ANALYSIS AND STRATEGIES TO ENHANCE INTENSITY-BASED IMAGE REGISTRATION

by

Roshni R. Bhagalia

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2008

Doctoral Committee:

Professor Jeffrey A. Fessler, Co-chair
Research Asst. Prof. Boklye Kim, Co-chair
Professor Alfred O. Hero III
Professor Charles R. Meyer

To nana, who taught me to work while I work and to play while I play.

ACKNOWLEDGEMENTS

Many people have been instrumental in bringing this dissertation to fruition, making things at once easier and harder. I take this opportunity to thank some of the kindest, smartest, funniest and quirkiest people I know. Prof. Jeffrey Fessler for being an exceptionally patient advisor, understanding mentor and guide. Co-advisor Dr. Boklye Kim for her guidance, patience in the teething years and support. Prof. Charles Meyer for the help and encouragement, the dark chocolates and coffee and for being so wonderful to my mom when I went 'missing' in Florida. Desmond for the camaraderie and for never complaining about my hellish room-temperature preferences. Bing for the laughs and her enthusiasm. Tarry Goble, Debbie Smith, Sharon and Becky Turanski for taking the guess work out of all things administrative. Gary (resident computer-doctor), Peyton, Jin, Sarad (and his bad jokes), Jeremy and Rushali for making DIPL the fun place it is. Ms. Fox, Mrs. Dinshaw, Mrs. Thadani and Mrs. Vergese for making me fall in love with school and science. Shyam for being a sounding board for crazy ideas, for the good food and for putting up with my tantrums. Ajji and nana, my first teachers, for their love, affection and kindness. Mom and dad for making this happen and always being there.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

ANALYSIS AND STRATEGIES TO ENHANCE INTENSITY-BASED IMAGE
REGISTRATION

by

Roshni R. Bhagalia

Co-chairs: Jeffrey A. Fessler and Boklye Kim

The availability of numerous complementary imaging modalities allows us to obtain
a detailed picture of the body and its functioning. To aid diagnostics and surgical plan-
ning, all available information can be presented by visually aligning images from different
modalities using image registration. This dissertation investigates strategies to improve the
performance of image registration algorithms that use intensity-based similarity metrics.

Nonrigid warp estimation using intensity-based registration can be very time consum-
ing. We develop a novel framework based on importance sampling and stochastic ap-
proximation techniques to accelerate nonrigid registration methods while preserving their
accuracy. Registration results for simulated brain MRI data and human lung CT data
demonstrate the efficacy of the proposed framework.

Functional MRI (fMRI) is used to non-invasively detect brain-activation by acquiring
a series of brain images, called a time-series, while the subject performs tasks designed
to stimulate parts of the brain. Consequently, these studies are plagued by subject head
motion. Mutual information (MI) based slice-to-volume (SV) registration algorithms used

to estimate time-series motion are less accurate for end-slices (i.e., slices near the top of the head scans), where a loss in image complexity yields noisy MI estimates. We present a strategy, dubbed SV-JP, to improve SV registration accuracy for time-series end-slices by using joint pdf priors derived from successfully registered high complexity slices near the middle of the head scans to bolster noisy MI estimates.

Although fMRI time-series registration can estimate head motion, this motion also spawns extraneous intensity fluctuations called spin saturation artifacts. These artifacts hamper brain-activation detection. We describe spin saturation using mathematical expressions and develop a weighted-average spin saturation (WASS) correction scheme. An algorithm to identify time-series voxels affected by spin saturation and to implement WASS correction is outlined.

The performance of registration methods is dependant on the tuning parameters used to implement their similarity metrics. To facilitate finding optimal tuning parameters, we develop a computationally efficient linear approximation of the (co)variance of MI-based registration estimates. However, empirically, our approximation was satisfactory only for a simple mono-modality registration example and broke down for realistic multi-modality registration where the MI metric becomes strongly nonlinear.

CHAPTER 1

Introduction

The advent of various medical imaging modalities has allowed us to obtain a more detailed glimpse of the brain's functioning and its anatomy. The information afforded by diverse imaging modalities is usually complementary. For example, Magnetic Resonance Imaging (MRI) systems give a detailed description of brain anatomy, while Positron Emission Tomography (PET) techniques depict the functioning and metabolic activity of the brain.

Often it is advantageous to visually align images from different modalities so as to be simultaneously presented with all the available information content. This requires some spatial transformation of structures in the various images so as to bring them all into a common frame of reference. Hence, it is necessary to establish some type of one-to-one mapping between the points in each image. This mapping may be applied to an image partially or in its entirety; however to be useful, it should include all points of medical (diagnostic or surgical) importance. In image processing terminology, the process of finding this spatial transformation is called Image Registration.

To accurately describe image registration we start by defining what constitutes an image. We will restrict ourselves to medical images, such as either tomographic images like Computed Tomography, Magnetic Resonance, Ultrasound images or projection images

like conventional X-ray images. In this framework, we define an image as an array of discrete samples of a continuous function that assigns scalar intensities to two or three dimensional spatial coordinate locations. The image is usually displayed by assigning varying levels of brightness known as gray levels, to each point in the image space.

Our interest in geometrical shapes and their interrelationships requires us to impose a coordinate system on each participating image space. The points in the image space are specified by the usual Cartesian coordinates, i.e., as distances from the orthogonal coordinate system axes. Medical image registration can now be defined as the process of finding the one-to-one mapping between the coordinates in the image spaces of interest such that the points so transformed will correspond to the same anatomical point.

This spatial mapping may be modelled by rigid, affine or non-rigid transformations [48]. Rigid transformations allow only rotations and translations and preserve the distance between any two points in the image. If the transformation maps parallel lines into parallel lines it is called an affine transformation. Transformations that map lines into lines are called projective while those that map lines onto curves are called curved or non-rigid transformations. The transformations as mentioned above form a sequence of increasing sets, in that each transformation is a special case of the one succeeding it. Further, a transformation is said to be global if it applies to the entire image and is local if it is constrained to small sub-regions within the image. Most registration methods treat rigid or affine transformations as global, while using curved transformations to model local deformations.

After estimating the transformation, it is applied to the image(s) in question so as to view it in the transformed image space. The images available to us are digital, that is two types of quantization processes have been effected on them, viz. spatial quantization or sampling and intensity quantization. To view the transformed image, we need to retrieve

intensity values at its transformed coordinates from those of the acquired image. However these transformed coordinates may not correspond to those of any sample point in the digital image and hence may not have an intensity value associated with them. In such cases the intensity value at these transformed coordinates is interpolated from the pixels or voxels in their neighborhoods. This method of determining intensities at new locations not necessarily corresponding to sample points of the acquired image is called resampling.

Registration algorithms estimate the transformations either directly in a one step process or, as is more often the case, as the optima of a transformation dependent objective function. The objective function is typically some similarity measure between the two images, given a particular guess of the transformation between them. Objective functions are usually fairly simple for the single modality case; examples include correlation coefficients, correlation functions or a sum of absolute differences. However, in some situations using these similarity measurements may lead to erroneous mappings. This is because the criterion values may not account for some physically observed variations, such as changes in the amount of contrast medium during angiography or the presence of a tumor in only one image. These objective functions are not as useful when registering images from different modalities. This is due to the lack of a direct relationship between pixel or voxel intensities in images acquired using varying modalities. The multi-modality case can be tackled with objective functions that are robust to variations in the intensity value correspondences or the amount of contrast in the two images; examples include information theory related metrics.

The objective function should be constructed to be well-behaved, so that the transformation parameters optimizing it will closely approximate the true mapping that transforms one coordinate system into the other. Numerous optimization techniques are commonly used to search for the transformation parameters; non-gradient based methods like the

Nelder-Mead simplex or Powell's method may be used if the objective function is not differentiable. Gradient based techniques like Steepest-descent, Conjugate Gradients and the Levenberg-Marquardt optimization are popular when the gradient of the objective function (or an estimate of it) is available.

The registration process estimates the transformation either at each group of corresponding points in the images or only on certain specified groups of points or landmarks and then applies it to a larger region of interest. In the latter case, the registration algorithm is extrinsic, i.e., based on foreign markers introduced into the image space. These markers are designed to be well-defined and clearly visible for the relevant imaging modality. The transformation is calculated based only on the relative orientations of these markers and is then applied globally to the entire image.

In contrast, intrinsic registration methods operate on image content obtained from the subject only. If complexity is not an issue, intrinsic registration algorithms may operate directly on some function of all image gray scale values. For larger images, to reduce the search space and time complexity of the optimization process, the registration may be based on the alignment of segmented object surfaces or a smaller set of identified 'landmarks'. Landmarks are anatomical, accurately locatable points of the morphology of the visible anatomy and are usually identified interactively by the user. Landmark based registration is flexible in that, at least in theory it can be applied to any image. Also, a priori information from the user's knowledge is straightforwardly introduced in the registration process. Another possibility to reduce computation time is to use a coarse-to-fine optimization strategy that starts by estimating simple transformations for downsampled images and progessively increases both image and transformation complexity [44].

Finally, image registration can be categorized with respect to patient space [58], as intra-subject, where all the images in the registration process are from the same patient;

inter-subject where images from different patients are to be registered and atlas-based, where the images from one patient are to be registered with a statistically determined atlas.

As described above image registration is a rich field offering numerous options. One can choose from a multitude of objective functions, transformation models and optimizers. Chapter 2 gives an overview of image registration algorithms that use non-rigid transformation models and gradient-based optimizers. These registration methods have been found to be capable of handling many types of deformations. However their versatility necessitates transformations with high degrees of freedom, i.e., many parameters. This makes the computation of the gradient of the objective functions with respect to these transformation parameters very time consuming. Chapter 3 introduces a strategy employing the 'importance sampling' technique, to accelerate a class of non-rigid registration algorithms that use intensity-based objective functions. Stochastic Approximation (SA) optimization methods amenable for use with such random sampling methods are described. We conclude with registration results comparing the performance of SA optimization with importance sampling versus SA with a commonly used uniform sampling scheme and a deterministic gradient descent optimizer. Experiments include applications of importance sampling to mono-modality and multi-modality registration for both simulated and real image datasets.

Chapter 4 describes functional Magnetic Resonance Imaging (fMRI), a non-invasive imaging modality used to study brain function. This is achieved by acquiring a long sequence of images, called a time series, while the subject is performing some tasks designed to stimulate (parts of) the brain. Statistical analysis of the time series is used to detect active brain regions. Brain activation detection is plagued by subject head movement during data acquisition. Head motion alone can be estimated and compensated for by using, for

instance, slice-to-volume (SV) registration. However head motion also spawns voxel intensity fluctuations called spin saturation artifacts and further processing is necessary to combat their effect. Chapter 5 develops a method, dubbed WASS correction, to identify and correct time series voxels affected by spin saturation artifacts. We present mathematical expressions describing the spin saturation artifact and design the WASS correction starting from Bloch equations. A procedure to implement WASS correction based on SV registration motion estimates is described. Results comparing the statistical analysis of a realistic simulated fMRI time series following SV-based motion and WASS correction demonstrate the efficacy of these methods in improving activation detection.

Given the vital role of image registration to estimate motion in fMRI time series, Chapter 6 compares the performance of existing slice-to-volume (SV) and volume-to-volume (VV) time series registration methods. We analyze the shortcomings of these registration approaches and discuss possible techniques to improve SV registration. Lastly we draw on these techniques to propose a new registration scheme that we believe will combine the advantages of existing methods. The performance of the new method is evaluated using simulated time-series data.

The main contributions of this dissertation are summarized below:

1. A novel framework to accelerate nonrigid intensity-based image registration methods that use gradient optimization schemes is developed in Chapter 3. For nonrigid warps, the computation of the gradient of the similarity metric with respect to the warp parameters is very time consuming. To save time this gradient is approximated using a small random subset of image voxels [67]. We use importance sampling to improve accuracy and reduce the variance of the gradient approximation. Our framework is based on an edge-dependent adaptive sampling distribution designed for use with intensity-based registration algorithms. Results on simulated and real data show

that a combination of stochastic approximation methods and importance sampling can improve the speed of registration while preserving accuracy.

2. Spin saturation artifacts in fMRI time-series data are a manifestation of the effect of head motion during data acquisition on spin magnetization. These artifacts are motion-dependant voxel intensity fluctuations that hamper brain-activation detection using fMRI time-series. In Chapter 5 we describe the spin saturation effect using mathematical expressions and develop a weighted-average spin saturation (WASS) correction scheme starting from Bloch equations. An algorithm to identify fMRI voxels affected by spin saturation artifacts and to compensate their intensities using WASS correction is described. Results using simulated time-series data show that WASS correction can improve brain-activation detection using fMRI time-series.

3. An improved mutual information (MI) based registration method for fMRI time-series data is developed in Chapter 6. Commonly used MI-based slice-to-volume (SV) registration is shown to be less accurate at time-series end-slices (i.e., slices near the top of the head scan). This is because a loss of image complexity yields noisy probability density function (pdf) estimates, affecting the MI approximation. Results on simulated data show that using joint pdf priors derived from registered high complexity center-slices (i.e., slices near the middle of the head scan) to bolster noisy pdf estimates can improve SV registration accuracy for time-series end-slices.

4. A computationally efficient linear approximation for the covariance of registration estimates obtained by completely maximizing a differentiable plug-in MI estimate is developed in Appendix A. Such an approximation, if reasonably accurate, can be used to efficiently find 'optimal' tuning parameters (such as the window width in kernel density pdf estimates or the bin width in histogram-based pdf estimates) to improve the performance of MI-based registration. While our approximation was sat-

isfactory for a simple registration estimating a single translation between 2D mono-modality images, it broke down for more realistic multi-modality registration for which the MI similarity metric becomes strongly non-linear.

CHAPTER 2

Nonrigid Image Registration

Given a set of images of the same subject obtained in distinct coordinate systems, image registration is the process of finding transformations or warps between the image coordinates so that anatomically similar image features are in alignment. Typically, to reduce the search space the desired spatial warps are parameterized based on their nature and domain. In such cases, image registration estimates the parameters that characterize the warps.

For simplicity, consider registration between a pair of images $\{\tilde{u}_i\}_{i=1}^N$ and $\{\tilde{v}_j\}_{j=1}^M$, called the reference and homologous image respectively. These images are assumed to be arrays of discrete samples from continuous intensity functions $u(.)$ and $v(.)$ at coordinates $x_i \in \mathbb{R}^3, i = 1, 2, \ldots N$ and $y_j \in \mathbb{R}^3, j = 1, 2, \ldots M$. Let $T_{\theta_*} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ with unknown parameters θ_*, be the optimal warp that maps the homlogous image onto the reference image. Registration algorithms iteratively obtain an estimate of these parameters $\hat{\theta}$ by maximizing some similarity metric $\Psi(\theta)$ between the two images. Since only discrete image samples $\{\tilde{v}_j\}$ are available, for each parameter guess θ an approximation of the homologous image at transformed coordinates $\{y_i^\theta = T_\theta(x_i)\}$ is used to compute the similarity metric $\Psi(\theta)$.

In this framework registration consists of four major components: the deformation model used to model the warp, the interpolation kernel used to approximate the trans-

formed homologous image, the similarity metric and the optimization scheme used to estimate $\hat{\theta}$. This chapter briefly discusses some of the popular choices for each of these four components in nonrigid registration applications.

2.1 Deformation Models

Nonrigid registration is a prolific technique applied to a variety of medical image data. Numerous deformation models, including locally affine warps with very few degrees of freedom, smooth elastic deformations and models that allow each voxel to be transformed differently, have been utilized to describe the observed distortions. The more flexible the warp, the larger the number of parameters to be estimated, making nonrigid registration time consuming. In general while rigid registration may take only a few seconds; its nonrigid counterpart requires minutes or in some cases even hours depending on the deformation model.

Often in cases where bone meets soft tissue (e.g., neck, lower abdomen) the appropriate deformation model for the bone is rigid while that for soft tissue is a nonrigid warp [61, 67]. Polyrigid and Polyaffine transforms [1] are diffeomorphic (i.e., invertible and differentiable), locally rigid/affine deformation models that depend on very few parameters. Each affine transform component is specified using coordinates of its center, the associated affine transform and its radius of influence. A collection of such components is utilized, with the influence of each component waning over its radius of influence according to a smooth, spatially decaying weighting function. The effective displacement at each image coordinate is calculated by integrating the instantaneous average speed at that location due to each affine component. These transformations can be applied only to situations well-modelled by locally affine deformations.

