is the task of finding $\mathbf{X}$ so that

$$\begin{aligned} \mathbf{F} &= \mathbf{G} + \nabla \mathbf{T}(\mathbf{G}, 0)\mathbf{X} \\ &= \mathbf{G} + \mathbf{A}\mathbf{X} , \end{aligned} \tag{3}$$

where $\mathbf{A}$ and $\nabla \mathbf{T}(\mathbf{G}, 0)$ have six columns, each column holding the partial derivatives of all voxels in $\mathbf{G}$ with respect to one of the motion parameters. In the presence of noise (and other sources of variation), we do not expect to be able to solve Eqn. (3) exactly. Instead, we look to minimize some measure of the difference between the two sides of the equal sign. For least-squares registration, we solve

$$\min_{\mathbf{X}} \| \mathbf{F} - (\mathbf{G} + \mathbf{A}\mathbf{X}) \| , \tag{4}$$

where $\| \cdot \|$ is the Frobenius norm (sum of the squares of the elements). The solution to Eqn. (4) can be represented analytically as

$$\mathbf{X} = \mathbf{A}^{\dagger} (\mathbf{G} - \mathbf{F}) , \tag{5}$$

where $\mathbf{A}^\dagger$ is defined as $\left(\mathbf{A}^\mathrm{T}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}$. Recall that $\mathbf{X}$ holds the motion parameter estimates that bring $\mathbf{G}$ closer to alignment with $\mathbf{F}$. The volumes in $\mathbf{F}$ are resampled using the inverse of these motion parameters. Since we are approximating the non-linear rigid-body transformation with a linear operator (known as "linearization"), this process needs to be iterated until the system converges or it reaches a maximum number of iterations. This technique is known as fixed point iteration, and it is possible to show convergence under certain mild conditions.

2.2. Activation

The presence of BOLD activation can be represented using the general linear model (GLM). Let $\mathbf{B}$ be a $p \times n$ matrix with each of its p rows holding a time-series regressor (one or more of which is the stimulus function). The $m \times p$ matrix $\mathbf{Y}$ holds the corresponding spatial maps for the regressors. Each column of $\mathbf{Y}$ holds a spatial distribution specifying the region of influence for a particular regressor, while each row of $\mathbf{Y}$ holds the contribution of each regressor to a particular voxel. An unactivated volume $\mathbf{G}$ is related to an activated volume $\mathbf{F}$ by the formula

$$\mathbf{F} = \mathbf{G} + \mathbf{YB} \ . \tag{6}$$

Equation (6) is only valid in the absence of motion, and can also be solved in the least-squares sense yielding

$$\mathbf{Y} = \left(\mathbf{F} - \mathbf{G}\right)\mathbf{B}^\dagger \ , \tag{7}$$

where $\mathbf{B}^\dagger$ is equal to $\mathbf{B}^\mathrm{T}\left(\mathbf{BB}^\mathrm{T}\right)^{-1}$.

2.3. Combined model

Equation (3) does not incorporate temporal fluctuations such as BOLD contrast, while Eqn. (6) assumes that the volumes are perfectly aligned. In the standard implementation of fMRI processing, registration is performed first, followed by activation detection. Because of this ordering, the registration algorithm has no information about any expected intensity changes due to activation. This fact may violate the conditions under which the least-squares cost function is effective. That is, the residuals resulting from the active voxels may not fit the model of Gaussian noise, and act as outliers in the least-squares fit.

Combining Eqns. (3) and (6) gives a single model that includes both motion and activation,

$$\mathbf{F} = \mathbf{G} + \mathbf{AX} + \mathbf{YB} \quad \Leftrightarrow \quad \mathbf{AX} + \mathbf{YB} - \mathbf{C} = 0 \ , \tag{8}$$

where $\mathbf{C}$ is equal to $\left(\mathbf{F} - \mathbf{G}\right)$. Solving Eqn. (8) means finding the $\mathbf{X}$ and $\mathbf{Y}$ that minimize

$$\|\mathbf{AX} + \mathbf{YB} - \mathbf{C}\| \ . \tag{9}$$

2.4. Simultaneous solution

A simultaneous solution of Eqn. (9) yields the $(\mathbf{X}, \mathbf{Y})$ pair that minimizes the error due to mis-registration and activation at the same time. The solution can be found by using QR decomposition.[10] The QR decomposition of $\mathbf{B}^\mathrm{T}$ is

