Improved fMRI Time-series Registration Using Joint Probability Density Priors

Roshni Bhagalia[1], Jeffrey A. Fessler[1,2], Boklye Kim[2] and Charles R. Meyer[2]

[1]Department of Electrical Engineering and Computer Science and [2]Department of Radiology
The University of Michigan, Ann Arbor, MI 418019, USA

ABSTRACT

Functional MRI (fMRI) time-series studies are plagued by varying degrees of subject head motion. Faithful head motion correction is essential to accurately detect brain activation using statistical analyses of these time-series. Mutual information (MI) based slice-to-volume (SV) registration is used for motion estimation when the rate of change of head position is large. SV registration accounts for head motion between slice acquisitions by estimating an independent rigid transformation for each slice in the time-series. Consequently each MI optimization uses intensity counts from a single time-series slice, making the algorithm susceptible to noise for low complexity end-slices (i.e., slices near the top of the head scans). This work focuses on improving the accuracy of MI-based SV registration of end-slices by using joint probability density priors derived from registered high complexity center-slices (i.e., slices near the middle of the head scans). Results show that the use of such priors can significantly improve SV registration accuracy.

Keywords: fMRI timeseries, slice-to-volume registration, mutual information, probability density priors

1. INTRODUCTION

Statistical analysis of fMRI data uses stimulus correlated variation in voxel intensities to identify brain activation. In the absence of head movement, voxels at fixed coordinate locations along the time-series can be assumed to correspond to the same physical brain region. However given the substantial time requirements and nature of fMRI studies, most subjects display varying rates of head motion with respect to the scanner. Due to the presence of the skull, and ignoring field inhomogeneity artifacts, head motion can be estimated by rigid body registration. Freire et al.[1,2] show that while fMRI time-series registration using L2 metrics may give rise to strongly biased motion estimates due to activation related intensity fluctuations, mutual information (MI) based similarity metrics are robust to such intensity variations. Thus we concentrate only on rigid time-series registration using MI.

In some block stimulus fMRI studies subjects can maintain relatively slow motion. However in other studies such as verbal tasks, jaw and mouth movement may result in fast head motion during the stimulus cycle when the subject enunciates responses. Furthermore, even non-verbal fMRI studies may elicit stronger motion in aged or invalid subjects. Two types of time-series motion estimation strategies are commonly used, viz. volume-to-volume (VV) and slice-to-volume (SV) registration. VV registration assumes that significant subject movement occurs only between volume acquisitions. Consequently a single rigid transformation is estimated and applied to all the slices in each fMRI volume, i.e., the estimated motion is constrained to be piecewise constant. On the other hand, SV registration methods allow for substantial inter-slice head motion.[3] A distinct and independent rigid motion estimate is obtained for each fMRI slice, i.e., SV registration can estimate more elaborate motion trajectories.

Although SV registration can handle a wider range of head motion, each MI approximation and optimization uses voxel intensity pairs corresponding to only one time-series slice. Hence the method is less accurate at end-slices (near the top of the head scan), where the MI-based registration is noisy due to low image complexity. Reliable end-slice registration is vital to accurately identify and correct time-series voxels affected by spin-saturation artifacts. Further, in cases where subject head motion at successive slice-acquisition time-points is correlated, accurate end-slice registration can be used to improve subsequent head motion trajectory estimation.

Recently, different ways of incorporating prior information from previously registered datasets to improve the accuracy of MI-based registration for new images from similar modalities have been proposed.[4–6] This work focuses on improving SV registration accuracy for end-slices, by using joint probability density function (pdf) priors derived from successfully registered center-slices (near the middle of the head scan) in the same time-series. We compare the accuracy of VV and SV registration with and without joint pdf priors for times-series with simulated fast and slow head motion. Results show that the proposed framework can be effective in improving the accuracy of SV time-series registration.

2. EXISTING TIME-SERIES REGISTRATION METHODS

We briefly describe the imaging model and similarity metric used to compare the performance of different registration strategies in this work. VV and SV rigid motion estimates were obtained using a Gradient Descent (GD) optimizer. The registration schemes were variations of the MIAMI-fuse[7] and MSV[3] algorithms respectively. Our implementation used a GD optimizer, instead of the Nelder-Mead simplex optimizer used in,[3,7] to improve the speed of registration. However, the inherent hill-climbing ability of the simplex optimizer may make it less susceptible to local minima than the GD optimizer. Both VV and SV registration used a high resolution anatomical T1 dataset as the homologous volume. Each VV registration used an entire fMRI volume as the reference image, while each SV registration used a single fMRI slice.