In many medical applications where a globally nonrigid deformation model suffices,

spline-based warps using control points are common. Thin Plate Splines (TPS) popu-larized by Bookstein [4], are 2D/3D interpolating surfaces that relate control point pairs between the two images. Control points represent locations at which the spline remains fixed. The TPS warp is based on the function $U(r)$, where, r is the distance from the Carte-sian origin, $U(x, y) = U(r) = r^2 \log(r^2), r = \sqrt{x^2 + y^2}$ for 2D warps and $U(x, y, z) = U(r) = |r|^3, r = \sqrt{x^2 + y^2 + z^2}$ for 3D warps. This function is the fundamental so-lution of the biharmonic equation $\Delta^2 U = 0$, where Δ is the Laplace operator given by $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ in 2D and $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ in 3D.

For simplicity, we only describe the 2D TPS warp here. The 3D warp can be con-structed similarly using $U(r) = |r|^3$ and three dimensional control points. Given P con-trol points $(x_i, y_i); i = 1...P$ in one image, a TPS $f(x, y)$ represents a thin metal sheet that passes though all the control points and minimizes the bending energy

$$\int \int_{\mathbb{R}^2} ((\frac{\partial^2 f}{\partial x^2})^2 + 2(\frac{\partial^2 f}{\partial xy})^2 + (\frac{\partial^2 f}{\partial y^2})^2) dx dy.$$

The TPS is a sum of a linear affine part and a non-linear part given by

$$f(x, y) = \sum_{i=1}^{P} w_i U(|(x - x_i, y - y_i)|) + a_0 + a_1 x + a_2 y, \text{ where } U(r) = r^2 \log(r^2).$$

Two separate TPS functions $f_x(x, y)$ and $f_y(x, y)$ are used to model the displacements in the x and y coordinates respectively. Given P corresponding control points $z_i = (x_i', y_i'); i = 1...P$ in the other image, the weights w_i and coefficients a_0, a_1, a_2 of the func-tions are calculated to give an exact correspondence solution, i.e., $z_i = (f_x(x_i, y_i), f_y(x_i, y_i))$. A regularization term may be added when the data is noisy and an exact solution is not desirable. This warp is computationally expensive due to the increased degrees of free-dom. Furthermore, since the basis functions $U(r)$ are global, a change in any control point affects the deformation at each location in the image. A more compact local control on the warp is afforded by B-splines.

B-splines have been used to develop free-from deformations (FFD). The deformation at a particular voxel location is governed by a set of uniformly placed control points and the support of B-spline basis functions used to define the warp. For a set of $A + 1$ control points, $p_i, i = 0...A$, a smooth 1D curve can be defined using polynomial functions of degree $k; 1 \leq k \leq (A+1)$. These basis functions $B_{i,k}$ are defined using a vector of internal knots $(t_0, t_1, ..., t_{A+k})$ with $t_i < t_{i+1}, \forall i$. The curve, a piecewise continuous function whose order is independent of the number of control points, is given by:

$$C(t) = \sum_{i=0}^{A} p_i B_{i,k}(t), \ t \in (t_0, t_{A+k}).$$

These B-spline curves are affine invariant, i.e., constructing a curve from an affine image of the control points is equivalent to applying the affine transform to the original curve. The basis function $B_{i,k}$ within each knot interval is given by a recursive formula:

$$B_{i,k}(t) = \frac{t - t_i}{t_{i+k-1} - t_i} B_{i,k-1}(t) + \frac{t_{i+k} - t}{t_{i+k} - t_{i+1}} B_{i+1,k-1}(t)$$

$$\text{with } B_{i,1}(t) = \begin{cases} 1 & t_i < t < t_{i+1} \\ 0 & \text{else.} \end{cases}$$

While the spacing of the knots can be irregular, most registration algorithms use uniform knots, so that $t_{i-1} - t_i$ is constant for all i. In particular for a given spline of degree k the basis functions simply become shifted versions of each other. The above formulation can be used to generate a parameterized FFD (where the parameters are the control point locations) as follows. For simplicity we assume registration of 2D images with extents (r_0, s_0) and (r_N, s_M) in image space. Consider a grid of $A \times B$ control points $p_{ij} \in \mathbb{R}^2; i = 0...A - 1, j = 0...B - 1$; in parametric space each control point p_{ij} is initially given by $p_{ij}^0 = (i, j)$. Let the spacing of the control points in image space be Q_r and Q_s. Then given a set of control point locations $\{p_{ij}'\}$, a B-spline warp of degree k maps each

image coordinate (r, s) to new location $w(r, s) \in \mathbb{R}^2$ such that

$$(2.1) \qquad w(r, s) = \sum_{l=0}^{k} \sum_{m=0}^{k} B_{l,k}(u) B_{m,k}(v) p'_{(i+l)(j+m)}.$$

Where image space coordinate (r, s) is given by $((r - r_0)/Q_r, (s - s_0)/Q_s)$ in parametric space, $i = \left\lfloor \frac{r-r_0}{Q_r} \right\rfloor - 1$, $j = \left\lfloor \frac{s-s_0}{Q_s} \right\rfloor - 1$ and $u = \frac{r-r_0}{Q_r} - i + 1$, $v = \frac{s-s_0}{Q_s} - j + 1$. The initial mapping with $p'_{ij} = p^0_{ij}$ is given by $w(r, s) = (r, s)$.

The B-spline warp in (2.1) is $k - m$ times differentiable at locations where the same knot is duplicated m times and is one-to-one as long as it does not fold. A sufficient condition to prevent folding in a 2D cubic B-spline warp, is approximately given by $|p_{ij} - p^0_{ij}|_\infty \leq 0.48$ in parametric space. While, $|p_{ijk} - p^0_{ijk}|_\infty \leq 0.40$ is sufficient to avoid folding in a 3D cubic B-spline warp [7]. Since the deformation at image coordinates (r, s) depends only on its neighborhood of $k + 1 \times k + 1$ control points, the deformation has fine local control. Thus if a subset of the control points changes, only the affected part of the homologous image needs to be updated.

Lastly as discussed in [13] other nonrigid warps like Elastic Models treat the anatomical structures to be deformed as elastic solids. The solids are deformed according to a deformation force derived from an intensity based similarly metric between the reference and homologous images. The deforming force is opposed by an internal force dictated by the elastic model. The deformation proceeds until the two forces are in equilibrium. This model works well only with small deformations, since the linear elasticity assumption is violated for large deformations. Viscous Fluid models on the other hand can be used for very flexible deformations where each voxel can be transformed differently, however this large flexibility may lead to large mis-registration errors.

2.2 Interpolation

At each iteration, nonrigid registration algorithms use the current estimate of warp parameters θ to find homologous image coordinates $\{y_i^\theta = T_\theta(x_i)\}$ that map onto reference coordinates $\{x_i\}$. The warped homologous intensity map $\{\hat{v}_i^\theta \approx v(T_\theta(x_i))\}$ is approximated from discrete samples $\{\tilde{v}_j\}$ by modeling the underlying continuous intensity function $v(.)$ using an interpolation kernel. The approximation either exactly interpolates the discrete values or in case of noisy images approximates (smoothes) them. Various functions such as the truncated Sinc, Gaussian, Bi-linear interpolation kernels can be used for this task. Lower degree interpolants like the Bi-linear kernel with a narrow support are computationally efficient but introduce aliasing artifacts. In contrast kernels with a large support (e.g. Sinc) reduce aliasing considerably; however, apart from being computationally expensive they increase ringing artifacts. A reasonable compromise can be obtained by using a differentiable B-spline B_k of degree $k > 2$ as the interpolation kernel [70]. The continuous function $v(.)$ is approximated by a curve similar to that used in (2.1),

$$\hat{v}_i^\theta = \hat{v}(T_\theta(x_i)) = \sum_{j=1}^{M} b_j B_k(T_\theta(x_i) - y_j).$$

The coefficients b_j are computed such that $\hat{v}(y_j) = \tilde{v}_j$ and are consistent with certain boundary conditions (e.g., extending the images on either side using mirror images). Unser et al. [73] describe an efficient filter designed to calculate these coefficients from $\{\tilde{v}_j\}$. In case of noisy data, smoothing can be incorporated in the above representation.

Differentiability of the interpolation kernel is necessary when using fast gradient based optimization methods. Due to its finite support and twice-differentiability the cubic B-

spline B_4, mathematically given by

$$B_4(t) = \begin{cases} \frac{(4-6|t|^2+3|t|^3)}{6} & 0 \leq |t| \leq 1 \\ \frac{(2-|t|)^3}{6} & 1 \leq |t| \leq 2 \\ 0 & \text{else,} \end{cases}$$

is often used as the interpolation kernel.

2.3 Similarity Metrics

Image registration algorithms aim to find 'accurate' estimates of the unknown warps that will bring a set of images into alignment. These images typically have a common theme in that they are either images of the same scene taken over time or from different poses with respect to the imaging devise. While the practitioner restricts the warp estimates to a family of (usually parameterized) deformation models to reduce their search space, the chosen deformation model may not adequately describe the unknown warp. Hence, except in the case of simulations where the 'true' warp is known, the only indication of the quality of registration is some quantification of how the similarity between the image sets has improved as a result of their undergoing the estimated deformation.

Based on the type of images being registered various similarity metrics can be employed. Registration of images acquired using the same imaging modality use the simplest gauges of similarity, such as the sum of squared differences, correlation and metrics that rely on the correspondences of voxel gray-level intensities in the images. However when the images belong to multiple modalities such gray-level correspondences are lost; for instance the same tissue may appear bright in one image and greyish in another. Thus more complicated similarity measures are needed for multimodality image registration. The most prominent of these is an information theoretic approach using Mutual Information and its variations.

Mutual Information (MI) between a pair of random variables is indicative of the amount of information one random variable gives about the other and is a function of their individual and joint entropies. The entropy of a random variable is a measure of its uncertainty and quantifies the amount of information required to describe it [12]. Assuming that the reference and homologous intensity images are observations of random variables with unknown joint and marginal probability distributions P_{uv}, P_u and P_v, their marginal entropies H_u and H_v are given by

$$H_u = -\int P_u \log(P_u) dP_u \text{ and } H_v = -\int P_v \log(P_v) dP_v$$

respectively. Similarly their joint entropy H_{uv} is

$$H_{uv} = -\int P_{uv} \log(P_{uv}) dP_{uv}.$$

The MI I_{uv} between the two images is the relative entropy between their joint probability distribution and the product of their marginals

$$I_{uv} = H_u + H_v - H_{uv}.$$

In practice a plug-in estimate of MI is obtained by approximating the probability densities and replacing integrals by summations,

$$\hat{I}_{uv}^{\theta} = -\sum_{k=1}^{K} \hat{P}_u(f_k) \log(\hat{P}_u(f_k)) - \sum_{l=1}^{L} \hat{P}_v(g_l; \theta) \log(\hat{P}_v(g_l; \theta))$$

(2.2)
$$+ \sum_{l=1}^{L} \sum_{k=1}^{K} \hat{P}_{uv}(f_k, g_l; \theta) \log(\hat{P}_{uv}(f_k, g_l; \theta)).$$

$\hat{P}_u(f_k)$ is an approximation of the probability that $\tilde{u}_i \in [f_k - \epsilon, f_k + \epsilon]$, $f_k = f_1, f_2 \ldots f_K$. Similarly $\hat{P}_v(g_l; \theta)$ approximates the probability that $\hat{v}_i^{\theta} \in [g_l - \eta, g_l + \eta]$, $g_l = g_1, g_2 \ldots g_L$ and $\hat{P}_{uv}(f_k, g_l; \theta)$ is the corresponding joint probability approximation. The intensity levels $\{f_k\}_1^K$ and $\{g_l\}_1^L$ are chosen so that the probability density functions are sampled sufficiently finely. For each guess of the warp parameters θ, joint observations $(\tilde{u}_i, \hat{v}_i^{\theta})$ are

drawn from the (fixed) reference and the interpolated (warped) homologous image. Hence these joint observations vary with changes in θ; correspondingly updating the similarity metric $\hat{I}_{uv}^{\theta}$ through variations in the plug-in probability estimates $\hat{P}_{v}(.;\theta)$ and $\hat{P}_{uv}(.;\theta)$.

The differentiability of the MI approximation $\hat{I}_{uv}^{\theta}$ in (2.2) depends on the method used to approximate the joint and marginal pdfs. To ensure that this similarity measure is differentiable, kernel density estimation (given by (3.14)) can be used to estimate the pdfs [57, 74], as opposed to the histogramming method [47]. In some situations MI is known to be sensitive to the varying overlap between the reference and warped homologous image, this sensitivity may be reduced by using Normalized Mutual Information (NMI) [68]. Lastly, other information theoretic measures such as alpha-entropy can be approximated using entropic graphs and require no probability density estimation [31].

2.4   Optimization Strategies

Registration is an optimization problem that depends on the properties of the similarity metric $\Psi(\theta)$ and the deformation model used. Estimated warp parameters are obtained such that $\hat{\theta} = \arg\max_{\theta} \Psi(\theta)$.

The Nelder-Mead simplex optimizer [53] is common when the similarity metric is not differentiable and uses only cost function evaluations. A simplex in p dimensions has $p+1$ vertices and is a generalized triangle. To find the local maxima of $\Psi(\theta), \theta \in \mathbb{R}^{p}$, the optimizer is initialized with $p+1$ metric values that form the vertices of the simplex. The vertex with the smallest value is replaced by a new vertex to form a new simplex. The process continues forming a series of simplexes with varying shapes and concludes when the size of the simplex reduces significantly. The coordinates of the vertex with the largest value is the estimated function maxima.

Another optimizer often used with non-differentiable similarity metrics is Powell's

Method. Given an initial guess $\theta_0 \in \mathbb{R}^p$, this method proceeds by generating a set of search directions $(u_1, u_2...u_p)$. At each iteration n, the scheme successively finds the maxima of the function along the directions in the set, i.e., $\theta_1 = \phi_1...\phi_p = \theta_p$, where ϕ_i is the maxima along u_i. The new guess θ_{n+1} is given by the maxima of the function along the new direction $u' = \theta_1 - \theta_p$. The set of directions is updated using u' to replace one of the older directions and the process is repeated until a stopping criterion is reached.

Both Nelder-Mead and Powell's method may become inefficient when the number of parameters p to be estimated is large, since each iteration requires $p + 1$ similarity metric evaluations. First or second order differentiable similarity metrics can be employed to improve optimization speeds by using higher order gradient information. Steepest Descent (SD) is the simplest gradient based optimizer. The SD algorithm drives iterates in the direction in which the similarity metric increases fastest. The update rule is given by:

$$(2.3) \qquad \theta_{n+1} = \theta_n + a_n \nabla_\theta \Psi(\theta)|_{\theta=\theta_n} = \theta_n + a_n g(\theta_n),$$

where a_n is the step-size. The speed of convergence can be improved by choosing a_n such that update θ_{n+1} is the maxima of $\Psi(\theta)$ in the chosen direction, that is

$$\frac{\partial \Psi(\theta_{n+1})}{\partial a_n} = g(\theta_{n+1})^T \frac{\partial \theta_{n+1}}{\partial a_n} = g(\theta_{n+1})^T g(\theta_n) = 0.$$

Thus the step-size should be selected to make $g(\theta_{n+1})$ orthogonal to $g(\theta_n)$. This value of the step-size is typically computed using a line search. In applications where the line search is computation intensive, a_n is a pre-determined decreasing sequence of positive reals or is set to a fixed value. Though only one gradient calculation is needed per iteration, if the number of parameters p is large, the gradient calculation can take very long and is often the bottle-neck of the optimizer. Further SD can be plagued with very slow convergence when $\Psi(\theta)$ does not have strong gradients especially when nearing the function optima and in some cases even for long narrow peaks the optima can be strongly dependent

on the initialization.
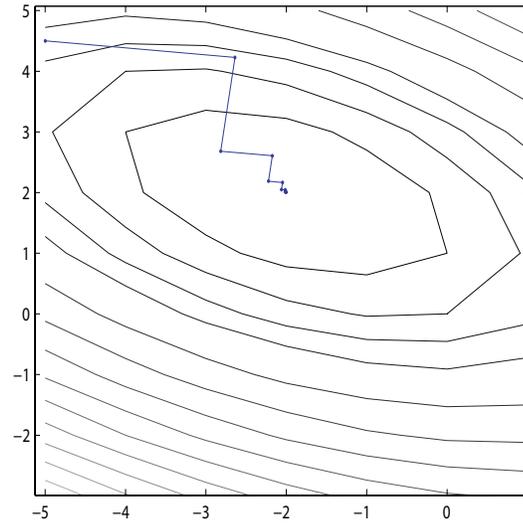
Figure 2.1: Search directions for the Steepest Descent optimizer for a simple quadratic function. Step sizes were computed to keep consecutive search directions orthogonal.