$$\mathbf{B}^\mathrm{T} = \mathbf{Q}\left[\begin{array}{c} \mathbf{R} \\ 0 \end{array}\right] \ , \tag{10}$$

where $\mathbf{Q}$ is an $n \times n$ orthogonal matrix, and the second matrix is a 2×1 block matrix in which $\mathbf{R}$ is $p \times p$ and upper-triangular, and the remaining $(n-p)$ rows are all zeros. Making the substitution for $\mathbf{B}$ in Eqn. (9), and right-multiplying by $\mathbf{Q}$ (which does not change the norm, since $\mathbf{Q}$ is unitary) gives us

$$\|\mathbf{A}\bar{\mathbf{X}} + \mathbf{Y}\left[\mathbf{R}^\mathrm{T}\ 0\right] - \bar{\mathbf{C}}\| \ , \tag{11}$$

where $\bar{\mathbf{X}}$ and $\bar{\mathbf{C}}$ are equal to $\mathbf{XQ}$ and $\mathbf{CQ}$, respectively. The cardinal feature that this algebraic manipulation elicits is the fact that the elements of $\mathbf{Y}$ are only present in the first p columns of the matrix expression in Eqn. (11). We will take advantage of this detail later on.

At this point, it is important to note that this problem does not have a unique solution. To demonstrate this, assume that $(\mathbf{X}, \mathbf{Y})$ is an exact solution to Eqn. (8). Now, consider a perturbation of that solution, $(\mathbf{X} + \delta_{\mathbf{X}}, \mathbf{Y} + \delta_{\mathbf{Y}})$. Plugging the perturbation into Eqn. (8) yields

$$
\begin{aligned}
\mathbf{A}\left(\mathbf{X} + \delta_{\mathbf{X}}\right) + \left(\mathbf{Y} + \delta_{\mathbf{Y}}\right)\mathbf{B} - \mathbf{C} &= 0 \\
\Rightarrow \mathbf{AX} + \mathbf{YB} - \mathbf{C} + \mathbf{A}\delta_{\mathbf{X}} + \delta_{\mathbf{Y}}\mathbf{B} &= 0 \\
\Rightarrow \mathbf{A}\delta_{\mathbf{X}} + \delta_{\mathbf{Y}}\mathbf{B} &= 0 .
\end{aligned}
\tag{12}
$$

Hence, the perturbation is also a solution as long as $(\delta_{\mathbf{X}}, \delta_{\mathbf{Y}})$ satisfies Eqn. (12). The solution of Eqn. (12) is $\delta_{\mathbf{X}} = \alpha\mathbf{B}$ and $\delta_{\mathbf{Y}} = -\mathbf{A}\alpha$, where α is any $6 \times p$ matrix. Thus, the general solution of Eqn. (9) is the space

$$
\left\{ (\mathbf{X} + \alpha\mathbf{B}, \mathbf{Y} - \mathbf{A}\alpha) \mid \forall \alpha \in \mathbb{R}^{6 \times p} \right\} .
\tag{13}
$$

To find a unique solution to our problem, we need to add a constraint. In our case, since we know the form of the general solution, we can tackle the problem with a two-step approach. First, we find any particular solution, $(\mathbf{X}, \mathbf{Y})$, and use it as the anchor for our general solution. Then, we use an additional constraint to choose a unique solution from our solution space.

Since any solution will do for the first step, it is convenient to set the $6p$ elements in the first p columns of $\bar{\mathbf{X}}$ to zero. This choice merely forces our corresponding solution for $\mathbf{X}$ to be orthogonal to the rows in $\mathbf{B}$. We can do this without loss of generality because our general solution gives us the opportunity to add these components back in later. Now the solution is easy to compute because $\bar{\mathbf{X}}$ and $\mathbf{Y}$ are completely decoupled:

$$
\begin{aligned}
&\min_{(\bar{\mathbf{X}}, \mathbf{Y})} \left\| \mathbf{A}\begin{bmatrix} 0 & \bar{\mathbf{X}}_{p+1:n} \end{bmatrix} + \mathbf{Y}\begin{bmatrix} \mathbf{R}^{\mathrm{T}} & 0 \end{bmatrix} - \begin{bmatrix} \bar{\mathbf{C}}_{1:p} & \bar{\mathbf{C}}_{p+1:n} \end{bmatrix} \right\| \\
&= \min_{\bar{\mathbf{X}}} \left\| \mathbf{A}\bar{\mathbf{X}}_{p+1:n} - \bar{\mathbf{C}}_{p+1:n} \right\| + \min_{\mathbf{Y}} \left\| \mathbf{Y}\mathbf{R}^{\mathrm{T}} - \bar{\mathbf{C}}_{1:p} \right\| .
\end{aligned}
\tag{14}
$$