Let $\{x_i^s\}_{i=1}^M$, $x_i^s \in \mathbb{R}^3$ be coordinates of voxels in slice $s = 1, 2, \ldots S$ of an fMRI volume with corresponding intensities $\{u_i^s\}_{i=1}^M$. Similarly let the T1 volume have coordinates $\{y_j\}_{j=1}^N$, $y_j \in \mathbb{R}^3$ with intensities $\{v_j\}_{j=1}^N$. At each GD iteration, SV registration for fMRI slice s used the current estimate of the rigid transform T_{θ_s} to find transformed coordinates $\{y_i^{\theta_s} = T_{\theta_s}(x_i^s)\}_{i=1}^M$ in the T1 volume. Corresponding intensities $\{\hat{v}(y_i^{\theta_s})\}_{i=1}^M$ were then approximated using a cubic B-spline interpolation kernel.[8]

A plug-in estimate of the MI between the reference and homologous images, given by

$$
\begin{aligned}
\hat{\Psi}_{\mathrm{MI}}(\theta_s) &= \hat{H}_u + \hat{H}_v(\theta_s) - \hat{H}_{uv}(\theta_s) \\
&= -\sum_{k=1}^{K} \hat{P}_u(g_k) \log(\hat{P}_u(g_k)) - \sum_{l=1}^{L} \hat{P}_v(h_l; \theta_s) \log(\hat{P}_v(h_l; \theta_s)) \\
&\quad + \sum_{l=1}^{L} \sum_{k=1}^{K} \hat{P}_{uv}(g_k, h_l; \theta_s) \log(\hat{P}_{uv}(g_k, h_l; \theta_s)),
\end{aligned}
\tag{1}
$$

was used as the similarity metric. $\hat{H}_{uv}(\theta_s)$ is an estimate of the joint entropy between the reference and homologous images and $\hat{H}_u$ and $\hat{H}_v(\theta_s)$ are the marginal entropy estimates. $\hat{P}_v(h_l; \theta_s)$ is the approximate probability that a homologous intensity voxel $\hat{v}(y_i^{\theta_s}) \in [h_l - \eta, h_l + \eta]$; $\hat{P}_u$ and $\hat{P}_{uv}$ are defined similarly over intensity levels $g_k = g_1, g_2, \ldots, g_K$ and $h_l = h_1, h_2, \ldots, h_L$. These sets of intensity levels were chosen to span the dynamic intensity range of the reference and homologous images respectively. Our use of a GD optimizer requires that we approximate these pdfs using differentiable kernel density estimates.[8,9] VV registration followed the same framework, but used intensity pairs from all fMRI slices with rigid transformations $T_{\theta_s} = T_{\theta}, \forall s$ to estimate the joint and marginal pdfs in eq. (1).

2.1 Time-series Simulation

Two short fMRI time-series with 'slow' and 'fast' head motion were simulated to compare the performance of competing registration methods. These time-series were derived from a synthetic high resolution T2-weighted volume with $1 \times 1 \times 1$ mm^3 voxels, downloaded from the International Consortium of Brain Mapping (ICBM).[10] The data were assumed to be acquired using echo planar imaging (EPI). Head motion was simulated by rotating and translating the T2 volume in three dimensional space prior to extracting each EPI slice to form an fMRI volume. Low resolution EPI voxels were obtained by averaging voxel intensities in the corresponding $2 \times 2 \times 6$ neighborhood of the T2 volume. Gaussian noise ($N(0, 49)$) and Rayleigh noise ($\sigma = 7$) was added to voxels with non-zero and no signal intensities respectively.[11,12] Further, each slice was blurred with a 5×5 Gaussian kernel. This process was repeated to obtain sets of 40 time-series volumes with 14 slices in each volume. Each

EPI volume acquisition was assumed to be interleaved with a TR of 3000 ms, i.e., neglecting TE, one EPI slice was acquired every TR/14 ms. For registration purposes, a T1-weighted volume from ICBM was used as the anatomical reference. This T1 volume was in complete registration with the initial T2 volume used to simulate the time-series.

The applied motion was designed to be smooth without being periodic. Translations along the three axes were assumed to be linear in time with a small fixed gradient. Euler angles describing the orientation of the head at N equi-spaced time points along the entire time-series were generated by drawing three sets of N uniform i.i.d. random numbers from a $[-a, a]$ degree range. To ensure smoothness, a piecewise cubic interpolating polynomial was fit to the sequence of random angles for each coordinate axis. The orientation of the head during each intermediate EPI slice scan was obtained by sampling the polynomials at appropriate time points. The rate of change of head position was controlled by changing N and the range of rotation angles $[-a, a]$. To simulate 'slow' head motion $N = 4$ and $a = 2$ were used, while for 'fast' motion N and a were 16 and 5 respectively. The average speed of head motion at a point on the circumference of the head (assuming an average head radius of 87.5 mm) was 0.14 mm/sec for slow motion and 1.35 mm/sec for fast motion.

2.2 VV versus SV registration

To compare VV and SV registration, rigid motion estimates were obtained for each center-slice (a single slice at the middle of the head scan) and each end-slice (the superior most slice of the head scan) in both simulated time-series using both registration methods. To avoid local minima each registration was repeated 30 times with a randomly perturbed initial guess. The rigid transform estimate corresponding to the largest of the 30 similarity metric values was treated as the best estimate and used in the comparison presented here.