As shown in Fig. 2.1 the SD search direction at each point is orthogonal to the previous search direction. Hence the algorithm ends up looking for function optima in directions parallel to previously used search directions. This is one of the reasons for its slow convergence and is remedied by the Conjugate Gradient (CG) algorithm. The update rule for CG has the form of (2.3). However, the new direction is chosen to be an A conjugate of the old direction, i.e. such that $g(\theta_{n+1})^T A g(\theta_n) = 0$, where A is the Hessian of $\Psi(\theta)$. If the Hessian is accurate CG prevents any search direction from being repeated; specifically, CG finds the optima of a p dimensional quadratic function in exactly p iterations. In most practical cases however, $\Psi(.)$ is not quadratic and computing its Hessian is very costly. Further, approximate or inaccurate Hessians make the search directions lose their conjugacy. Variations of the search direction update rule like the Fletcher-Reeves and the Polak-Ribiere formula try to deal with this issue.

Finally though we do not discuss it here, the Levenberg-Marquardt (LM) algorithm is

commonly used in the optimization of non-linear functions. It uses a blending parameter to gradually switch between an SD-like iteration and an inverse Hessian based step-size. The LM scheme enjoys an improved rate of convergence since it uses SD-like step-sizes away from the optima and gradually switches to the inverse Hessian based step-size in low gradient regions near the function optima.

CHAPTER 3

Accelerated Intensity-based Nonrigid Image Registration [1]

Nonrigid registration algorithms estimate a warp or deformation with many degrees of freedom that appropriately maps one image onto another. This ill-posed problem is often facilitated by parameterizing the warp. Mathematically, image registration is an optimization problem:

$$(3.1) \qquad\qquad \hat{\theta} = \arg\max_{\theta} \Psi(\theta);$$

where Ψ is the similarity metric and $\hat{\theta}$ is the estimate of the p dimensional vector of warp parameters.

In registration scenarios that use differentiable intensity-based similarity metrics and gradient optimization methods, it is possible to derive an analytical expression for the gradient of the similarity metric $\nabla_{\theta}\Psi(\theta)$. However for large image volumes, the large number of warp parameters in most nonrigid registration methods makes the gradient calculation time consuming. A simple strategy to reduce this computation time is to use a small random subset of image voxels to approximate the gradient [40].

Since this randomization of the gradient in effect makes the search direction a random variable, these techniques cannot be used with algorithms like Congugate Gradients

---

[1]This chapter is based on material from [2].

that endeavor to maintain the conjugacy of successive search directions. Furthermore while it is possible to approximate the Hessian, because the random sample-size is small, its accuracy is suspect. Hence step-sizes based on the inverse of the Hessian, as in the Levenberg-Marquardt scheme, may not be reliable. It was reported in [40] that an analytical gradient-based optimizer [49, 70] using a random sub-sampling technique to approximate the gradient, performed better than that using gradient approximations based on finite differences [37] and simultaneous perturbation [66].

The speed and accuracy of such registration algorithms depend on the quality of the gradient approximation obtained via random sampling. The subset of random voxel locations is typically drawn using uniform sampling (US). Here we present an alternative data-driven, non-uniform sampling strategy that can be used efficiently to improve these gradient approximations. We argue that image edges strongly influence intensity-based registration estimates. Consequently, we propose the use of importance sampling (IS) based on a sampling distribution that emphasizes image edges to improve the gradient approximations.

Section 3.1 casts image registration in a Stochastic Approximation framework. Importance sampling is described in Sec. 3.2; a non-uniform sampling distribution for intensity-based registration is developed in Sec. 3.3; and an efficient implementation strategy is outlined in Sec. 3.5. In Sec. 3.6 we use simulated 3D MRI volumes to compare the performance of multi-modal image registration using both IS and US with that using a deterministic gradient descent optimizer. Lastly we demonstrate the application of IS to register real inhale-exhale lung CT data using deformable B-spline warps. The quality of the registration for CT data is quantified using expert identified landmarks. These results suggest that IS based on the sampling distribution designed in this work can accelerate intensity-based nonrigid registration algorithms while preserving accuracy.

3.1 Stochastic Approximation

In the random sampling framework, the registration procedure becomes a stochastic approximation technique, with the following updates:

$$\theta_{k+1} = \theta_k + a_k \hat{g}(\theta_k); \tag{3.2}$$

where θ_k is the warp parameter estimate at the kth iteration, $\hat{g}(\theta_k)$ is an approximation of the gradient $\nabla_\theta \Psi(\theta)$ at θ_k and a_k is the step-size. Stochastic approximation (SA) is used to find the zeros of a function when only noisy function evaluations are available [37,42]. SA methods aim to find the unknown zeros by successively reducing the inaccuracy in their estimates. They have been applied successfully to numerous applications in the fields of statistical modeling and controls. In gradient-based image registration, SA techniques can be used to estimate warp parameters that maximize the similarity metric by steadily reducing the imprecision introduced in successive gradient approximations.
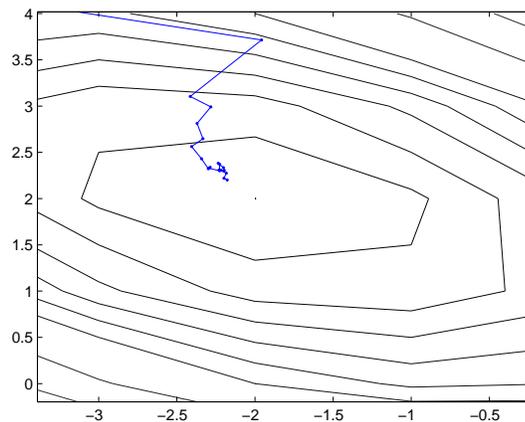
Figure 3.1: Search directions for a Stochastic Approximation optimizer for a simple quadratic function. The step-size at iteration n was $0.2/n$.

A now common SA approach was first introduced by Robbins and Monro [59]. This method aims to reduce the inaccuracy in warp parameter estimates by gradually reducing the step-size of the iterations; for brevity we call this technique Step-SA. Step-SA re-

quires that the number of points (image voxels) used to approximate the gradient, i.e., the sample-size, remains fixed over iterations. The step-size sequence, designed to guarantee convergence of the optimizer, is a non-increasing non-zero sequence $\{a_k\}, k \in \mathbb{N}$ such that $\sum_{k=1}^{\infty} a_k = \infty$ and $\sum_{k=1}^{\infty} a_k^2 < \infty$. Clearly there are numerous sequences that describe a valid step-size progression. In practice the step-size sequence is chosen heuristically for a given application.

Unlike Step-SA, sample-size controlled SA (Samp-SA) [17] keeps the step-size constant. Errors in parameter estimates are reduced by progressively increasing the sample-size used to approximate the gradient. The slowest sample-size growth rate that ensures convergence is proportional to $\ln(k)$ where k is the iteration number [17]. Using a slow growth rate should reduce computation time. In our implementation we use $K_0 k^c \ln(k + (e - 1)); 0 < c < 1$ where K_0 is the initial sample-size, as the growth rate. Both techniques effectively average out the approximation error as the iterations progress, yielding convergence.

Empirical results [2] comparing Samp-SA and Step-SA for registration of simulated brain data indicated that under identical conditions Samp-SA has faster initial convergence than Step-SA. However Step-SA appeared to be more stable at later iterations than Samp-SA. Solid lines in Fig. 3.2 indicate the mean behavior of thirty realizations for each SA method; dotted lines flanking the solid lines are +/- one standard deviation plots.

Irrespective of the SA scheme used, the efficiency of these methods for image registration applications depends on the bias and variance properties of the underlying gradient approximation based on a small random subset of image voxels. This work focuses on the use of importance sampling to enhance the performance of registration algorithms by reducing the variance of such gradient approximations without introducing any additional bias. In the following section we briefly review the variance reduction properties
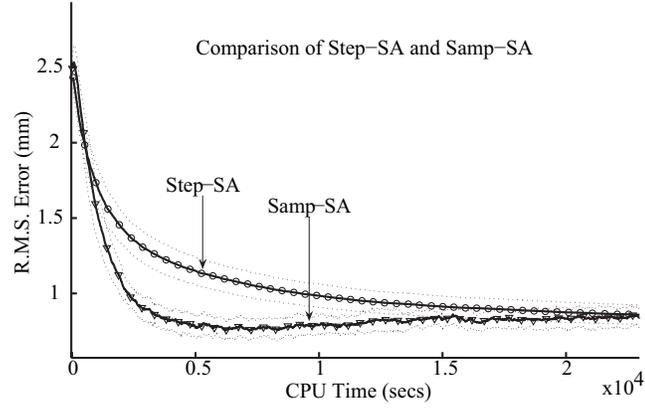
Figure 3.2: Empirical comparison of the convergence properties of Samp-SA and Step-SA under identical simulation conditions.

of importance sampling and identify image regions that strongly influence intensity-based registration. Subsequently we describe an appropriate adaptive sampling distribution that emphasizes samples from these regions. Further, a simple strategy to efficiently implement the sampling distribution is discussed.

3.2 Importance Sampling

Importance sampling (IS) is a variance reduction technique capable of incorporating knowledge of the quantity being approximated into the sampling process. IS recognizes that certain types of random samples can affect the approximation more than others and utilizes a sampling distribution that emphasizes these important samples. Such a biased distribution would produce a biased estimator; however by weighting the samples appropriately this bias can be preempted. For completeness we briefly summarize IS along the lines of [41]. To study the variance reduction afforded by IS, consider estimating a computationally intractable integral $\Phi = \int_\Omega f(u)du$. This integral can be expressed as the expectation of a (non-linear) function of a continuous uniformly distributed random vector X up to a known constant such that,

$$(3.3) \qquad \Phi = \int_\Omega f(u)du \propto E_X(f(X)), \ X \backsim P_X,$$

where $P_X(x) \triangleq U(\Omega)$ is the uniform distribution over Ω. This expectation can be re-written as,

$$
\begin{aligned}
E_X(f(X)) &= \int_\Omega f(x) P_X(x) dx \\
&= \int_\Omega \frac{f(y)}{w(y)} \hat{P}_Y(y) dy, \ w(y) = \frac{\hat{P}_Y(y)}{P_X(y)} \\
&= E_Y\left(\frac{f(Y)}{w(Y)}\right), \ Y \backsim \hat{P}_Y.
\end{aligned}
$$

(3.4)

The random vector Y is distributed according to the non-uniform distribution $\hat{P}_Y(y) = P_X(y)w(y)$. To gain any advantage by using $E_Y(.)$ over $E_X(.)$, the function $w(y)$ should be chosen carefully.

In practice, the expectations above are approximated by their sample means using N i.i.d. samples of random vectors $X \backsim U(\Omega)$ and $Y \backsim \hat{P}_Y$. Ignoring the proportionality constant, we obtain the following estimates of the integral in (3.3);

$$
\hat{\Phi}_{\text{uni}} \triangleq \frac{1}{N} \sum_{n=1}^N f(x_n) \approx E_X(f(X))
$$

$$
\hat{\Phi}_{\text{imp}} \triangleq \frac{1}{N} \sum_{n=1}^N \frac{f(y_n)}{w(y_n)} \approx E_Y\left(\frac{f(Y)}{w(Y)}\right).
$$

$\hat{\Phi}_{\text{uni}}$ corresponds to the uniform sampling case and $\hat{\Phi}_{\text{imp}}$ is the estimate obtained by importance sampling. Both $\hat{\Phi}_{\text{uni}}$ and $\hat{\Phi}_{\text{imp}}$ are unbiased with expectations proportional to the original integral in (3.3). Since the random samples are i.i.d., the variances of the two estimates are given by

$$
\text{var}(\hat{\Phi}_{\text{uni}}) = \frac{1}{N}\text{var}(f(X)) \text{ and } \text{var}(\hat{\Phi}_{\text{imp}}) = \frac{1}{N}\text{var}\left(\frac{f(Y)}{w(Y)}\right).
$$

IS based on the sampling distribution $\hat{P}_Y$ is beneficial only if $\hat{P}_Y(y) = w(y)P_X(y)$ is chosen to ensure that $\text{var}(\hat{\Phi}_{\text{imp}}) \ll \text{var}(\hat{\Phi}_{\text{uni}})$. This is possible if and only if the function $\frac{f(.)}{w(.)}$ has lower variance than $f(.)$ alone. Thus the weights $w(.)$ and correspondingly the sampling distribution $\hat{P}_Y$ should be chosen to be similar in shape to the original integrand $f(.)$, ensuring that the function $\frac{f(.)}{w(.)}$ is smooth.

3.3 Importance Sampling for Image Registration

To use importance sampling in an image registration context, we briefly outline the basic assumptions and imaging model used in a registration framework. Consider registration between a pair of intensity images, namely the reference image and the homologous image. These images are assumed to be sets of samples $\tilde{u}_i = u(x_i)$, $i = 1, 2, \ldots N_u$, and $\tilde{v}_j = v(y_j)$, $j = 1, 2, \ldots N_v$, drawn from continuous intensity functions $u(.)$ and $v(.)$ respectively. These continuous functions are sampled at coordinates $x_i \in \mathbb{R}^3$ and $y_j \in \mathbb{R}^3$ respectively.

Most nonrigid registration algorithms assume that these coordinates are related by a warp $T_{\theta_*} : \mathbb{R}^3 \to \mathbb{R}^3$. The vector of unknown warp parameters $\theta_* \in \mathbb{R}^p$ is estimated iteratively by the algorithm. At each iteration, the current estimate $\theta = \theta_k$ is used to find intensities at coordinates $\{y_i^\theta = T_\theta(x_i)\}_{i=1}^{N_u}$ in the homologous image corresponding to each reference voxel location. These transformed coordinates rarely lie on the sampling grid points and hence their corresponding intensity values $\{\hat{v}_i^\theta \approx v(T_\theta(x_i))\}$ are not known. Intensity-based similarity metrics commonly approximate these unknown intensities by modeling the continuous intensity function $v(.)$ using an appropriate interpolation kernel. Specifically, we use

$$(3.5) \qquad \hat{v}_i^\theta = \sum_{j=1}^{N_v} b_j B\big(T_\theta(x_i) - y_j\big), \quad i = 1, \ldots N_u,$$

where B is a cubic B-spline and $\{b_j\}$ are spline coefficients obtained by pre-filtering the original image $\{\tilde{v}_j\}$ [73]. Similarity metrics Ψ employing this model can be written as

$$(3.6) \qquad \Psi(\theta) = \Psi(\{\tilde{u}_i, \hat{v}_i^\theta\}_{i=1}^{N_u}).$$

Assuming differentiability and using the chain rule, the gradient of Ψ is given by

$$(3.7) \qquad g(\theta) \triangleq \nabla_\theta \Psi(\theta) = \sum_{i=1}^{N_u} \frac{\partial \Psi(\theta)}{\partial \hat{v}_i^\theta} \nabla_\theta \hat{v}_i^\theta.$$

where $\nabla_\theta = [\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_2} \dots \frac{\partial}{\partial\theta_p}]$ denotes the gradient operator.

The large number of warp parameters p and the large number of voxels N_u over which the above summation is computed makes the gradient calculation very time consuming. The above gradient can be expressed as the expectation of a nonlinear function of i.i.d samples of a uniform random variable, up to a known proportionality constant. To accelerate the gradient computation, (3.7) may be approximated by evaluating the summation over a small random subset $R \subset \{1, 2, \dots N_u\}$ of image coordinates, i.e., $\{(x_i, T_\theta(x_i))\}$; $i \in R$ are i.i.d. draws from a *uniform* sampling distribution [40, 49]. Hence, given the reference and homologous images, the approximate gradient based on uniform sampling is given by

$$\hat{g}_{\mathrm{uni}}(\theta) = \sum_{i \in R} \frac{\partial \Psi(\theta)}{\partial \hat{v}_i^\theta} \nabla_\theta \hat{v}_i^\theta.$$

Thus any voxel pair is equally likely to be used to approximate the gradient, ensuring that the resulting approximation is unbiased. Further, since x_i; $i \in R$ are i.i.d. samples, functions of these samples (under certain regularity conditions) are i.i.d. themselves.