The notation $\mathbf{C}_{p+1:n}$ represents the sub-matrix of $\mathbf{C}$ containing only columns $(p+1)$ through n, and likewise for $\bar{\mathbf{X}}$. As before, the least-squares solution can be expressed analytically as

$$
\begin{aligned}
\bar{\mathbf{X}}_{p+1:n} &= \mathbf{A}^{\dagger}\bar{\mathbf{C}}_{p+1:n} \\
\mathbf{Y} &= \bar{\mathbf{C}}_{1:p}(\mathbf{R}^{\mathrm{T}})^{-1} .
\end{aligned}
\tag{15}
$$

Finally, we calculate $\mathbf{X}$ using the formula $\mathbf{X} = \bar{\mathbf{X}}\mathbf{Q}^{\mathrm{T}}$. As with the registration procedure outlined in section 2.1, the solution obtained from Eqn. (15) constitutes a single iteration in a fixed-point iterative method. The dataset $\mathbf{F}$ is resampled based on the inverse of the motion estimates, and the process is repeated until convergence.

2.5. Constrained simultaneous solution

As mentioned above, the solution to the combined least-squares problem is not unique. We know that if $(\mathbf{X}, \mathbf{Y})$ is a solution, then every element of the set $\left\{ (\mathbf{X} + \alpha\mathbf{B}, \mathbf{Y} - \mathbf{A}\alpha) \mid \forall \alpha \in \mathbb{R}^{6 \times p} \right\}$ is also a solution. However, we can narrow down the set of plausible solutions and extract the one that is appropriate for our problem context by adding a constraint. In the case of fMRI, we expect relatively small clusters of activation, suggesting that most elements of $\mathbf{Y}$ should be zero. However, the particular solution we obtain from the method outlined in section 2.4 is likely to contain a significant $\mathbf{A}$ component. Recall that the matrix $\mathbf{A}$ holds the partial derivatives of the reference volume with respect to the motion parameters (see Eqn. (3)). These volumes highlight edges of the reference volume, and contain extended non-zero regions. We wish to remove these $\mathbf{A}$ components so

that we are left with an activation volume, $\mathbf{Y}$, that is very close to zero everywhere except at small clusters of activation.

One way to enforce this preference is to minimize the ℓ_1 norm of $(\mathbf{Y} - \mathbf{A}\alpha)$. The ℓ_1 norm has been shown to favour images with large black regions.[11]   Other formulations that yield the same desired effect are also commonly used. In our implementation, we use the constraint

$$\min_\alpha \left\{ \sum \arctan\big(c|\mathbf{Y} - \mathbf{A}\alpha|\big) \right\} \ , \tag{16}$$

for suitably-chosen constant c. Equation (16) is similar in essence to the Geman-McClure robust estimator.[12] However, we point out that we are using this constraint to negotiate through our solution space for the least-squares problem. Hence, any single solution we arrive at based on Eqn. (16) is still a least-squares solution of Eqn. (8).

Once the desired, unique $(\mathbf{X}, \mathbf{Y})$ solution is found, a single iteration of the solution process is complete. The dataset $\mathbf{F}$ is then resampled, and the process continues until the element in $\mathbf{X}$ with the largest magnitude is smaller than a user-given threshold, typically 0.001.

3. METHODOLOGY

Both the standard least-squares algorithm and the SRA algorithm were implemented in C++ for testing purposes. For the SRA method, we used the Nelder/Mead simplex method[13] to minimize Eqn. (16). Both methods used Fourier interpolation for dataset resampling.[14]

We compared the performance of the two algorithms on simulated fMRI datasets. Simulated datasets are an effective tool for assessing registration accuracy because the motion and activation are known. Datasets of 40 frames were created by taking an original $64 \times 64 \times 30$ EPI volume (voxel dimensions $3.75 \times 3.75 \times 4$ mm), applying a $3 \times 3 \times 3$ median filter, and then duplicating the volume 40 times with varying amounts of BOLD activation and motion. The motion was applied by the AFNI[15] package, using the Fourier interpolation option. The activation mask was hand-drawn over portions of the occipital and parietal regions, covering a total of 13% of the brain volume. During two stimulus epochs spanning frames 5 to 15 and frames 25 to 35, a 5% signal increase was added to activated voxels. Then, Gaussian noise with a standard deviation of 2.5% was added to all voxels, followed by smoothing using a 5 mm full-width at half-maximum Gaussian kernel.