Registration accuracy was gauged by computing RMS errors between the known ground truth and estimated rigid motion. For rotation and translation parameters $\theta_s = [\phi_x, \phi_y, \phi_z, t_x, t_y, t_z]$, let denote T_{θ_s} the rigid SV estimates for time-series slice s in a given volume. The RMS registration error for slice s, containing M voxels at coordinates $x_i^s \in \mathbb{R}^3, i = 1, 2, \ldots M$, is given by:

$$\text{RMS error} = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left\| T_{\theta_s^*}(x_i^s) - T_{\theta_s}(x_i^s) \right\|^2}, \tag{2}$$

where $T_{\theta_s^*}$ is the known ground truth rigid transform used to simulate slice s. As VV registration computes a single rigid transform estimate T_θ for each EPI volume, RMS errors for this method were computed using $T_{\theta_s} = T_\theta, \forall s$. RMS errors for VV and SV registration for all the center-slices and end-slices are shown in Fig. 1 (slow motion) and Fig. 2 (fast motion).

These plots show that VV registration has lower error than SV for slow head motion. The piecewise constant motion approximation in VV registration is capable of capturing a slow rate of change of head position. This coupled with the reliability of the MI approximation in VV registration, which is based on a large number of intensity counts, results in better registration accuracy. In contrast the dearth of meaningful intensity counts in SV registration, especially for slices near the top of the head, makes the corresponding joint pdf estimates susceptible to spurious matches. This results in SV motion estimates that oscillate about their optimal value. However, when the velocity of head motion is high, the piecewise constant motion approximation used in VV registration is no longer valid. For rapid head motion trajectories, the approximation error in VV motion estimates is larger than the error introduced in SV motion estimates due to noisy pdf estimates.

3. IMPROVING FMRI TIME-SERIES REGISTRATION

The comparison of registration methods in the previous section illustrates a trade-off between the number of intensity counts available for joint pdf estimation and the space of allowable motion trajectories. VV registration restricts motion trajectories to be piecewise constant and uses all available voxel intensity pairs to estimate MI, while SV registration estimates an independent rigid transform for each EPI slice using MI approximations based on voxel intensity pairs corresponding to a single EPI slice only.

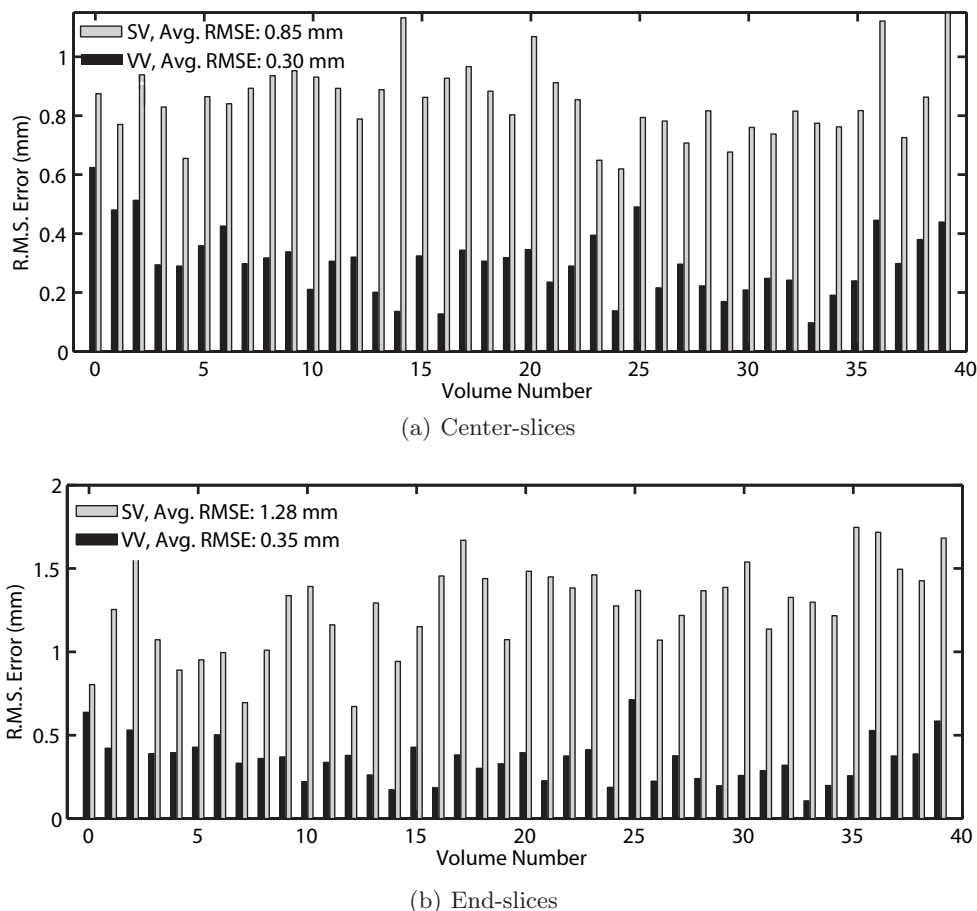(a) Center-slices

(b) End-slices

Figure 1. Comparison of RMS errors in time-series rigid motion estimates for simulated slow head motion.