Reducing the variance of this gradient approximation (without introducing any bias) will not only improve the convergence of the SA optimizer but may also facilitate the use of smaller sample-sizes. This may be possible by using IS to encourage denser sampling from image regions that strongly influence the gradient given by (3.7). To design a meaningful sampling distribution for gradient-based image registration, we first identify image regions that contribute significantly to the gradient of the similarity metric. These 'important' image regions can be identified by differentiating (3.5):

(3.8)
$$\nabla_\theta \hat{v}_i^\theta = \left\{ \sum_{j=1}^{N_v} b_j \dot{B}\big(T_\theta(x_i) - y_j\big) \right\} [\nabla_\theta T_\theta(x_i)],$$

where $\dot{B}(y) = \nabla_y B(y)$, $y \in \mathbb{R}^3$ is the 1×3 vector gradient of the B-spline kernel. The term in the braces contains the directional gradients or edges of the homologous intensity

image along the three coordinate axes. Recalling (3.7), only voxel intensities that lie on an edge in the homologous image $\{\hat{v}_i^\theta\}$ will contribute significantly to $g(\theta)$.

Now consider registration by swapping the two images, i.e., treating $\{\tilde{v}_j\}$ as the reference image and $\{\tilde{u}_i\}$ as the homologous image. This corresponds to finding an 'inverse' warp. In this case, the continuous function $u(.)$ will be modeled using an interpolation kernel. Repeating the above analysis, we see that edges in the swapped reference image $\{\hat{u}_j^\theta\}$ will now be vital in the gradient calculation. This suggests that our importance sampling scheme should follow a distribution that emphasizes edges in both the reference and the homologous images.

At a given SA iteration with parameter guess θ, we base the design of our θ-dependent sampling distribution P_s^θ on the edge magnitudes of the two intensity images. The probability that a voxel pair with coordinates $(x_i, T_\theta(x_i))$ is selected is chosen as follows:

$$(3.9) \qquad P_s^\theta(i) \triangleq \frac{e_i^\theta}{\sum_{j=1}^{N_u} e_j^\theta}, \quad i = 1, 2, \ldots N_u,$$

where

$$e_i^\theta \triangleq \begin{cases} \dfrac{s_i}{\sum\limits_{j=1}^{N_u} s_j} + \dfrac{t_i^\theta}{\sum\limits_{j=1}^{N_u} t_j^\theta}, & \text{if } \dfrac{s_i}{\sum\limits_{j=1}^{N_u} s_j} + \dfrac{t_i^\theta}{\sum\limits_{j=1}^{N_u} t_j^\theta} \geq T \\ \\ \epsilon & \text{else.} \end{cases}$$

$\{s_i\}_{i=1}^{N_u}$ and $\{t_i^\theta\}_{i=1}^{N_u}$ are approximate edge magnitudes of the reference and interpolated homologous images respectively. T is a user-defined edge threshold and $\epsilon \in (0, T]$. If the normalized edge magnitudes in both images are smaller than T, the sampling distribution becomes uniform with each voxel pair having an equal chance of being selected.

Let $(x_i, T_\theta(x_i)); i \in S$ where $S \subset \{1, 2, \ldots N_u\}$, be coordinates of the voxel pairs chosen according to $P_s^\theta(i)$. Then the importance sampling-based approximate gradient to be used in (3.2) is given by

$$(3.10) \qquad \hat{g}(\theta) = \sum_{i \in S} \frac{1}{w(i)} \frac{\partial \Psi(\theta)}{\partial \hat{v}_i^\theta} \nabla_\theta \hat{v}_i^\theta,$$

where $w(i) = N_u P_s^\theta(i)$. This approximate gradient uses only $|S| \ll N_u$ voxel pairs; hence the time consuming sum in (3.7) is evaluated only at these $|S|$ sample points.

Interestingly, Sabuncu et al. [63] recently developed an edge-dependent sampling scheme to reduce the approximation error in their Euclidean Minimum Spanning Trees (EMST) based registration. However, they did not study the variance-bias properties of their approximation and assigned the same weight to all samples.

3.3.1 Application to Two Similarity Metrics

We demonstrate the use of IS for image registration with two commonly used intensity-based similarity metrics. Due to its simplicity, monomodality registration algorithms often use the (negative of) sum of squared differences (SSD) as a similarity metric. In this case, both the reference and homologous images are assumed to be noisy realizations drawn from the same continuous function. Let the reference image be given by a set of noisy samples $\{\tilde{u}_i\}_{i=1}^{N_u}$. Then the negative SSD similarity metric is

$$(3.11) \qquad \Psi_{\text{SSD}}(\theta) = -\frac{1}{N_u} \sum_{i=1}^{N_u} \left(\tilde{u}_i - \hat{v}_i^\theta \right)^2,$$

where the interpolated homologous image $\{\hat{v}_i^\theta\}_{i=1}^N$ is given by (3.5). By differentiating the above expression, image edges can be easily shown to be important in the gradient calculation of the negative SSD metric.

As discussed in Chapter 2, MI is a prevalent similarity metric for multimodality registration. The plug-in MI estimate between the two images given by (2.2) is repeated here for clarity;

$$\Psi_{\text{MI}}(\theta) \;=\; -\sum_{k=1}^{K} \hat{P}_u(f_k) \log(\hat{P}_u(f_k)) - \sum_{l=1}^{L} \hat{P}_v(g_l;\theta) \log(\hat{P}_v(g_l;\theta))$$

$$(3.12) \qquad + \sum_{l=1}^{L} \sum_{k=1}^{K} \hat{P}_{uv}(f_k, g_l; \theta) \log(\hat{P}_{uv}(f_k, g_l; \theta)).$$

$\hat{P}_v(g_l; \theta)$ is the approximate probability that a homologous intensity voxel $\hat{v}_i^\theta \in [g_l - \eta, g_l + \eta]$; $\hat{P}_u$ and $\hat{P}_{uv}$ are defined similarly over intensity levels $f_k = f_1, f_2, \ldots f_K$ and $g_l = g_1, g_2, \ldots g_L$. These sets of intensity levels $\{f_k\}_1^K$ and $\{g_l\}_1^L$ are chosen to span the dynamic intensity range of the reference and homologous images respectively. Our interest in gradient based optimizers requires that we approximate these pdfs using a differentiable kernel density estimate. In this case the gradient of the MI metric w.r.t. θ is given by

$$
\begin{aligned}
\nabla_\theta \Psi_{\mathrm{MI}}(\theta) \; = \; & -\sum_{l=1}^L \nabla_\theta \hat{P}_v(g_l; \theta) \big(\log(\hat{P}_v(g_l; \theta)) + 1 \big) \\
& + \sum_{l=1}^L \sum_{k=1}^K \nabla_\theta \hat{P}_{uv}(f_k, g_l; \theta) \big(\log(\hat{P}_{uv}(f_k, g_l; \theta)) + 1 \big).
\end{aligned}
$$

(3.13)

The differentiable kernel density estimate $\hat{P}_v(g_l; \theta)$ is given by [15],

(3.14)
$$
\hat{P}_v(g_l; \theta) = \frac{1}{N_u h} \sum_{i=1}^{N_u} \kappa \left(\frac{\hat{v}_i^\theta - g_l}{h} \right),
$$

where, $\kappa(.)$ is a differentiable density kernel that integrates to unity and $h > 0$ is the scalable width of the kernel. The gradient of $\hat{P}_v(g_l; \theta)$ with respect to the warp parameters is given by

(3.15)
$$
\nabla_\theta \hat{P}_v(g_l; \theta) = \frac{1}{N_u h^2} \sum_{i=1}^{N_u} \dot{\kappa} \left(\frac{\hat{v}_i^\theta - g_l}{h} \right) \nabla_\theta \hat{v}_i^\theta;
$$

where $\dot{\kappa}(.)$ is the derivative of the density kernel. This expression is similar to (3.7); substituting (3.8) for $\nabla_\theta \hat{v}_i^\theta$ above yields

$$
\nabla_\theta \hat{P}_v(g_l; \theta) = \frac{1}{N_u h^2} \sum_{i=1}^{N_u} \dot{\kappa} \left(\frac{\hat{v}_i^\theta - g_l}{h} \right) \left\{ \sum_{j=1}^{N_v} b_j \dot{B}(T_\theta(x_i) - y_j) \right\} \nabla_\theta T_\theta(x_i).
$$

The term in the parenthesis contains the edges of the homologous image along each coordinate axis. At a fixed intensity level g_l, only voxels that lie on an edge in the homologous image and whose intensity is within $[g_l - h/2, g_l + h/2]$ will contribute to $\nabla_\theta \hat{P}_v(g_l; \theta)$. Since the intensity levels $\{g_l\}_1^L$ are chosen to span the range of homologous image intensities, every voxel in the edge map of this image will belong to the neighborhood of at least

one intensity level. This implies that the entire edge map influences the gradient calculation. Similar considerations apply to $\hat{P}_{uv}(.; \theta)$, indicating that edges in the homologous image are important for its gradient approximation too.

Consider now a registration scheme to find the 'inverse' mapping from the reference to the homologous image, by swapping the two images. In this case, repeating the above analysis illustrates that the reference image edges are also significant players in the gradient approximation.

In our MI implementation we use a cubic B-spline $B(.)$ as the differentiable density kernel [70]. The number of intensity levels K and L at which to compute $\hat{P}_u(.), \hat{P}_v(.; \theta)$ and $\hat{P}_{uv}(.; \theta)$ was chosen to be proportional to the number of voxel pairs used to compute the gradient of MI. The number of levels was approximately given by Scott's normal reference rule [64]:

$$\text{No. of intensity levels} \approx \frac{\text{range of intensity values}}{3.49 \hat{\sigma} n^{-1/3}},$$

where, $\hat{\sigma}$ is the approximate standard deviation and n is the number of voxel pairs used to estimate the pdf. Intensity levels $\{f_k\}_{k=1}^K$ and $\{g_l\}_{l=1}^L$ were chosen to uniformly span the range of intensities in the reference and homologous image repectively.

As discussed in [70], an added advantage of using the cubic B-spline $B(.)$ as the density kernel, is that it satisfies the partition unity:

(3.16)
$$\sum_{i \in \mathbb{I}} B(i - z) = 1; \quad \forall z \in \mathbb{R}.$$

The kernel width h was chosen so that $\frac{f_k}{h} \in \mathbb{I}, \forall k$ and $\frac{g_l}{h} \in \mathbb{I}, \forall l$. This choice of h coupled with (3.16) ensures that $\hat{P}_u(.)$, which is computed only from the fixed reference image voxel intensities, remains independent of θ.

3.4 Optimization Scheme

As discussed previously we have two SA methods, Step-SA and Samp-SA, that can estimate the unknown warp parameters. We explored two schemes to combine the advantages of these two SA methods; (i) an 'Hybrid-SA' scheme that started with Samp-SA for a fixed number of iterations and switched to Step-SA for later iterations and (ii) a 'Pyramid-SA' scheme that employed a variable combination of step and sample-sizes using a multi-resolution pyramid approach [70].

When the number of unknown warp parameters is very small, it may be sufficient to empirically identify a single step-size value for SA algorithms. However for large-dimensional vector valued parameters, the optimal step-size for each vector component may vary widely. To remedy this, we adopt an adaptive step-size estimation technique that has been shown to be convergent [36]. Let θ_k be the estimate of warp parameters at iteration k, with elements $\{\theta_k^i\}, i = 1, 2, \ldots p$. The adaptive step-size strategy assumes that for a stationary point θ_* of the similarity measure, rapid changes in the sign of $(\theta_k^i - \theta_*^i) - (\theta_{k-1}^i - \theta_*^i) = \theta_k^i - \theta_{k-1}^i$ indicate that θ_k^i is closer to its optima. Similarly, fewer sign changes are indicative of a greater distance from θ_*^i. Thus the step-size associated with the ith warp parameter component is kept inversely proportional to the number of sign changes of $\theta_k^i - \theta_{k-1}^i$. Our implementation estimates the step-size for the ith component θ_k^i as follows: $a_k^i = a_0/(A + Q_k^i)$, where Q_k^i is the number of sign changes in $\{\theta_m^i - \theta_{m-1}^i\}, m = 2, \ldots k$ and $Q_1^i = 0$. A and a_0 are positive non-zero constants.

Initial experiments comparing the different SA techniques used a pair of 2D 256×256 T1 and T2 MRI brain images obtained from ICBM, with pixel dimensions $1\text{mm} \times 1\text{mm}$. These slices were initially in registration. We applied different nonrigid warps T to the T2 weighted image, resulting in ground truth coordinates $T(x_i), i = 1 \ldots N_u$. This T2 image

was treated as the reference and the undeformed T1 slice was the homologous image. Warp estimates $T_{\hat{\theta}}$ were obtained by registering the homologous image onto the reference. These warp estimates were modeled using cubic B-spline basis functions given by (2.1) and were obtained by maximizing a plug-in estimate of the mutual information (MI), given by (3.12), between the two of images. The quality of the estimated warp $\{T_{\hat{\theta}}(x_i)\}_{i=1}^{N}$ was evaluated by computing the RMS error between the warp estimate and the ground truth,

$$(3.17) \qquad \text{RMS error} = \sqrt{\frac{1}{N_u}\sum_{i=1}^{N_u}\|T(x_i) - T_{\hat{\theta}}(x_i)\|^2}.$$

To compare Hybrid-SA with Step-SA, a known B-spline warp (ground truth) using 5×5 equally spaced control points, was applied to generate the reference T2 image. This ground truth warp represented zero model mismatch. We estimated the B-Spline warp that mapped the homologous image onto the reference, using (i) Step-SA with $a_0 = 1500$, $A = 15$ and fixed sample-size $= 5\%$ of the total number of pixels and (ii) Hybrid-SA using Samp-SA with $(K_0 = 2\%)$ and step-size $= 75$ for the first 159 (of 2000) iterations. For the remaining iterations, Step-SA used a fixed sample size set to the average sample size of the first 159 iterations. The step-size sequence parameter was $a_0 = 75 \times \min_i Q_{159}^i$ and $A = 1$.

The two SA methods were tested using both uniform sampling (US) and importance sampling (IS) with P_s^θ defined by (3.9). Thirty realizations of both SA methods with US and IS were obtained. Registration accuracy was quantified using the RMS error between the estimated warp and the ground truth, given in (3.17). Fig. 3.3 shows the mean performance of the SA techniques. Hybrid-SA reduces RMS error much faster than Step-SA. In this and subsequent figures error bars have been omitted to improve clarity; all $\pm$ one standard deviation error bars were within 0.25 pixels of the mean behavior plots.

To compare the performance of Hybrid-SA with US and IS against deterministic Gra-
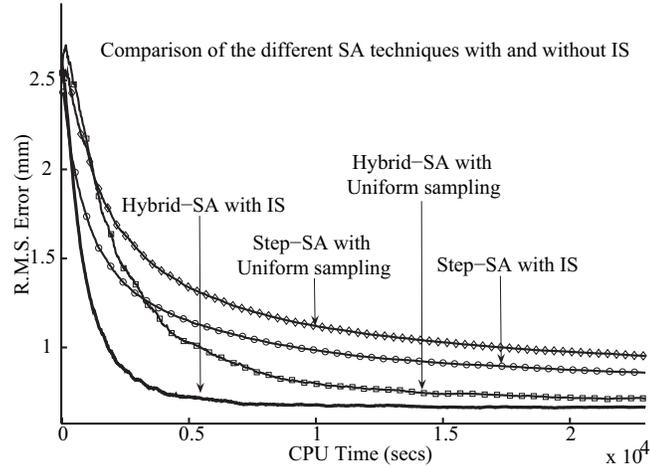
Figure 3.3: Comparison of Hybrid-SA and Step-SA.

dient Descent (GD), we applied a known warp using randomly placed Gaussian blobs to

the T2 image. This known warp had an inherent mismatch associated with the B-spline

warp model used to register the two images. For simplicity, registration was performed at

a single resolution, using 64 intensity levels to approximate the pdfs. Hybrid-SA optimiza-

tion with US and IS, used Samp-SA with $K_0 = 0.5\%$ for the first 159 of 2000 iterations.