To test the algorithms' behaviour on datasets involving varying amounts of BOLD signal and patient motion, 10 datasets were created for each of the following four scenarios.

Scenario 1: The dataset contains activation and random motion.
Scenario 2: The dataset contains 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 dataset contains no activation, but contains stimulus-correlated random motion.
Scenario 4: The dataset contains activation, but no motion.

The two registration methods were each used to generate motion estimates for each simulated dataset. These estimates were then used to resample the datasets for motion "correction".

The corrected datasets were used to generate binary activation masks, classifying each voxel as either "active" or "inactive". Classification was based on thresholding of each voxel's linear fit coefficient and correlation coefficient. A voxel's linear fit coefficient is the value of its corresponding element in the matrix $\mathbf{Y}$, and reflects
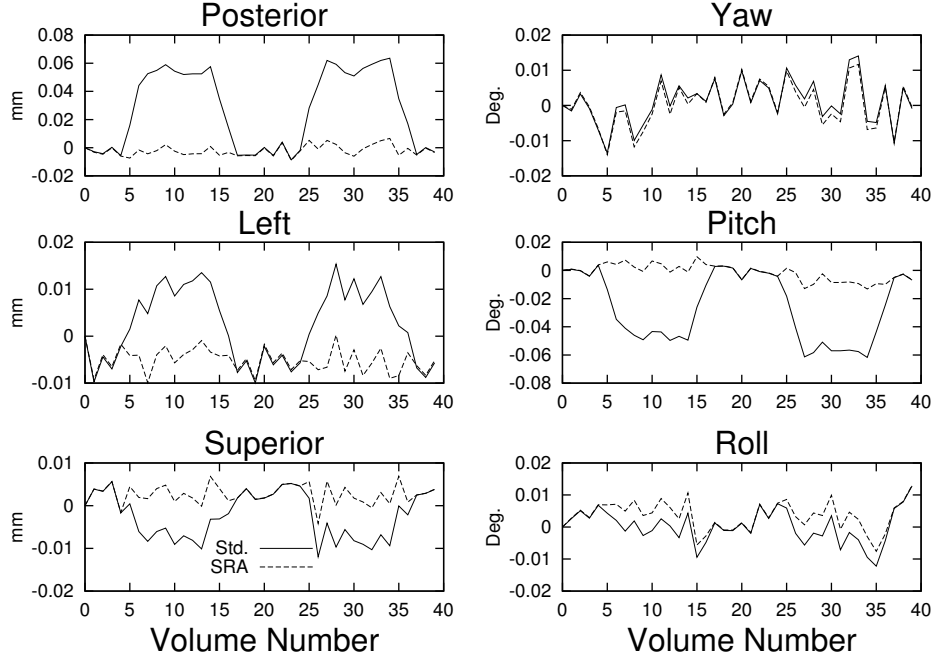
Figure 2. Typical motion detection error plots for the standard least-squares algorithm (Std) and the SRA algorithm. Notice that some of the motion parameters for the standard registration method are highly correlated to the stimulus.

the size of the stimulus-correlated component present in the voxel's time-course. The correlation coefficient between a voxel time series $\{x_i\}$ and the stimulus function $\{b_i\}$ is calculated using the formula

$$CC(x,b) = \frac{\sum_i (x_i - \bar{x})(b_i - \bar{b})}{\sqrt{\left(\sum_i (x_i - \bar{x})\right)^2 \left(\sum_i (b_i - \bar{b})\right)^2}} \ , \tag{17}$$

where $\bar{x}$ and $\bar{b}$ are the average values for the time series. If a voxel's correlation coefficient was greater (in magnitude) than 0.505 (corresponding to $P < 0.001$), and its linear fit coefficient was greater than a given threshold (about 5% of the maximum coefficient), then the voxel was classified as active and included in a binary activation mask.

A true activation map was created by a similar means, but from a dataset that was neither motion-corrupted nor motion corrected. In the true activation map, a voxel was considered active if its correlation coefficient was greater (in magnitude) than 0.505. The true activation map contained 3454 active voxels in total. Based on these activation maps, false-positive and false-negative activation counts were tabulated for each of the 40 trials.