It appears that a registration scheme that works well for both slow and fast head motion should handle more elaborate motion trajectories than simply piecewise constant while maintaining some level of continuity. To improve pdf estimation in SV-based registration, either more intensity pairs should be used to estimate the joint pdf or prior information about the nature of the joint pdf at registration should be employed to bolster pdf estimates when fewer intensity counts are available.

The following strategies can be used to improve SV registration of fMRI time-series data:

1. Reduce the search space, i.e., use constrained motion trajectories.[13] Alternatively, the use of suitable prior motion models may also be effective.

2. Use a pdf estimate that retains as much information about voxel intensities from the higher resolution anatomical data set as possible. In[14] such an approach was used to improve the accuracy of rigid registration between a 2D MR scout scan and a complete 3D MR brain volume.

3. Incorporate an informative prior on the nature of the joint pdf or joint histogram obtained from EPI time-series data (previously) registered onto a T1 anatomical volume.

Here we focus on using informative pdf priors to improve SV registration of time-series end slices.

3.1 Using Priors on Joint Intensity Histograms

As medical image registration becomes an integral part of surgical planning and diagnosis, large populations of registered multi-modality medical images have become available. Numerous ways of using these pre-registered
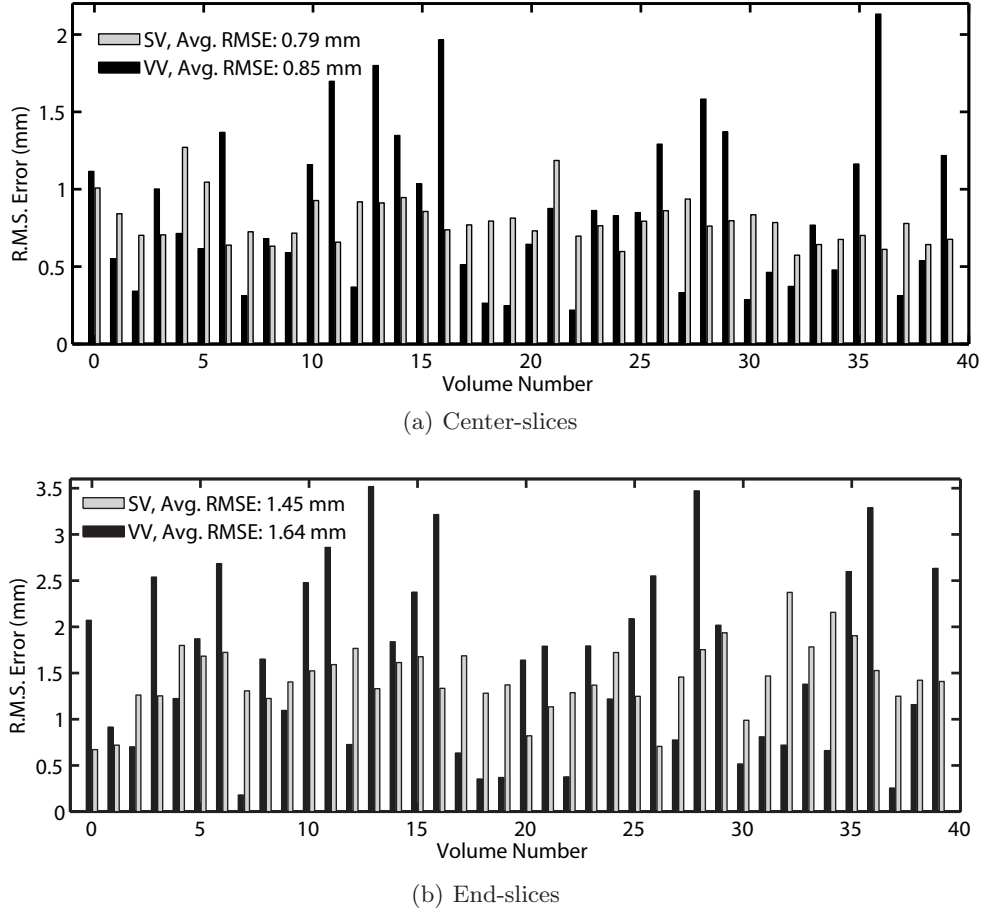
(a) Center-slices

(b) End-slices

Figure 2. Comparison of RMS errors in time-series rigid motion estimates for simulated fast head motion.

datasets to improve the accuracy of registration of new images from similar modalities have recently been proposed.[4, 5, 15]