The remaining iterations used Step-SA with $a_0 = 20 \times \min_i Q_{159}^i$ and $A = 1$. Determin-

istic gradient descent was found to perform best by using an adaptive step-size sequence,

similar to that described earlier, with $a_0 = 1500$ and $A = 15$. Thirty realizations were

obtained for each of the three optimization methods, with each realization of the determin-

istic GD method initialized with a small random warp guess. Mean values of the RMS

error obtained using the three optimizers shown in Fig. 3.4, indicate that Hybrid-SA with

IS outperforms Hybrid-SA with Uniform sampling and deterministic gradient descent. To

account for the effect of warp model mismatch, we computed a least squares fit between

the applied B-spline warp model and the known ground truth warp. The model mismatch

RMS error was 0.615 mm. Thus the effective registration error of Hybrid-SA with IS, after

accounting for model mismatch, was less than 0.5 mm.
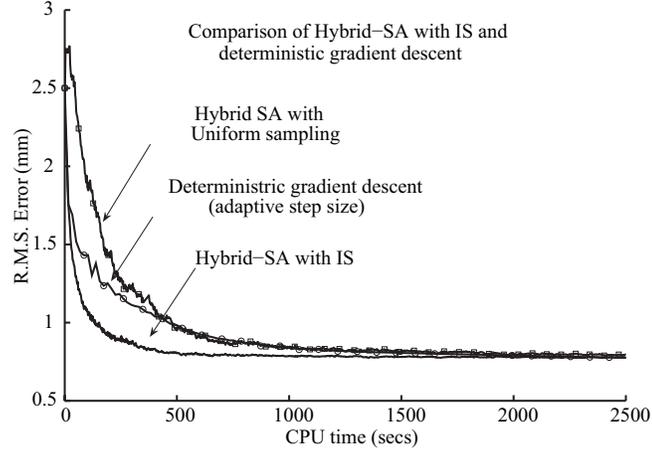
Figure 3.4: Comparison of Hybrid-SA with IS and deterministic gradient descent. The ground truth (applied) warp was based on randomly placed Gaussian blobs.

Lastly, to evaluate the efficacy of using IS-based SA in the widely used Pyramid registration scheme [70], we applied a B-spline warp with a grid of 11×11 control points to the T2 image and left the T1 image undistorted. The registration algorithm used a B-spline warp model. Our SA trials used a 3 level pyramid: The first level used 5×5 control points to model the deformation, 32 intensity levels to approximate the pdfs and both images were down-sampled by a factor of 4. Level 2 had 7×7 control points, 58 intensity levels and a down-sampling factor of 2. The last level used 9×9 control points, 64 intensity levels and no down-sampling. Levels 1 and 2 operated at 144 and 128 iterations of Samp-SA each. The initial sample-size K_0 was 1% of the total number of pixels at both levels and the step-sizes were fixed at 1 and 5 respectively. The last level used 256 iterations of Step-SA with $a_0 = 150$, $A = 1$ and sample-size $= 5\%$ of the total number of pixels at this level. The final warp estimate at a lower level was up-sampled and used to initialize the optimizer at the next level. As the highest level used only 9×9 control points to estimate the B-spline warp and the true (applied) warp was generated using a set of 11×11 control points, there was an inherent mismatch in the registration process. A least square fit of the applied B-spline model with 9×9 control points to the known ground truth warp revealed

a model mismatch RMS error of 0.553 mm.

Further, the same Pyramid structure and number of iterations were used for deterministic GD, which gave the best results by using an adaptive gain sequence with $a_0 = 10$ at level 1 and $a_0 = 100$ at levels 2 and 3. A was 1 at all levels of the pyramid. As before, thirty realizations were obtained for all three methods, with the deterministic optimization initialized by a random seed point for each realization. Results in Fig. 3.5 show that Pyramid-SA with IS performed well giving a large speed up in the rate of convergence. The effective registration error using Pyramid-SA with IS, after accounting for model mismatch, was less than 0.5 mm i.e. less than half a pixel.
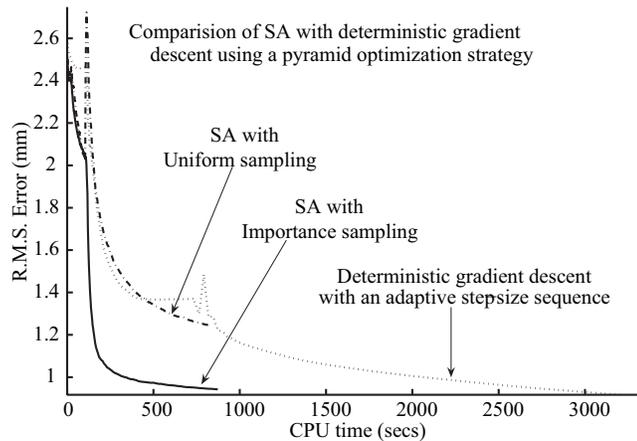
Figure 3.5: Improved convergence of Pyramid-SA with IS

These empirical results indicate that both Hybrid-SA and Pyramid-SA are viable candidates for faster nonrigid image registration using random sampling. However, given the recent prevalence of pyramid optimization schemes and their empirically demonstrated robustness to local minima [49, 70], we used Pyramid-SA for all subsequent experiments described here.

In our experiments all levels of Pyramid-SA used cubic B-spline representations of both images. Lower levels of the pyramid used coarse image approximations with small amounts of data to obtain initial warp estimates. These warp estimates were then refined

at higher levels of the pyramid using more precise image representations by including more intensity data. Since coarse image approximations are accompanied by a loss of detail, low level warp estimates capture gross global alignment and are explained using fewer parameters. As image detail increases with pyramid levels, the warps become more elaborate and depend on a larger number of parameters. Thus successive levels of the pyramid use an increasing number of intensity pairs to estimate the similarity metric. In an SA framework, this corresponds to implicitly increasing the sample-size between each level of the pyramid. 'Optimal' warp parameters within each pyramid level were estimated using Step-SA. For simplicity we call this optimization scheme 'Pyramid-SA'.

3.5 Implementation Issues

To use IS effectively for image registration, it is crucial to design a meaningful sampling distribution that requires minimal computational effort. The sampling distribution P_s^θ depends on the changing warp parameter estimates through $\{t_i^\theta\}_{i=1}^{N_u}$, so it has to be recomputed with significant variations in the SA estimates of θ. Thus it is important to use a fast and simple approximation of the edge maps. Since the reference image does not change throughout the registration, we pre-compute its (fixed) edge map $\{s_i\}_{i=1}^{N_u}$. However the homologous image geometry changes with updates in θ and corresponding edge magnitude values need to be recomputed. For large homologous images, edge maps based on higher order kernels such as the cubic spline in (3.5) can be computationally expensive. Hence we approximate edge magnitudes using fast lower order finite central differences of the intensity images along each image dimension.

As described in Sec. 3.4, the coarse-to-fine framework of the Pyramid-SA scheme inherently results in large scale changes in the warp estimate at lower levels of the pyramid, while finer warp adjustments occur at higher pyramid levels. At each iteration, large scale

(a) Example Sampling Distribution   (b) Importance Sampling   (c) Uniform Sampling
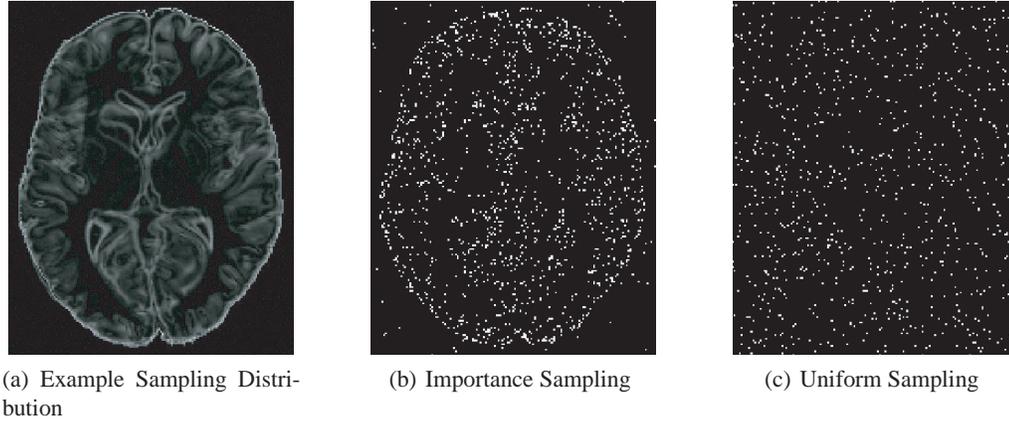
Figure 3.6: Comparison of samples obtained using the sampling distribution given by (3.9) versus samples obtained by Uniform sampling. Images were created when the algorithm was not near registration.

warp changes are more likely to significantly affect the edge map than finer refinements. Hence, we update the sampling distribution frequently at lower Pyramid-SA levels and increase the number of iterations m between updates as the optimizer switches to higher levels. SA algorithms are characterized by small steps along random search directions that point uphill (or downhill when minimizing a cost function) on average. Thus the sampling distribution P_s^θ is updated every m iterations to reflect the average change in these m warp estimates. At pyramid level $l = 1, 2, \ldots$ we used $m = 2^l$.

Lastly, at every update, the approximate homologous image edge map need be recomputed only at locations where the effective deformation is large enough to significantly change the edge magnitude. That is, we incrementally update our finite central difference based edge estimate only at geometric coordinates that move more than the dimensions of a voxel in any direction on average. These measures ensure that the overhead required to compute and update the sampling distribution is reasonably small. Further, this fractional overhead reduces steadily with increasing sample-sizes. Fig. 3.6 shows the sampling distribution and corresponding samples obtained using importance sampling for registration of simulated brain datasets.

In the following sections we used simulated and real data to study the performance of nonrigid registration using IS-based SA versus uniform sampling based SA. IS-based gradient approximations used a small subset of random samples drawn according to the sampling distribution designed in (3.9).

3.6 Results

We demonstrate the use of IS for image registration using both simulated and real data. Results include pair-wise monomodality and multimodality registration using two common intensity-based similarity metrics. All registration results using IS-based Pyramid-SA (IS-SA) and US-based Pyramid-SA (US-SA) described here employed the pyramid optimization framework detailed in Sec. 3.4. For comparison, registration was also performed using deterministic Gradient Descent (GD) in the same multi-resolution pyramid framework. GD used all image voxels to compute the analytical gradient at each iteration. All three methods utilized multi-resolution representations of both images using cubic splines and estimated deformable warps based on B-splines.

3.6.1 Behavior of IS-SA with Variations in Step-size

A limitation of SA approaches is their sensitivity to tuning parameters such as step-sizes. If the sampling distribution P_s^θ designed in (3.9) reduces the variance of $\hat{g}(\theta)$, IS-SA can be expected to have an increased tolerance to variations in step-sizes. Simulated datasets were used to compare the behavior of multi-modal registration using IS-SA and US-SA with various step-sizes.

Mutual Information (MI) based registration was performed between $180 \times 260 \times 60$ T1 and PD-weighted simulated MR volumes with $1 \times 1 \times 3$ mm^3 voxels, obtained from ICBM [9]. A plug-in estimate of MI between the two volumes, given by (3.12) , was used as the similarity metric. Analysis outlined in Sec. 3.3.1 shows that image edges strongly

influence the gradient of this MI estimate.

All results using IS-SA optimization schemes in this section used the sampling distribution given by (3.9). A known synthetic warp $T(.)$ derived using radial blobs of varying severity was applied to the T1 volume, yielding ground truth coordinates $T(x_i), i = 1, \ldots, N_u$. This warped volume was treated as the reference, while the unchanged PD volume was the homologous image. B-spline warps $T_{\hat{\theta}}(.)$ were estimated by mapping the homologous volume onto the reference volume. Quality of the estimated warp $\{T_{\hat{\theta}}(x_i)\}_{i=1}^{N_u}$ was evaluated using the RMS error between the warp estimate and ground-truth, as in (3.17).

A two level Pyramid-SA scheme was used to register the two data sets. Level one used 64 histogram bins, a B-spline control point spacing of $16 \times 16 \times 8$ voxels and both images were down-sampled by a factor of two in all dimensions. The second level had 128 histogram bins, an $8 \times 8 \times 4$ voxels B-spline control point spacing and no down-sampling. Both levels implemented 150 and 250 iterations of Step-SA respectively and used only a fixed percentage of all available voxel pairs at that level. The step-size a_k^i, corresponding to component θ_k^i of the warp parameters' estimate at iteration k, was $a_k^i = a_0/(10 + Q_k^i), i = 1, 2, \ldots, p$. Where, Q_k^i was the number of sign changes in $\{\theta_m^i - \theta_{m-1}^i\}, m = 2, \ldots, k$. Thus the only tuning parameter in the step-size sequence was a_0.

To study the effect of varying step-size parameter a_0, warp estimates from 10 registration runs were obtained using IS and US, for six systematically increasing values of a_0 from 1000 up to 16000 in increments of 3000. Independent realizations of Gaussian noise $N(0, 9)$ were added to both images prior to the registration runs. This process was repeated for four different sample sizes of $0.25, 0.5, 1$ and 2 percent respectively. Fig. 3.7 compares statistics of the final RMS errors obtained using the two sampling strategies for a fixed CPU time. As hypothesized, IS-SA yields lower errors than US-SA over the entire

range of step-sizes.

Empirically, IS-SA was significantly less sensitive to step-size variations, while consistently giving more accurate warp estimates. Further, US-SA required larger sample sizes to achieve accuracies comparable to those using IS. As sample-sizes increase both IS and US will capture similar levels of image complexity making their performance comparable. The minimum sample-size beyond which both sampling methods give similar results will depend on the complexity of the datasets. In general, US will be effective at smaller sample-sizes when image edge features are roughly uniformly dispersed.

3.6.2 Application to Real Data

Encouraged by the observations made in the previous section, we used IS to register real datasets. Intensity-based registration using B-spline warps was used to align CT inhale and exhale lung datasets from 8 subjects. These CT scan pairs were obtained using a helical CT scanner (CT/I, General Electric, Milwaukee, WI) with $0.187 \times 0.187 \times 0.5 \, \text{cm}^3$ voxels. Each scan pair was acquired during coached voluntary breath-hold periods of 18 to 35 secs; the first scan at normal exhale followed by one at normal inhale. A more detailed description of the data can be found in [11].

Monomodality registration was performed using the negative of Sum of Squared Differences (SSD) as a similarity metric. In this case, both the reference and homologous images are assumed to be noisy realizations drawn from the same continuous function. Let the reference image be given by a set of noisy samples $\{\tilde{u}_i\}_{i=1}^{N_u}$. Then the negative SSD similarity metric is

$$(3.18) \qquad \Psi_{\text{SSD}}(\theta) = -\frac{1}{N_u} \sum_{i=1}^{N_u} \left(\tilde{u}_i - \hat{v}_i^\theta \right)^2,$$

where the interpolated homologous image $\{\hat{v}_i^\theta\}_{i=1}^{N_u}$ is given by (3.5). Differentiating the above expression shows that image edges are important to the gradient of Ψ_{SSD}.
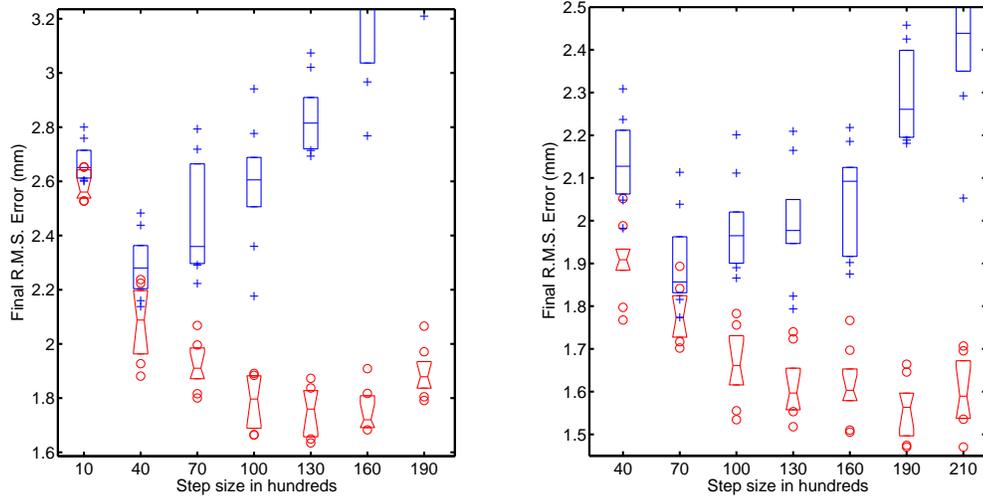
(a) Each pyramid level used 0.25% of all available voxels.