4. RESULTS

Figure 2 shows a typical set of motion errors resulting from the different methods. Notice that the standard least-squares method often exhibits large excursions from zero during the two stimulus epochs (frames 5 to 15, and frames 25 to 35). The motion errors for the SRA method exhibit no visible stimulus-correlation.

The false-positive and false-negative activation rates for the two methods are graphed in Fig. 3. For the datasets in which activation was present (scenarios 1, 2 and 4), the SRA algorithm reduced the number of
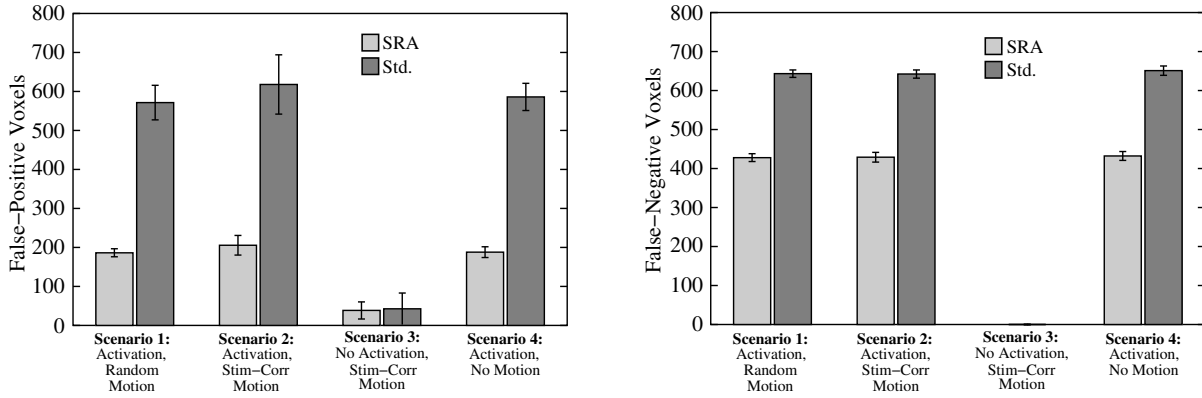
Figure 3. False-positive and false-negative activation rates for the standard least-squares algorithm and the SRA algorithm. Scenario 3 has no false-negatives because there is no activation added to the dataset and the true activation mask is blank.
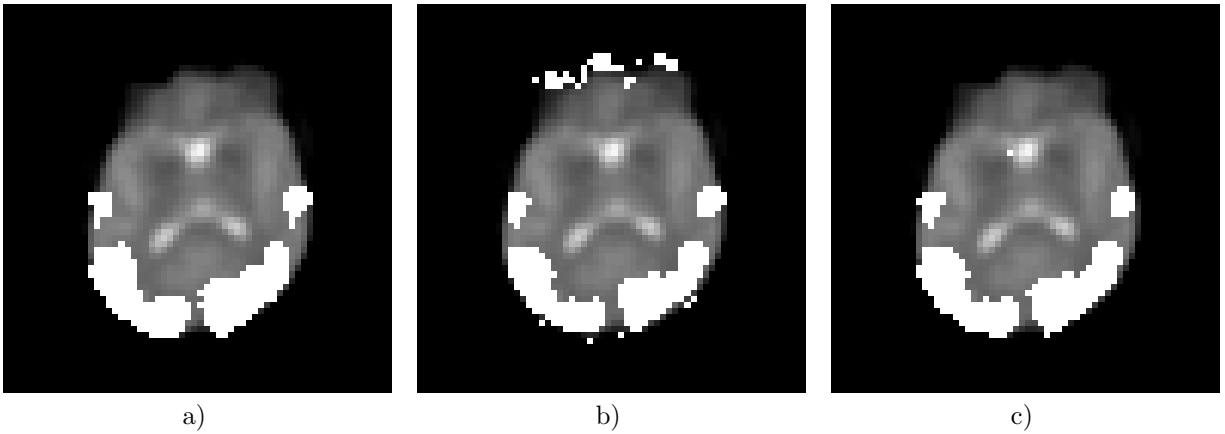

Figure 4. Activation masks for the actual (a), standard least-squares (b), and the SRA method (c) for a central slice of a scenario 1 dataset, overlaid on the corresponding EPI slice for anatomic reference.

false-positives by approximately two-thirds, and reduced the number of false-negatives by approximately one-third compared to the standard method. The two methods performed equally well on scenario 3 datasets, each averaging about 40 false-positives per trial.