In particular for joint pdf estimates based on discrete joint histograms, the vector of histogram bin counts $\{d_{kl}\}_{k=1,l=1}^{K,L}$, can be modelled as a Multinomial random vector with parameters $\{P_{uv}^{kl} = P_{uv}(g_k, h_l; \theta_s)\}_{k=1,l=1}^{K,L}$ and M trials:

$$P(\{d_{kl}\}_{k=1,l=1}^{K,L}) = M! \prod_{k=1,l=1}^{K,L} \frac{(P_{uv}^{kl})^{d_{kl}}}{d_{kl}!}, \tag{3}$$

where, the bin counts are computed using the M voxel pairs from EPI slice s. The corresponding joint pdf estimates given by

$$P_{uv}^{\mathrm{ML}}(g_k, h_l; \theta_s) = \frac{d_{kl}}{\sum_{kl} d_{kl}} = \frac{d_{kl}}{M}$$

are the Maximum Likelihood (ML) solution of the parameters of the Multinomial distribution. Toews et al.[5] note that such ML pdf estimation techniques (including kernel density estimation) are unreliable in the absence of sufficient intensity counts. They use uninformative uniform priors to replace ML pdf estimates by more robust Maximum a Posteriori (MAP) estimates. These uniform priors in effect discourage spurious noise matches in sparse histograms. In contrast Zollei et al.[4] propose the use of informative pdf priors. To facilitate MAP pdf estimation, they use priors based on the Dirichlet distribution with parameters $\{\alpha_{kl}\}_{k=1,l=1}^{K,L}$, $\alpha_{kl} > 0 \ \forall k, l$ given

by

$$P\big(\{P_{uv}^{kl}\}_{k=1,l=1}^{K,L}; \{\alpha_{kl}\}_{k=1,l=1}^{K,L}\big) = \Gamma\bigg(\sum_{k,l} \alpha_{kl}\bigg) \prod_{k=1,l=1}^{K,L} \frac{(P_{uv}^{kl})^{\alpha_{kl}-1}}{\Gamma(\alpha_{kl})} \tag{4}$$

This distribution is a conjugate prior on the Multinomial distribution. That is

$$\text{if} \quad P\big(\{d_{kl}\}_{k=1,l=1}^{K,L}|\{P_{uv}^{kl}\}_{k=1,l=1}^{K,L}\big) \quad \sim \quad \text{Multinom}\big(\{P_{uv}^{kl}\}_{k=1,l=1}^{K,L}, M\big)$$

$$\text{and} \quad P\big(\{P_{uv}^{kl}\}_{k=1,l=1}^{K,L}; \{\alpha_{kl}\}_{k=1,l=1}^{K,L}\big) \quad \sim \quad \text{Dirichlet}\big(\{\alpha_{kl}\}_{k=1,l=1}^{K,L}\big),$$

$$\text{then} \quad P\big(\{P_{uv}^{kl}\}_{k=1,l=1}^{K,L}|\{d_{kl}\}_{k=1,l=1}^{K,L}; \{\alpha_{kl}\}_{k=1,l=1}^{K,L}\big) \quad \propto \quad \text{Dirichlet}\big(\{\alpha_{kl}+d_{kl}\}_{k=1,l=1}^{K,L}\big);$$

where the Multinomial and Dirichlet distributions are given by (3) and (4). This formulation presents a natural way to include prior information in (sparse) histogram estimates, yielding a MAP estimate of the pdf $\{P_{uv}^{kl}\}_{k=1,l=1}^{K,L}$ given by

$$P_{uv}^{\text{MAP}}(g_k, h_l; \theta_s) = \frac{d_{kl} + \alpha_{kl} - 1}{\sum_{k,l}(d_{kl} + \alpha_{kl} - 1)},$$

where the parameters of the Dirichlet distribution $\{\alpha_{kl}\}_{k=1,l=1}^{K,L}$ represent prior histogram bin counts. In practice one uses $\alpha_{kl} \geq 1$, $\forall k, l$ to ensure that the estimated probabilities are non-negative. This approach essentially corresponds to obtaining joint pdf estimates by combining *fixed* intensity counts from pre-registered datasets and *changing* transformation-dependant intensity counts from the new un-registered datasets. In the following section we briefly outline a framework to improve SV registration of sparse end-slices by using joint pdf priors derived from information-rich center-slices in the same time-series.

3.2 Slice-to-Volume Registration with Joint Probability Density Priors

The approaches discussed in the previous section rely on pre-registered datasets from a given population to derive histogram priors for use with new datasets from the same or similar populations. However, since SV registration has low accuracy for time-series end-slices, it may not be possible to derive appropriate priors from end-slices previously aligned using SV registration. Results in Sec. 2.2 indicate that SV registration is reasonably accurate for high complexity center-slices from both simulated time-series. Hence we investigate an approach to improve SV registration for end-slices by using a joint pdf prior based on intensity counts from registered center-slices.