(b) Each pyramid level used 0.5% of all available voxels.
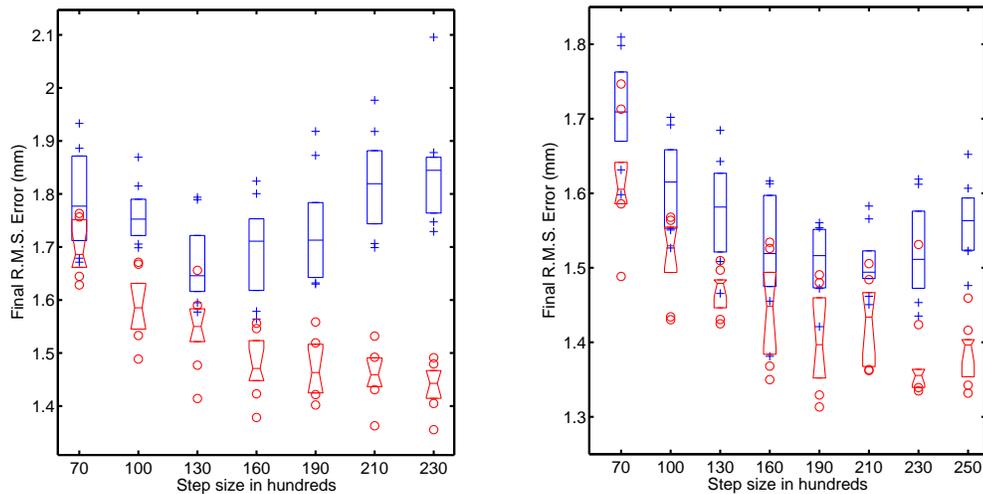
(c) Each pyramid level used 1% of all available voxels.

(d) Each pyramid level used 2% of all available voxels.

Figure 3.7: Comparison of the performance of IS-SA (red/notched) versus US-SA (blue/plain) with variations in step-sizes. Figures show RMS error statistics for 10 nonrigid multimodality registration runs at six step-sizes and four $(0.25, 0.5, 1$ and $2\%)$ sample-sizes. The line at the center of each boxplot shows the median RMS error value and top and bottom edges are the 75 and 25 percent quantile RMS errors. 'Outliers' are shown by (o) for IS and by (+) for US. IS does significantly better than US at all four sample-sizes. Specifically, IS results in lower variance values and shows better tolerance to variations in step-sizes. Trends in the four plots indicate that the performance of both sampling strategies will become comparable with an increase in sample-size.

Training

Effective use of US-SA or IS-SA to register a population of real datasets requires an efficient strategy to estimate the step-size parameter a_0. Here we outline a simple procedure to estimate this a_0 value using a single randomly chosen dataset from the target CT population. In the absence of known ground truth, B-spline warp estimates obtained using deterministic GD optimization were treated as the pseudo ground-truth. This is a reasonable assumption since the goal of our SA algorithms is to use only a small subset of strategically selected image voxels to attain registration accuracy comparable to that using GD with all image voxels. To mitigate local minima, registration estimates from multiple runs of a GD algorithm were used. Each run was initialized using a small randomly generated warp. The final registration estimate corresponding to the largest similarity metric value was treated as the best attainable warp. For a given sample-size, optimal a_0 values using both IS-SA and US-SA were chosen to consistently find warp estimates that yielded the smallest RMS error values with respect to this pseudo ground-truth warp.

For training purposes, we employed a two-level pyramid registration scheme. Level 1 downsampled the images by a factor of 2, estimated B-spline warps with a $16 \times 16 \times 8$ voxels control point spacing and used a_0 as the step-size parameter. The second level used no downsampling, a $8 \times 8 \times 4$ B-spline control point spacing and the step-size parameter was $1.5 \times a_0$. Each level used 1% of the total available voxels at that level. Ten registration wrap estimates were obtained using both IS-SA and US-SA for a set of five different a_0 values. Each registration run was terminated after 10 mins and at every iteration we recorded RMS errors of the estimated B-spline warp with respect to the pseudo ground-truth warp. Step-size parameter value $a_0 = 1$ was found to yield the best results for both SA methods. Fig. 3.8(a) shows statistics of RMS error values for all 10 IS-SA and US-SA registration runs at all five a_0 values. Fig. 3.8(b) shows speed and accuracy comparisons

of GD, IS-SA and US-SA (both using $a_0 = 1$) with respect to the pseudo ground-truth warp. All subsequent SA based registrations were performed using this trained pyramid scheme with $a_0 = 1$.

Validation

To gauge the performance of IS-SA and US-SA based on the trained pyramid scheme described above, we applied both methods to register all 8 CT inhale-exhale lung scan pairs. To quantify registration accuracy, six expert identified feature points were used per scan pair. These features included both bronchial and vascular bifurcations. For each subject, registration was performed by treating the exhale scan as the reference and the inhale scan as the homologous dataset. Following registration, the estimated B-spline warp was used to transform the six exhale feature point coordinates to obtain predicted inhale feature point coordinates. The average of the Euclidean distance between the coordinates of each predicted and expert identified inhale feature point was used as an error metric to quantify registration accuracy for each dataset.

Since in reality we wish to replace a single GD registration run by a single SA registration run it is important that the method of choice give consistently good warp estimates with as little variance as possible. To empirically demonstrate the estimate variance associated with both SA methods, each CT dataset registration was repeated ten times. For comparison each dataset was also registered using GD. Each of the ten GD repetitions was initialized with a small random independently generated warp. Each SA registration run was completed in approximately 5 to 8 mins on a moderate PC running C++ code; in contrast, each successful GD registration required about 30 to 90 mins. Fig. 3.9 summarizes statistics of the resulting feature point error metric for all ten registration warp estimates using IS-SA and US-SA for all 8 datasets. In general IS-SA resulted in better accuracy
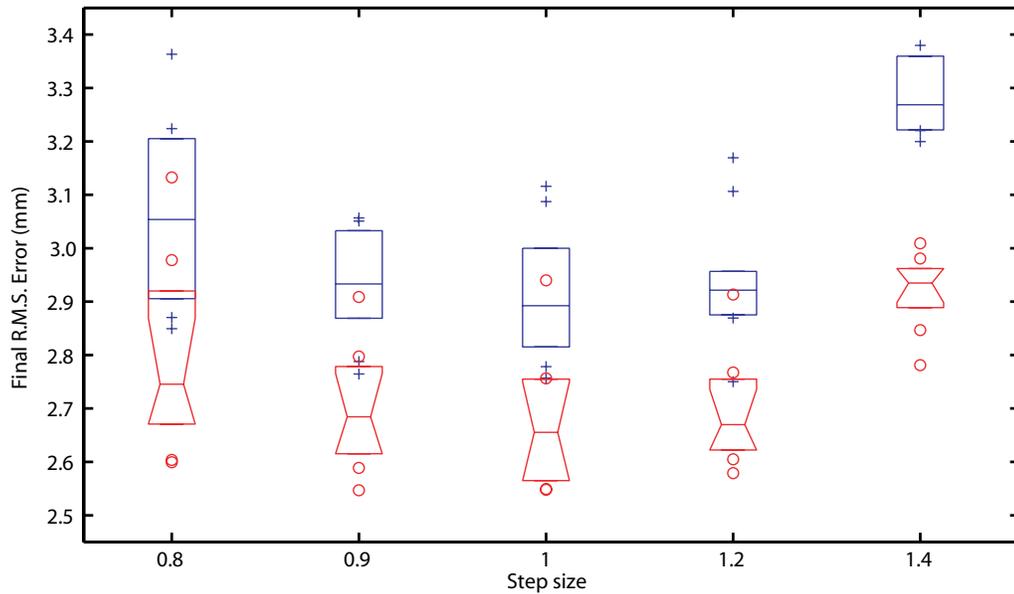
(a) Step-size parameter (a_0) estimation.
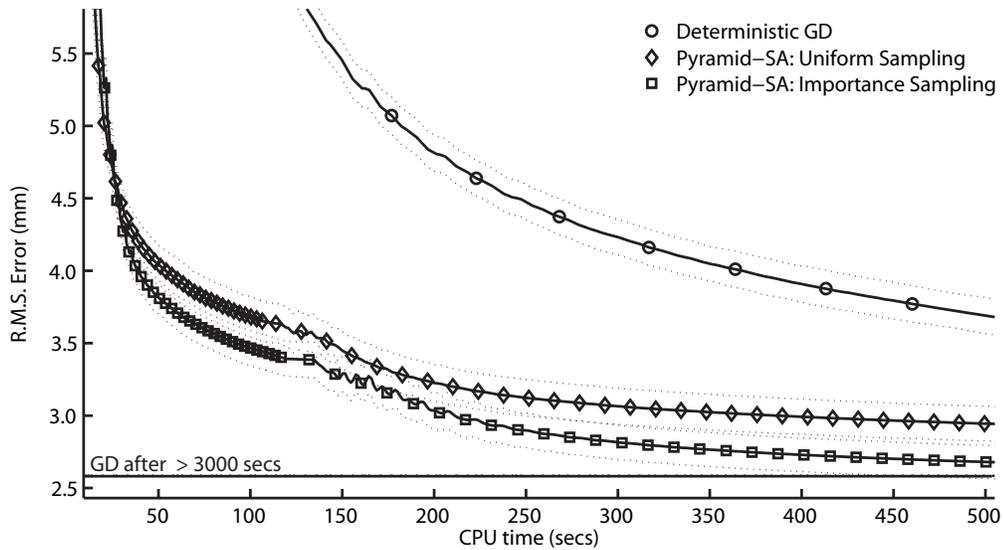
(b) Speed and accuracy comparison of IS-SA, US-SA and GD.

Figure 3.8: Comparison of the speed and accuracy of IS-SA (red/notched) and US-SA (blue/plain) for registration of CT Lung data. The optimal step-size parameter a_0 was empirically chosen to consistently produce warp estimates closest to the pseudo ground-truth warp in an RMSE sense. Fig. 3.8(a) shows that $a_0 = 1$ was the best value for both methods. The line at the center of each box-plot is the median RMS error, while top and bottom edges are 75 and 25 percent quantiles. Outliers are represented by ($\circ$) for IS-SA and ($+$) for US-SA. Fig. 3.8(b) shows how the speed and accuracy of the best IS-SA and US-SA schemes ($a_0 = 1$ and sample-size $= 1\%$) compare with those using GD (sample-size $= 100\%$) on average. Dotted lines are ± 1 standard deviation plots.

than US-SA and showed a reduction in estimate variance.
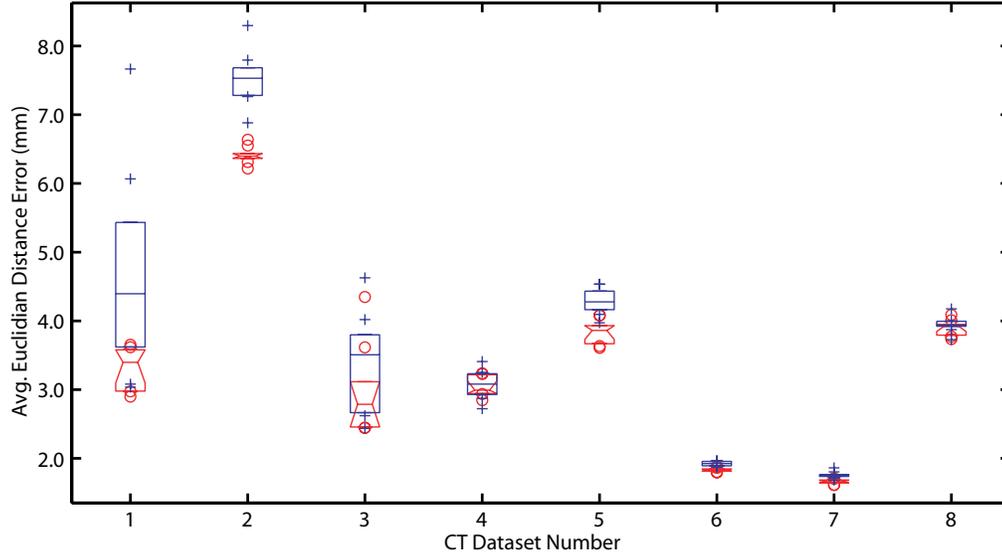
Figure 3.9: Comparison of the accuracy and variation in trained IS-SA (red/notched) versus US-SA (blue/plain) registration using expert identified feature points for CT inhale-exhale lung data. The line at the center of each box-plot is the median error metric, while top and bottom edges are 25 and 75 percent quantiles. Outliers are represented by (○) for IS-SA and (+) for US-SA. Dataset 5 was used in the training step.

The average Euclidian distance between the expert identified exhale and inhale feature points can be used as some measure of the severity of the initial deformation. Table 3.1 indicates that for datasets with larger deformations (datasets 1, 2 and 3) IS-SA showed a marked improvement in accuracy over US-SA. For datasets with smaller deformations (datasets 6, 7 and 8) both methods performed comparably with IS-SA doing only slightly better than US-SA. The datasets are presented in order of decreasing initial deformation for ease of comparison. For most datasets IS-SA showed accuracy comparable to that using GD. Empirically, for datasets with larger deformations, SA methods appeared to be less susceptible to local minima than GD. For datasets 1, 2 and 3 most repeated GD registration trials got stuck in local minima and terminated after 5 to 7 mins. These GD registrations resulted in poor inhale feature point predictions and were discarded as unsuccessful. In particular no GD registration run was successful for datasets 2 and 3, while only one run

managed to escape local minima for dataset 1.

Results in Table 3.1 indicate that on average GD registration was more accurate than SA registration for datasets 1, 4, 5 and 8. The accuracy of the SA-based registration schemes could have been improved by increasing the sample-size (1%) used in the training step to obtain registration warp estimates that fit the pseudo ground-truth warp more closely.

Avg. Error (mm)	CT Dataset Number							
	1	2	3	4	5	6	7	8
Initial	15.10	14.52	13.31	11.73	9.13	8.62	7.77	6.89
Final								
US-SA	4.64	7.52	3.40	3.06	4.29	1.92	1.76	3.95
IS-SA	3.31	6.41	2.97	3.05	3.84	1.83	1.66	3.89
GD	3.14	-	-	2.15	3.29	1.95	2.12	3.63

Table 3.1: Comparsion of the average Euclidian distance error for inhale feature points predicted using US-SA, IS-SA and GD.

3.7 Discussion and Conclusion

We have developed and validated an importance sampling based stochastic approximation (IS-SA) approach to accelerate nonrigid image registration. We leveraged the significant influence of image edges on gradients of intensity-based similarity metrics to design an adaptive non-uniform sampling distribution that encourages sampling from these regions. Results for both synthetic simulations and real CT lung data show that registration using IS-SA can yield better speed and accuracy than SA schemes that use uniform sampling (i.e., US-SA). In particular, Fig. 3.7 shows that the number of samples required to attain a particular registration accuracy was halved by using IS-SA. For a fixed sample-size in Fig. 3.8(b) IS-SA was more than 2 times faster than US-SA on average.

The use of SA methods in practical applications can be hindered by their dependence on the step-size parameter. To effectively apply these methods to populations of real data, we

introduced a training strategy to empirically estimate a reasonable value for this step-size parameter in the absence of ground-truth. The training method uses only a single randomly chosen dataset from the target population and its corresponding 'successful' deterministic GD registration warp estimate. This approach should be practical when several scans from the same protocol need to be registered. Finding automatic parameter selection methods for a single image pair is a challenging open problem.

Though we have demonstrated the efficacy of IS-SA only with B-spline warps, our framework is mostly independent of the warp model. Specifically for more global warps (such as Thin-plate Splines) where each warp parameter depends on a larger number of image voxels, we expect to see more marked improvements in registration performance using IS-SA.

The data used here to demonstrate improvements in registration using IS-SA had few or sparse edges. In a random subsample drawn using a uniform distribution, the fraction of voxels that lie on an image edge will on average equal the fraction of total image voxels that belong to edges. Thus as the percentage of edges increases the benefit of using an edge-based importance sampling distribution will be diluted. Fig. 3.10 shows the sampling distribution for a dataset with a large number of edges, for which both IS-SA and US-SA gave comparable registration speeds. In such cases it may be beneficial for IS-SA to use a more stringent criterion to retain fewer edges in the sampling distribution, however, more empirical experiments will be needed to quantify the approximate percentage of edges that need to be retained. Further, an edge-based sampling strategy may not be the best choice for registration when one image has significant strongly demarcated structures absent from the other image(s). Finally, we note that low discrepancy sequences were used in [69] to improve the performance of uniform sampling based registration by utilizing Highly Uniform Point-sets (HUPS). A similar strategy, i.e. transforming such HUPS to obtain

samples that follow the target sampling distribution in (3.9), may further augment the performance of importance sampling based registration.
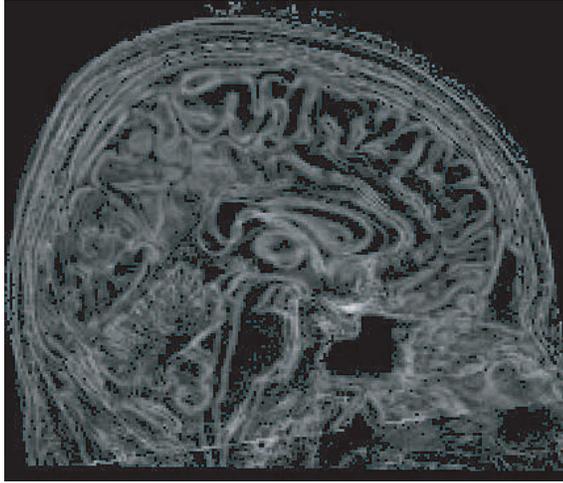
Figure 3.10: Sampling distribution for a high resolution brain volume with dense edges.