Examples of typical activation masks produced in this study are shown in Fig. 4. The activation mask produced by the SRA method is very similar to the actual activation mask, and does not show any additional large clusters of activation. The mask resulting from the standard least-squares registration method shows large clusters of false-positive activation. As expected, the clusters form near contrast borders.

5. DISCUSSION AND CONCLUSIONS

By combining the registration and activation detection steps into a single optimization problem, the SRA algorithm properly decomposes the dataset into the baseline image, the BOLD signal, and motion. This algorithm avoids stimulus-correlated registration errors while still allowing for true stimulus-correlated motion to be corrected.

The SRA method consistently outperformed the standard least-squares algorithm on all simulated datasets containing activation. Use of the SRA method instead of least-squares registration resulted in a drastic reduction in false-positive and false-negative activation rates, with a drop of 67% and 33% respectively. Both the standard and SRA methods handle the activation-free trials appropriately.

Some methods attempt to remove residual motion artifacts as a part of activation detection, but run the risk of ignoring activation if it is synchronized with motion. These methods are particularly problematic when the patient moves in step with the stimulus, a phenomenon that is fairly common.[3] The SRA method effectively handles these situations, performing equally well on datasets that contain either random or genuine stimulus-correlated motion.

At the core of the simultaneous method outlined in this paper is the well-justified assumption that regions of activation form localized clusters, while the rest of the activation map is close to zero. This principle is incorporated into the model by applying an additional constraint to the optimization problem. Although excellent results were attained with the aforementioned constraints (see Eqn. (16)), other forms may yield even better results. Study of these forms remains a topic for future investigation.

Solving the combined least-squares problem is feasible on a modest computer system. This is made possible in part by a careful consideration of the numerical aspects of our implementation. All of the analysis done for this study was executed on a desktop computer with a 1.2 GHz AMD Athlon CPU and 512 MB of memory.

In conclusion, the SRA method is able to overcome the pitfalls encountered when performing registration and activation detection sequentially. It is well motivated, has a solid theoretical basis, and is shown to be robust in all tested combinations of motion and activation. Further testing of the SRA algorithm is ongoing, including demonstration on *in vivo* data, as well as comparison to other registration methods such as mutual information, robust estimators, and a customized two-stage method developed by Freire *et al.*[16]

REFERENCES

1. K. J. Friston, S. Williams, R. Howard, R. S. J. Frackowiak, and R. Turner, "Movement-related effects in fMRI time-series," *Magn. Reson. Med.* **35**, pp. 346–355, 1996.
2. 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* **8**, pp. 240–248, 1998.
3. 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.* **31**, pp. 283–291, 1994.
4. 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., ch. 34, pp. 537–553, Academic Press, 2000.
5. L. Freire and J.-F. Mangin, "Motion correction algorithms may create spurious brain activations in the absence of subject motion," *NeuroImage* **14**, pp. 709–722, 2001.
6. J. Orchard and M. S. Atkins, "Theoretical analysis of the effect of fMRI brain activation on motion correction," in *10th Annual Meeting of ISMRM*, (Honolulu), May 2002.
7. P. R. Bannister, C. F. Beckmann, M. Jenkinson, S. Smith, and J. M. Brady, "Motion artefact decorrelation in fMRI analysis using ICA," in *10th Annual Meeting of ISMRM*, (Honolulu), May 2002.

8. 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* **11**, pp. 49–57, 2000.

9. J. Orchard, C. Greif, G. Golub, B. Bjornson, and M. S. Atkins, "Simultaneous registration and activation detection for fMRI," Tech. Rep. SCCM-02-05, SCCM, Stanford University, April 2002.

10. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Third Edition, Johns Hopkins University Press, Baltimore, MD, 1996.

11. J. Nocedal and S. Wright, *Numerical Optimization*, Springer, 1999.

12. 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* **8**, pp. 30–43, 1998.

13. J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal* **7**, pp. 308–313, 1965.

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

15. R. W. Cox, "AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages," *Computers and Biomedical Research* **29**, pp. 162–173, 1996.

16. L. Freire and J.-F. Mangin, "Two-stage alignment of fMRI time series using the experiment profile to discard activation-related bias," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI'02)*, T. Dohi and R. Kikinis, eds., **2489 of LNCS**, pp. 663–670, (Tokyo), September 2002.