Fig. 3 compares an estimate of the joint pdf based on intensity pairs from an EPI center-slice and corresponding locations in the T1 anatomical volume at registration with that based on intensity pairs from an end-slice. Both joint pdfs are similar in form, indicating that suitable joint pdf priors can be derived from registered EPI center-slices. For brevity we denote slice-to-volume registration with joint pdf priors by SV-JP.
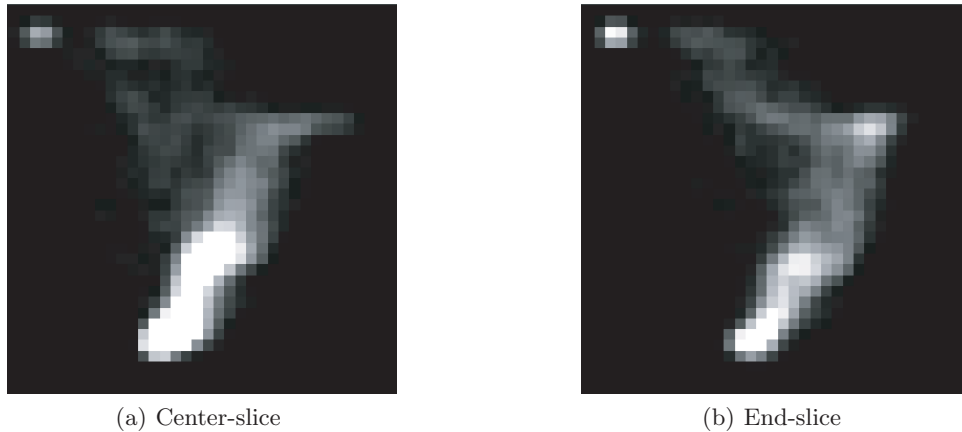

(a) Center-slice   (b) End-slice

Figure 3. Comparison of the estimated joint pdf using intensity counts from a center-slice and an end-slice at registration.

For SV-JP, we propose to replace the joint pdf $\hat{P}_{uv}(g_k, h_l; \theta_s)$ in (1) by:

$$\tilde{P}_{uv}(g_k, h_l; \theta_s, \beta) = (1 - \beta)\hat{P}_{uv}(g_k, h_l; \theta_s) + \beta P^*_{uv}(g_k, h_l); \quad (5)$$

where $g_k \in \{g_m\}_{m=1}^K$, $h_l \in \{h_n\}_{n=1}^L$ and $\beta \in [0, 1)$ is a user defined constant. The pdf estimate $\hat{P}_{uv}(g_k, h_l; \theta_s)$ depends on transformation parameters θ_s and changes with each parameter update, while $P^*_{uv}(g_k, h_l)$ is based only on prior intensity counts from pre-registered center-slices and remains fixed. Pdfs $\tilde{P}_v(h_l; \theta_s, \beta)$ and $\tilde{P}_u(g_k; \beta)$ are given by marginalizing $\tilde{P}_{uv}(g_k, h_l; \theta_s, \beta)$ over the appropriate index. The strength of the pdf prior is controlled by β. For $\beta = 0$ this approach reverts to SV registration, however, for $\beta \approx 1$ the sensitivity of $\tilde{P}_{uv}(.; \theta_s, \beta)$ to changes in $\hat{P}_{uv}(.; \theta_s)$ would be greatly reduced. Thus β should be kept smaller than 1. Further analysis of the effect of using joint pdf $\tilde{P}_{uv}(.; \theta_s, \beta)$ from (5) to compute the plug-in MI estimate in (1) is presented in appendix A.

In contrast to registration algorithms that jointly estimate constrained motion parameters for all (or a subset of) time-series slices, such as JMSV,[13] SV-JP allows for complete decoupling of slice-wise motion estimates. Hence SV-JP may be more computationally efficient and can be straight-forwardly parallelized over multiple CPUs. Further, as the similarity metric in JMSV is a sum of slice-wise MI estimates, the registration may be driven by the large number of intensity counts from time-series center-slices. This coupled with the regularization term encouraging smooth motion may make JMSV less sensitive to intensity counts from sparse end-slices. On the other hand SV-JP can be expected to show increased sensitivity to data from end-slices, as the only 'changing' intensity counts used to estimate pdf $\tilde{P}_{uv}(.; \theta_S, \beta)$ are drawn from slice S.

4. RESULTS

To compare the performance of VV, SV and SV-JP, rigid motion estimates were obtained for all end-slices in both time-series using SV-JP registration in addition to the estimates obtained using VV and SV registration in sec. 2.2. To avoid local minima each registration was repeated 30 times with a randomly perturbed initial guess, as before. Parameters corresponding to the largest of the 30 final similarity metric values for each registration were treated as the best rigid motion estimates and used in the results discussed here.

A single rigid transform was estimated for each of the 40 time-series volumes using VV registration, while SV registration was performed only for a center-slice and an end-slice from each volume. The pdf prior was obtained by averaging all 40 joint pdf estimates, obtained from each center-slice after SV registration, over time. This averaged prior pdf was then used to register all time-series end-slices using SV-JP. Prior probabilities $P^*_{uv}(g_k, h_l)$; $g_k \in \{g_m\}_{m=1}^K$, $h_l \in \{h_n\}_{n=1}^L$ with values below a user defined threshold were attributed to image noise and set to zero. SV-JP registration for end-slice S in each volume of both time-series used $\beta = \frac{N^*}{\tilde{N} + N^*}$, where $\tilde{N}$ was the number of valid intensity voxels in slice S and N^* was the average number of valid intensity voxels in a time-series center-slice.