CHAPTER 4

Functional Magnetic Resonance Imaging

This chapter briefly reviews the principles, acquision protocols and challenges of functional magnetic resonance imaging (fMRI), an increasingly popular modality used to non-invasively study brain function. The prevalence of hydrogen nuclei in the human body is exploited by Magnetic Resonance Imaging (MRI) to non-invasively obtain images of the spatial distribution of different tissues. MRI is based on the nuclear magnetic resonance properties of nuclei having an odd number of neutrons and/or protons (such as Hydrogen ^1H), which have an associated nuclear spin and magnetic moment. Each tissue type has a characteristic concentration of hydrogen nuclei and a corresponding bulk magnetic moment per unit volume, called 'magnetization'; in the absence of an external magnetic field this net magnetization is zero. However, when placed in the strong static magnetic field B_0 of an MRI scanner, the individual magnetic moments align themselves either parallel (low energy state) to the external magnetic field or anti-parallel (high energy state) to it. The number of magnetic moments in the low energy state is always slightly greater than those in the high energy state, resulting in a net magnetization M_0 in the direction of B_0. The magnitude of M_0 is a function of the proton density of the tissue type ρ_0, the magnitude of B_0 and the absolute temperature T i.e. $M_0 \propto \rho_0 \frac{B_0}{T}$.

To obtain contrast images of various tissues in the body, M_0 is tipped away from its

equilibrium state using an RF pulse of field strength B_1 applied perpendicular to the direction of B_0. The angle by which M_0 is tipped away from B_0 depends on the strength of B_1 and is called the flip angle α. In the absence of extraneous fields, such a tipped magnetization will precess about the static field at the Larmor frequency given by $\omega_0 = \gamma B_0$, where the gyromagnetic constant $\gamma \approx 2.68 \times 10^8$ rad/s/Tesla for Hydrogen. This precessing magnetization has a longitudinal component M_z along the direction of B_0 and a transverse component M_{xy} in the plane perpendicular to B_0.

Owing to thermal energy interactions of ^1H protons with the lattice of neighboring atoms, the longitudinal magnetization M_z exponentially grows back to its equilibrium state M_0 over time. This regrowth is characterized by a 'spin-lattice relaxation' time constant T1. In addition to B_0, spins experience variations in local fields due to the magnetic fields of their neighbors. Thus the local precession frequencies vary, causing the individual spins to fan out. This fanning out is commonly called 'dephasing' and over time results in an exponential decay in the transverse magnetization M_{xy}. This exponential decay is characterized by the 'spin-spin relaxation' time constant T2. Typical T1 and T2 values for some tissues [30] are given in table 4.1. MRI uses differences in T1 and T2 values for the

Tissue	T1 ms	T2 ms
gray matter (GM)	950	100
white matter (WM)	600	80
cerebrospinal fluid (CSF)	4500	2200
muscle	900	50
fat	250	60
blood	1200	100-200

Table 4.1: Typical T1 and T2 relaxation time constants, reproduced from [30].

various tissue types to create tissue contrast images, by employing various timing and RF excitation strategies.

An additional dephasing of the transverse magnetization may occur due to external magnetic inhomogeneities. This reduction in the initial M_{xy} magnitude is characterized

by a different time constant T2*, such that $\frac{1}{\text{T2}^*} = \frac{1}{\text{T2}} + \frac{1}{\text{T2}'}$; where T2$'$ accounts for the effect of external field inhomogeneities.

While MRI can be used effectively to create T1 or T2 weighted images of spatial distribution of different tissue in the body, the very abundance of Hydrogen in water that make this possible precludes the use of MRI to examine the subtle effects of brain function. To study brain activity, MRI should be able to detect direct effects of neural activity or indirect variations in metabolic activity due to brain function. Functional MRI (fMRI) satisfies the later condition, in that it can identify changes in blood oxygenation levels that depend on the energy requirements of active brain cells, i.e., the BOLD effect.

4.1 BOLD Effect

Neuronal membrane potentials required for signaling and responding to various stimuli need energy for their maintenance and restoration. Since there are few energy storage sites in the brain, this energy has to be supplied in the form of glucose and oxygen by blood flow to the brain. Oxygen binds to hemoglobin molecules in the blood stream, which in turn swap it for carbon dioxide in capillaries. fMRI studies try to discern brain activity by capturing changes in blood oxygenation levels in the brain correlated to external stimuli.

In 1938 Linus Pauling and Charles Coryell discovered that oxygenated hemoglobin (Hb) has no unpaired electrons and hence no magnetization, making it diamagnetic; however deoxygenated hemoglobin (dHb) is paramagnetic (i.e., has unpaired electrons and a considerable net magnetic moment) [35]. This Blood Oxygenation Level Dependant (BOLD) change in the magnetic properties of hemoglobin was shown to translate into MR signal changes by Thulborn et al. [71].

The paramagnetic nature of dHb implies that there is high extraneous magnetic susceptibility in the vicinity of deoxygenated blood; causing a larger dephasing of the transverse

magnetization M_{xy} due to T2* decay leading to MR signal loss. However when blood is oxygenated the same anatomical location will have a larger MR signal since Hb is diamagnetic and results in a reduction in external field inhomogeneity. Thulborn et al. demonstrated that this BOLD effect in the MR signal increases with the square of the strength of the static magnetic field B_0.

Blood consists of two main components viz. plasma and red blood cells (RBCs). The fraction of whole blood volume that is taken up by RBCs is called the hematocrit (Hct). The magnetic susceptibility of the entire blood system χ_{blood} for a fractional oxygenation of RBCs Y, can be expressed as [30];

$$\chi_{\text{blood}} = \text{Hct}(Y\chi_{\text{oxy}} + (1 - Y)\chi_{\text{deoxy}}) + (1 - \text{Hct})\chi_{\text{plasma}},$$

where χ_{oxy}, χ_{deoxy} and χ_{plasma} are the magnetic susceptibilities of oxygenated RBCs, deoxygenated RBCs and plasma respectively. Thus a change ΔY in the oxygenation level will affect the susceptibility of blood by

$$\Delta\chi_{\text{blood}} = -\Delta Y(\chi_{\text{deoxy}} - \chi_{\text{oxy}})\text{Hct};$$

where $\chi_{\text{deoxy}} - \chi_{\text{oxy}} = 4\pi \times 0.18$ ppm per unit Hct, assuming that oxygenation of plasma does not affect its susceptibility. Finally, for a given Y, the change ΔY can be expressed in terms of the relative change in blood flow δ and in the metabolic rate $\beta - 1$ as,

$$\Delta Y = \frac{1 + \delta - \beta}{1 + \delta}(1 - Y).$$

If the change in metabolic activity is negligible, $\beta \approx 1$ and $\Delta Y = \frac{\delta}{1+\delta}(1 - Y)$. However in some cases β may be large enough to mask the effect of change in blood flow δ.

fMRI is based on fluctuations in χ_{blood}. Changes in the oxygenation level ΔY due to brain activity are manifested as variations the in T2* of blood, resulting in differences in MR signal. For instance, in a finger tapping experiment fMRI signal would increase due to

an increase in blood flow to certain regions of the brain. Fortunately the metabolic activity in these areas does not change significantly, i.e., $\beta \approx 1$ ensuring that ΔY is dominated by δ. As a result the blood has higher oxygen content leading to a decrease in blood susceptibility and increase in MR signal.

4.2 Echo Planar Imaging

Commonly used gradient-echo pulse sequences can in theory be used to record strong BOLD signal in brain regions such as the visual or motor cortex. However each gradient-echo pulse sequence sequentially samples k-space line by line, increasing acquisition time. This makes these methods too slow to capture smaller BOLD contrasts that are characteristic of subtle behavioral or though related processes.

Echo Planar Imaging (EPI) is a fast image acquisition protocol that traverses all k-space, within a 2D slice, after a single excitation pulse. This reduction in acquisition time allows EPI to achieve the higher temporal resolution required to capture fleeting BOLD effects while allowing sufficient time for the BOLD contrast to develop. However the increase in temporal resolution is offset by a reduction in spatial resolution. EPI pulse sequences are based on the same echo-forming mechanism used in spin-echo or gradient-echo for ordinary MRI. Fig. 4.1 shows a generic EPI pulse sequence and the resulting k-space traversal for a thin 2D slice, adapted from bitc.bme.emory.edu. For this pulse sequence, the 2D slice is excited using slice-selection gradient G_{ss} and the entire k-space is traversed in a zig-zag manner. Each $\pm$ gradient cycle in the frequency encode direction G_{fe} acquires one line along K_x. Every blip in the phase encode direction G_{pe} advances the k-space location in K_y to the beginning of the next line to be acquired in K_x. Though each gradient cycle in the frequency encode direction G_{fe} has it own echo, the overall echo time TE is adjusted to coincide with the primary and strongest echo at the center of k-space with

$K_x = K_y = 0$. This is because the net reduction in the MR signal for the entire image is dependant on the signal loss incurred at the center of k-space. Having acquired one slice, the slice-select gradient G_{ss} is adjusted and the entire pulse sequence in repeated to excite and acquire the next slice.

(a) Generic EPI pulse sequence.   (b) EPI K-space traversal.
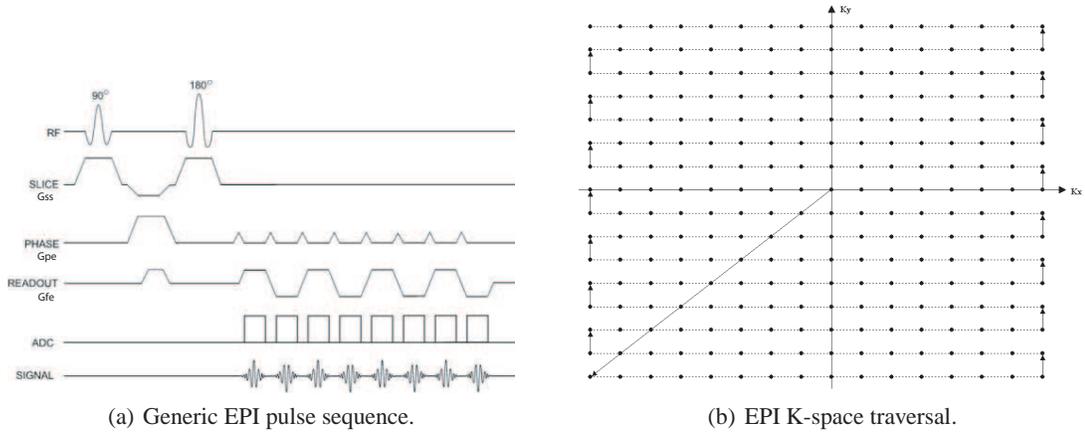
Figure 4.1: A generic Echo Planar Imaging pulse sequence and the corresponding k-space traversal, reproduced from bitc.bme.emory.edu.

Gradient-echo recalled EPI is very sensitive to local variations in T2*. As image contrast due to the BOLD effect stems from small field inhomogeneities, this EPI pulse sequence is commonly used to acquired fMRI data. Further, image quality can be improved by using smaller flip angles to reduce scan times. While spin-echo recalled EPI pulse sequences are more susceptible to inhomogeneities due to blood capillaries, they display reduced overall sensitivity to field inhomogeneities. An improvement in the sensitivity of these pulse sequences to the BOLD effect can be achieved by using an asymmetric spin-echo [18].

The main objective of EPI is to cover k-space quickly. It has been shown that a spiral k-space trajectory results in fast scanning techniques that produce comparatively higher BOLD signal to noise ratios (SNR). However the k-space data are acquired at non-uniform intervals and possibly non-uniform distributions. Hence spiral k-space data have to be re-gridded for use with conventional FFT reconstruction methods. Alternatively, nonuniform

FFT-based gridding can be used to reconstruct the images [20, 45].

4.3   Artifacts in fMRI

fMRI experiments measure changes in the magnetization of brain tissue via current induced in a detector coil in the MR scanner. Thus the recorded quantity is a mixture of relevant MR signal and noise introduced due to, among others, thermal interactions in the subject. Images reconstructed from this noisy k-space data will inevitably have an associated noise component. Raw signal-to-noise-ratio (SNR) is the ratio of the mean intensity of the image over a region containing MR signal (inside the brain) over the noise standard deviation, computed over a region without any MR signal (outside the brain). For MR datasets a more cogent measure is the contrast-to-noise-ration (CNR). Contrast in an MR image refers to the physical properties to which it is sensitive (e.g., T1, T2) [35]. The CNR is a measure of how the difference in the intensities of various tissues compares to the noise in their measurements. However, the most important measure for fMRI is the functional signal-to-noise ration (fSNR) which is the ratio of the magnitude of the change in signal intensity at two distinct states of a brain region (active versus inactive) and the associated noise.

At a given field strength the amount of magnetization determines the level of MR signal, whereas the noise associated variation in fMRI data is both spatial and temporal in character. Any spatial variability in fMRI data not corresponding to image contrast due to intrinsic tissue properties (such as T1, T2 etc.) and any temporal fluctuations along the time series not correlated to the stimulus of interest are treated as noise.

The most common and prevalent source of noise in MRI is the thermal fluctuations of electrolytes in the subject or body being scanned [55]. This thermal noise increases with body temperature. Collisions between free electrons and atoms in the electrical compo-

nents of the MR receiver are another (smaller) source of thermal noise. Receiver thermal noise increases with system temperature. Apart from body (or system) temperature, thermal noise is linearly proportional to the strength of the static magnetic field B_0. As discussed in [5], in MR magnitude images, thermal noise displays a Gaussian distribution inside the brain and a Rayleigh distribution outside it.

Imperfections in B_0 are another source of artifacts that can lead to geometric distortions and signal variation in fMRI data. Static field inhomogeneities, i.e., deviations in the actual strength of B_0 from its desired theoretical value at different spatial coordinates, result in unwanted variations in voxel spin frequencies at those spatial locations. Large changes in spin frequencies in k-space result in inaccurate spatial displacements of the voxel, while smaller B_0 inhomogeneities are embodied as loss of signal coherence due to extraneous T2* effects. Further fMRI artifacts are introduced by nonlinearities in the gradient fields. As x and y-gradients are used to control k-space trajectories, irregularities in these gradients skew the path traveled through k-space, introducing shear in the fMRI data. Similarly, discrepancies in the slice select or z-gradient will degrade slice thickness accuracy and signal strength.

Thermal noise and image distortions due to field inhomogeneities are evident in both animate and inanimate subjects. However, the BOLD effect studied by fMRI data, is a combination of the brain's response to a variety of intrinsic and extrinsic stimuli, voluntary and involuntary muscle activity and other metabolic activity driven chemical reactions. These factors result in fMRI signal being plagued by physiological noise and motion related artifacts.

Motion is a prominent source of noise in active fMRI subjects. During the course of fMRI experiments subjects may voluntarily move due to fatigue or may speak, swallow etc. Involuntary movements due to periodic activity such as breathing or the cardiac cy-

cle also cause motion related artifacts. As the subject's head moves with respect to the scanner the MR signal at a fixed voxel coordinate will vary. This effect is most evident at head boundaries, where a voxel that corresponded predominantly to CSF in one scan may depict grey matter in another scan along the time series. This variability at a given coordinate location along the time series confounds statistical analyses of fMRI data used to study brain function. In many instances head movement is small and can be corrected retrospectively using registration algorithms. However in some cases head motion can be significant enough to render the fMRI data unusable.

Further, head motion may cause some magnetic spins to be re-excited at irregular time intervals. Voxels excited prematurely do not get sufficient time to relax to their equilibrium states resulting in signal loss, while others are allowed a longer time to regain their longitudinal magnetization resulting in a stimulus-independent increase in intensity. This effect results in spin-history or spin-saturation artifacts of the same order of magnitude as the BOLD effect. These spin saruration artifacts reduce the fSNR and may result in an incorrect activation analysis. We address this effect in Chapter 5.

Lastly, subject responses to various impertinent stimuli, such as scanner noise or unrelated memory and thought related stimuli, result in incidental neural activity. Other sources of variability include changes in the attention span and response time of the subject over the duration of the scan.