The quality of registration for end-slices was quantified by computing RMS errors of the rigid motion estimates obtained from all three methods. SV and SV-JP registration RMS errors were computed by comparing the estimated transformed coordinates $\{T_{\theta_S}(x_i^S)\}_{i=1}^M$ with ground truth coordinates $\{T_{\theta_S^*}(x_i^S)\}_{i=1}^M$. As VV registration estimates a single rigid transform T_θ for each EPI volume, its RMS errors were obtained using $T_{\theta_S} = T_\theta$. Table 1 lists average RMS errors over both time-series.

	Avg Speed (mm/sec)	Avg. RMS Error (Std. Error) (mm)			
		No Correction	VV	SV	SV-JP
Slow motion	0.14	2.34 (0.49)	0.35 (0.13)	1.28 (0.27)	0.90 (0.26)
Fast motion	1.35	5.96 (1.56)	1.64 (0.98)	1.45 (0.37)	0.87 (0.26)

Table 1. Comparison of average RMS error values of motion estimates for times-series end-slices using VV, SV and SV-JP registration. Errors were computed for simulated slow (first row) and fast (second row) head motion.

SV-JP registration was significantly more accurate than SV registration for both slow and fast head motion trajectories. This indicates that the use of joint pdf priors derived from time-series center-slices following SV

registration can improve the accuracy of motion estimation for sparse end-slices. For very slow head motion with almost negligible inter-slice motion, VV registration was more accurate than both SV and SV-JP registration.

5. CONCLUSION

This work focused on improving MI-based SV registration of fMRI time-series for low complexity end-slices. We leveraged the reliability of SV registration of information rich center-slices to derive suitable joint pdf priors from these slices at registration. These priors were then used to improve the registration of sparse end-slices without the introduction of explicit motion constraints. Results on simulated data demonstrate a significant improvement in SV registration accuracy using this approach.

A similar approach may be used to improve SV registration for center-slices, if reliable joint pdf priors can be estimated from previously 'well-registered' time-series data. Alternatively a synthetic model of the expected joint distribution of EPI and T1 voxel pairs at registration could be used as the pdf prior.

As seen in Table 1, VV registration is well-suited to estimate slow head motion while SV and SV-JP registration are more accurate when there are faster changes in head position. Thus, an adaptive strategy that uses some supplementary information about the rate of change of head position to employ a combination of VV and SV-JP registration may further improve time-series motion correction.

APPENDIX A. EFFECT OF PDF PRIORS ON THE MI SIMILARITY METRIC

To gain some insight into how the similarity metric in SV-JP differs from that in SV we re-write the registration optimization along the lines of.[4] The estimate of rigid motion parameters obtained using SV-JP registration for fMRI slice s, $\tilde{\theta}_s$ is given by:

$$
\begin{aligned}
\tilde{\theta}_s &= \arg\max_{\theta_s} \tilde{\Psi}_{\mathrm{MI}}(\theta_s, \beta) \\
&= \arg\max_{\theta_s} \tilde{H}_v(\theta_s, \beta) - \tilde{H}_{uv}(\theta_s, \beta) \\
&= \arg\max_{\theta_s} \sum_{l=1}^{L} \sum_{k=1}^{K} \tilde{P}_{uv}(g_k, h_l; \theta_s, \beta) \log\left(\frac{\tilde{P}_{uv}(g_k, h_l; \theta_s, \beta)}{\tilde{P}_v(h_l; \theta_s, \beta)} \right),
\end{aligned}
\tag{6}
$$

where, $\tilde{H}_u$ was dropped as it is independent of the registration parameter θ_s. Using (5) to split $\tilde{P}_{uv}(g_k, h_l; \theta_s, \beta)$ the entropy terms above can be written as

$$
\tilde{H}_{uv}(\theta_s, \beta) = (1-\beta)\left(D_{\mathrm{KL}}\big(\hat{P}_{uv}(\theta_s) \| \tilde{P}_{uv}(\theta_s, \beta)\big) + \hat{H}_{uv}(\theta_s) \right) + \beta\left(D_{\mathrm{KL}}\big(P_{uv}^* \| \tilde{P}_{uv}(\theta_s, \beta)\big) + H_{uv}^* \right),
\tag{7}
$$

where, the Kullback-Leibler (KL) divergence D_{KL} is given by

$$
D_{\mathrm{KL}}\big(\hat{P}_{uv}(\theta_s) \| \tilde{P}_{uv}(\theta_s, \beta)\big) = \sum_{l=1}^{L} \sum_{k=1}^{K} \hat{P}_{uv}(g_k, h_l; \theta_s) \log \frac{\hat{P}_{uv}(g_k, h_l; \theta_s)}{\tilde{P}_{uv}(g_k, h_l; \theta_s, \beta)}.
$$

Using (7) and dropping all terms that do not depend on θ_s, the rigid motion parameters estimate $\tilde{\theta}_s$ obtained using SV-JP registration is given by

$$
\begin{aligned}
\tilde{\theta}_s &= \arg\max_{\theta_s} \Bigg\{ (1-\beta)\hat{\Phi}_{\mathrm{MI}}(\theta_s) - \beta \sum_{l=1}^{L} P_v^*(h_l) D_{\mathrm{KL}}\big(P_{u|v}^*(.|h_l) \| \tilde{P}_{u|v}(.|h_l; \theta_s, \beta)\big) \\
&\quad - (1-\beta) \sum_{l=1}^{L} \hat{P}_v(h_l; \theta_s) D_{\mathrm{KL}}\big(\hat{P}_{u|v}(.|h_l; \theta_s) \| \tilde{P}_{u|v}(.|h_l; \theta_s, \beta)\big) \Bigg\};
\end{aligned}
\tag{8}
$$

where, $\hat{\Phi}_{\mathrm{MI}}(\theta_s) \triangleq \hat{H}_v(\theta_s) - \hat{H}_{uv}(\theta_s)$. The parameter β controls the trade-off between an MI-like term that is independent of the prior and two prior-dependent KL-divergence terms. Since the only θ_s-dependent component in the KL-divergence terms is $\hat{P}_{u|v}$, these terms encourage θ_s values for which the form of the 'observed' conditional pdf, $\hat{P}_{u|v}$, is similar to that of the 'expected' conditional pdf, $P_{u|v}^*$.

When $\beta = 0$ we revert to SV registration based only on MI. SV registration finds the θ_s that yields the largest MI value, i.e., $\hat{P}_{u|v}(\theta_s)$ should be 'well-clustered'. In contrast SV-JP requires that the resulting conditional pdf $\hat{P}_{u|v}(\theta_s)$ be both 'well-clustered' and similar to the expected pdf $P_{u|v}^*$. Since SV-JP implicitly places constraints on the form of the observed pdf, it should discourage θ_s values that maximize MI but result in $\hat{P}_{u|v}(\theta_s)$ estimates that differ significantly from $P_{u|v}^*$.

REFERENCES

[1] Freire, L. and Mangin, J.-F., "Motion correction algorithms of the brain mapping community create spurious functional activations," in [*Information Processing in Medical Im.*], M F Insana, R. M. L., ed., 246–48, Springer, Berlin (2001).

[2] Freire, L., Roche, A., and Mangin, J.-F., "What is the best similarity measure for motion correction in fMRI time series?," *IEEE Trans. Med. Imag.* **21**, 470–84 (May 2002).

[3] Kim, B., Boes, J. L., Bland, P. H., Chenevert, T. L., and Meyer, C. R., "Motion correction in fMRI via registration of individual slices into an anatomical volume," *Mag. Res. Med.* **41**, 964–72 (May 1999).

[4] Zollei, L. and Wells, W., "Multi-modal image registration using dirichlet-encoded prior information," *WBIR* **LNCS 4057**, 34–42 (2006).

[5] Toews, M., Collins, D., and Arbel, T., "Maximum a posteriori local histogram estimation for image registration," *MICCAI* **LNCS 3750**, 163–170 (2005).

[6] Sabuncu, M. R. and Ramadge, P. J., "Gradient based nonuniform sampling for information theoretic alignment methods," in [*Proc. Int'l. Conf. IEEE Engr. in Med. and Biol. Soc.*], **3**, 1683–6 (2004).

[7] Meyer, C. R., Boes, J. L., Kim, B., Bland, P. H., Zasadny, K. R., Kison, P. V., Koral, K., Frey, K. A., and Wahl, R. L., "Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin plate spline warped geometric deformations," *Med. Im. Anal.* **1**, 195–206 (Apr. 1997).

[8] Thevenaz, P. and Unser, M., "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Im. Proc.* **9**, 2083–99 (Dec. 2000).

[9] Duda, R. O., Hart, P. E., and Stork, D. G., [*Pattern classification*], Wiley, New York (2001).

[10] Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., and Evans, A. C., "BrainWeb: Online interface to a 3D MRI simulated brain database," in [*Proc. 3rd Intl. Conf. on Functional Mapping of the Human Brain*], S425 (1997). NeuroImage, vol. 5 (4, part2/4), May.

[11] Buxton, R. B., [*Introduction to functional Magnetic Resonance and Imaging: Principles and Techniques*], Cambridge University Press, New York (2002).

[12] Nishimura, D. G., "Principles of magnetic resonance imaging," (1996). Unpublished textbook.

[13] Park, H., Meyer, C., and Kim, B., "Improved motion correction in fMRI by joint mapping of slices into an anatomical volume," *Proc. 7th MICCAI*, 745–751 (2004).

[14] Chandler, A. G., Netsch, T., Cocosco, C. A., Schnabel, J. A., and Hawkes, D. J., "Slice-to-volume registration using mutual information between probabilistic image classifications," in [*Medical Imaging 2004: Image Processing*], Fitzpatrick, J. M. and Sonka, M., eds., *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference* **5370**, 1120–1129 (May 2004).

[15] Sabuncu, M. and Ramadge, P., "Graph theoretic image registration using prior examples," *Proceedings of European Signal Processing Conference* (2005).