4.4 Statistical Analysis

FMRI experiments are designed carefully to increase the fSNR. Typically, experiments use an epoch based scheme. Each epoch consists of an 'on' time period, when the external stimulus is presented; followed by an 'off' period during which the stimulus is absent. The duration of the on period is designed to be sufficient to allow the BOLD contrast to

develop while that of the off period is long enough to allow it to subside. This on-off cycle is repeated a fixed number of times to improve the fSNR. To reduce the total scan time during which the subject is assumed to remain still and attentive, the number of repetitions is kept at a minimum. This series of repeated on and off scans constitutes an fMRI time series.

Having processed the time series to remove some motion and field inhomogeneity related noise, it remains to be determined whether the observed differences in voxel intensity along the time series are statistically significant and can be attributed to stimulus related BOLD contrast. Statistical analysis of fMRI data is aimed at discriminating between the research hypothesis and the null hypothesis at each voxel location. The null hypothesis assumes that the external stimulus being examined has no effect on the voxel intensity while the research hypothesis postulates that intensity variations are correlated to the presence or absence of external stimuli. Statistical tests are designed to evaluate the probability, called a p-value, that intensity differences at each image location occur by pure chance, i.e., can be explained under the null hypothesis. Only voxels with p-values below a user defined alpha-threshold are marked as significant or active.

The probability of a false positive or labeling a voxel active when in reality it does not respond to the stimulus, is given by the alpha-threshold. Similarly accepting the null hypothesis at voxels that are actually active results in false negatives. Clearly statistical tests that minimize the number of false positives while increasing the probability of a true positive (i.e., 1-false negative) are desirable.

The simplest statistical tests for epoch based fMRI paradigms examine the difference between the means of voxel intensities at the on and off states of the experiments. The Student's t statistic, given by (4.1), is commonly used to alleviate the effect of intrinsic

stimulus independent intensity variation along the time series;

$$(4.1) \qquad t = \frac{\bar{X}_{\mathrm{on}} - \bar{X}_{\mathrm{off}}}{\hat{\sigma}_{\bar{X}_{\mathrm{on}} - \bar{X}_{\mathrm{off}}}}, \quad \hat{\sigma}_{\bar{X}_{\mathrm{on}} - \bar{X}_{\mathrm{off}}} = \sqrt{\frac{(n_1 - 1)\sigma^2_{\bar{X}_{\mathrm{on}}} - (n_2 - 1)\sigma^2_{\bar{X}_{\mathrm{off}}}}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

where, n_1 and n_2 are the number of on and off time series volumes with sample mean and sample variance $\bar{X}_{\mathrm{on}}, \hat{\sigma}_{\bar{X}_{\mathrm{on}}}$ and $\bar{X}_{\mathrm{off}}, \hat{\sigma}_{\bar{X}_{\mathrm{off}}}$ respectively.

When the activation pattern is expected to have a specific form, such as when prior information about the haemodynamic response is available, a correlation test between the observed and expected activation patterns can be used. Other statistical tests include the Fourier transform to identify voxels with components in the frequency domain that correspond to the stimulus frequency, use of the General Linear Model (GLM) [26, 29], Principal Component Analysis and clustering techniques. The result of combining such statistical tests at all voxel locations is a statistical parametric map (SPM) of brain activity. Without any priors on the anatomical location of active voxels, the entire SPM is assessed for significant effects related to the BOLD contrast. This assessment of the SPM also accounts for the multiplicity that arises by testing all voxel locations simultaneously.

All the methods mentioned above assume that fMRI data are normally distributed and are hence parametric. Randomization or permutation tests introduced by Holmes et al. [33], present a simple non-parametric alternative that can handle multiple comparisons [54]. If the 'on' and 'off' states of an fMRI time series are treated as condition labels, then under the null hypothesis randomly permuting these labels over the time series volumes should not significantly affect our test statistic (e.g., Student's t test). In this sense the acquired time series data are fixed while the 'on' and 'off' condition labels are assumed to be random realizations. Hence by randomly permuting these condition labels we can obtain a non-parametric distribution of the test-statistic, given the acquired time series. The value of the test-statistic corresponding to the actual arrangement of the 'on' and 'off'

labels is called the observed statistic T_{obvs}. The uncorrected p-value is given by the fraction of values in the test-statistic distribution that are greater than T_{obvs}. A thorough discussion of permutation tests with applications to fMRI data analysis has been presented by Nichols et al. [54].

Lastly we note that as the acquisition and analysis of fMRI data improve, it has become possible to analyze variations in BOLD contrast following a single stimulus presentation. Event-related fMRI experiments typically present different events at irregular intervals in a random order. These experiments measure transient changes in brain activity in response to the discrete stimuli, as opposed to the steady-state brain activity examined by epoch-based paradigms. Statistical analysis methods for event-related fMRI experiments are different from those discussed above, examples include the serial t-test and an analysis of the variance of voxel intensities described in [8].

In the following Chapter we focus on the effect of head motion on spin magnetization, i.e., spin saturation artifacts. We describe spin saturation artifacts using mathematical expressions and develop a correction scheme, called WASS correction, starting from Bloch equations. An algorithm to implement WASS correction using slice-to-volume (SV) registration is described. Lastly statistical analyses for two simulated fMRI time-series before and after SV motion and WASS correction illustrate the role of these methods in improving time-series activation detection.

CHAPTER 5

fMRI Time-series Spin Saturation Artifact Correction [1]

Functional MRI (fMRI) is a non-invasive tool for imaging brain function. The functionality of the brain relative to a particular stimulus is assessed by measuring stimulus triggered blood oxygenation level dependent (BOLD) signal intensity changes along a series of MR images acquired for the duration of the activation study, i.e., an fMRI time series. However, due to subject head motion, voxel intensities may be altered causing signal intensity changes dependent on positions of the head with respect to the scanner [28]. Though head restraints may reduce motion in some fMRI tasks, their use is impractical for patient studies in which discomfort may cause adverse reactions. Since the activation hypotheses for various brain regions are tested using statistical measures [25, 27, 54] to identify significant fMRI signal changes, this analysis is greatly skewed due to head motion during scans.

One manifestation of head motion is its confounding effect on spin magnetization, leading to an increase in signal modulation that is not related to the BOLD effect. Most fMRI time series are acquired by a multi-slice scheme using single shot Echo Planar Imaging (EPI). Fast acquisition of an EPI slice makes it possible to safely neglect head motion during a single slice excitation. However, due to changes in head position during the

---

[1]This chapter is based on material from [3].

multi-slice acquisition, slices in the EPI volumes no longer remain parallel to each other. Thus some nuclear spins in overlapping slice areas are re-excited before being allowed sufficient time to recover to their equilibrium states. Since EPI data are acquired at low spatial resolution, each EPI voxel can be approximated by a mixture of Gray Matter (GM), White Matter (WM) and Cerebrospinal Fluid (CSF) isochromats. Patient head motion may result in a loss of equilibrium magnetization states for some or all isochromats contributing to a particular EPI voxel, due to excitation at irregular time intervals. Thus these spin saturation artifacts, also called spin history artifacts [28, 52], make signal intensities a function of the subjects' movement history and adversely affect statistical analyses of fMRI data.

The spin saturation artifact at a particular single tissue voxel is a function of head position relative to B_0, repetition time (TR), echo time (TE), the effective flip angle and T1-dependent signal amplitude. Although recognized as a potential problem in fMRI activation analyses [28], to our knowledge, spin saturation artifacts have been handled using only volume-to-volume registration estimates. Volume-to-volume registration schemes assume that all slices in an acquired volume are aligned parallel to each other and may not provide a reasonable approximation of inter-slice head motion. This precludes the use of such volumetric motion estimates to assess and correct spin saturation artifacts induced by relative changes in slice positions in the same EPI volume. Muresan et al. [52] proposed a spin correction scheme that is applied prior to motion estimation. However, their treatment was restricted to the correction of spin saturation effects arising due to head translation between volume acquisitions only.

We devise a weighted average spin saturation (WASS) correction scheme that uses slice-to-volume (SV) registration motion estimates. The correction can handle full rigid motion and tries to account for the mélange of different brain tissue isochromats at each EPI voxel location, in removing spin saturation artifacts. SV rigid motion estimates are

obtained by mapping each EPI slice onto a higher resolution anatomical volume acquired from the same subject. Since anatomical data voxels are sampled finely enough to be approximated as tissue isochromats, they can be used to estimate unknown fractions of GM, WM and CSF contributing to each mapped EPI voxel. SV motion estimates and data acquisition parameters (TR, TE, etc.) can be used to identify voxels with spin saturation artifacts. To alleviate partial volume effects, the WASS correction factor for each affected EPI voxel is approximated by a weighted average of the correction factors of its constituent brain tissue isochromats as identified by its mapping onto the anatomical volume.

Sections 5.1 and 5.2 describe the spin saturation effect using mathematical expressions and develop the WASS correction to compensate EPI voxels showing spin saturation artifacts. Two realistic fMRI time series with known rigid motion and corresponding spin saturation artifacts were simulated to evaluate the WASS correction method. Ranges of the induced 3D rotational head motion were $\pm 5°$ and $\pm 2°$, respectively. Activation was introduced in manually selected brain regions assuming an epoch based experimental paradigm with a box-car stimulation sequence. WASS correction using SV motion estimates was used to correct these simulated data for spin saturation related voxel intensity variation. Receiver operating characteristic (ROC) curves were used to compare activation detection for both simulated times series before and after motion and WASS correction with that obtained from artifact-free data. Results indicate that retrospective WASS correction based on SV motion estimates may have a significant role in improving activation detection.

5.1 Spin saturation artifacts in tissue isochromats

To acquire a multi-slice fMRI time series, longitudinal magnetizations of spins in slice-like regions of the subject's brain are selectively excited and flipped onto the transverse plane using an α-angle RF pulse. The intensity of a tissue isochromat at a given coor-

dinate in an MR volume is proportional to the effective flip angle and the magnetization component in the XY plane, at that location [32]. For now we focus on the effect of patient head motion on the intensity of a single tissue isochromat. Let $m_z^0(\vec{v})$ be the initial magnetization (at rest) of a single tissue isochromat at location $\vec{v} \in \mathbb{R}^3$, characterized by time constants T1 and T2. Let $m_{z,i}^-(\vec{v})$ be the longitudinal magnetization just before the ith RF pulse with effective flip angle $\alpha_i(\vec{v})$. Imperfections in the slice excitation profile and B_0 inhomogeneities may make the effective flip angle vary spatially over the slice. The longitudinal magnetization $m_{z,i}^+(\vec{v})$, just after the ith excitation pulse is given by

$$(5.1) \qquad m_{z,i}^+(\vec{v}) = m_{z,i}^-(\vec{v}) \cos(\alpha_i(\vec{v})),$$

where $m_{z,1}^-(\vec{v}) = m_z^0(\vec{v})$. Let $t_i, i = 1, 2, ...$ be the time interval between the ith and $i+1$th consecutive excitation of the tissue isochromat, then the longitudinal magnetization between excitations recovers according to;

$$
\begin{aligned}
m_{z,i+1}^-(\vec{v}) &= m_{z,i}^+(\vec{v})e^{-\frac{t_i}{T1}} + m_z^0(\vec{v})(1 - e^{-\frac{t_i}{T1}}) \\
(5.2) \qquad &= m_{z,i}^-(\vec{v})\cos(\alpha_i(\vec{v}))e^{-\frac{t_i}{T1}} + m_z^0(\vec{v})(1 - e^{-\frac{t_i}{T1}}).
\end{aligned}
$$

The brain volume of interest is repeatedly excited M times before conducting the activation study so as to force all magnetization vectors to achieve an incoherent steady state [30]. In such a state, if each tissue isochromat is excited every TR seconds, $m_{z,i}^-(\vec{v}) = m_z^{\text{ss}}(\vec{v}), \forall i \geq M$. This steady state magnetization $m_z^{\text{ss}}(\vec{v})$ can be expressed in terms of the initial magnetization using (5.2),

$$
\begin{aligned}
m_{z,M+1}^-(\vec{v}) &= m_{z,M}^-(\vec{v})\cos(\alpha_M(\vec{v}))e^{-\frac{TR}{T1}} + m_z^0(\vec{v})(1 - e^{-\frac{TR}{T1}}) \\
&= m_z^{\text{ss}}(\vec{v}), \quad \alpha_i(\vec{v}) = \alpha(\vec{v})\forall i \\
(5.3) \quad \Rightarrow m_z^{\text{ss}}(\vec{v}) &= m_z^0(\vec{v})f_{\text{ss}}(\vec{v}) \quad \text{where } f_{\text{ss}}(\vec{v}) \triangleq \frac{1 - e^{-\frac{TR}{T1}}}{1 - \cos(\alpha(\vec{v}))e^{-\frac{TR}{T1}}}, \ i \geq M.
\end{aligned}
$$

Hence in the absence of head motion the artifact free intensity I_{true} is proportional to the transverse magnetization, $m_{xy}^{\text{ss}}(\vec{v}) = m_z^{\text{ss}}(\vec{v})\sin(\alpha(\vec{v}))$,

$$(5.4) \qquad I_{\text{true}}(\vec{v}) \propto m_{xy}^{\text{ss}}(\vec{v})e^{-\frac{TE}{T2}} = m_z^0(\vec{v})\frac{1 - e^{-\frac{TR}{T1}}}{1 - \cos(\alpha(\vec{v}))e^{-\frac{TR}{T1}}}\sin(\alpha(\vec{v}))e^{-\frac{TE}{T2}}.$$

In the presence of head motion between slice acquisitions, EPI slices in a volume are not parallel to each other. Hence the time difference between the successive excitations of some spins cannot be maintained at TR ms, causing the incoherent steady state established by (5.3) to break down. Thus (5.4) is rarely an accurate representation of isochromatic tissue intensities in an fMRI study. In particular, the intensity of a given tissue isochromat will drop if less than TR ms have elapsed since its last excitation causing the longitudinal magnetization at that location to be re-excited before relaxing to $m_z^{\text{ss}}(\vec{v})$. The effect of n irregularly spaced excitations on $m_{z,n}^-(\vec{v})$ can be found by repeatedly using eqs. (5.1) and (5.2). Specifically it can be shown that

$$(5.5) \qquad\qquad\qquad m_{z,n}^-(\vec{v}) = m_z^0(\vec{v})f_n(\vec{v});$$

where $f_n(\vec{v})$, a function of tissue T1 the effective flip angles and previous head positions via $t_i, i = 1, 2, ...n - 1$, is recursively given by

$$(5.6) \qquad\qquad f_{i+1}(\vec{v}) = f_i(\vec{v})\cos(\alpha_i(\vec{v}))e^{-\frac{t_i}{T1}} + (1 - e^{-\frac{t_i}{T1}}), \; i = 1, 2, ...$$

with $f_1(\vec{v}) = (1 - e^{-\frac{TR}{T1}})/(1 - \cos(\alpha(\vec{v}))e^{-\frac{TR}{T1}})$, assuming that the isochromat was initially in its incoherent steady sate given by (5.3). The corresponding observed intensity affected by the spin saturation phenomenon, $I_{\text{obvs}}(\vec{v}, n)$, is proportional to the transverse magnetization $m_{xy,n}(\vec{v})$;

$$(5.7) \qquad\qquad I_{\text{obvs}}(\vec{v}, n) \propto m_{xy,n}(\vec{v})e^{-\frac{TE}{T2}} = m_z^0(\vec{v})f_n(\vec{v})\sin(\alpha_n(\vec{v}))e^{-\frac{TE}{T2}}.$$

The recursive (5.6) is valid only for tissue isochromats or small voxels. However, to gain temporal resolution, fMRI scans typically have relatively large voxel sizes. Thus the in-

tensity of each fMRI voxel is proportional to the average transverse magnetization of a mixture of GM, WM and CSF isochromats.

5.2   Spin saturation artifact correction for EPI voxels

Spin saturation artifacts are a direct consequence of head motion and voxels affected by them can be identified using head trajectory estimates, head geometry and fMRI time series attributes. As noted earlier EPI voxels have low spatial resolution. Thus the intensity of a single EPI voxel in an fMRI time series reflects the average transverse magnetization over a small brain volume made up of a mélange of different brain tissue. Consequently the effective time constants $T1^{epi}$ and $T2^{*epi}$ of the EPI voxel cannot be estimated satisfactorily using a single brain tissue.

Consider the acquisition of an fMRI time series using EPI with repetition time TR, echo time TE and α-angle RF pulses. Let the time series contain V volumes with S slices in each volume. Let $n_{\text{epi}} = (v-1)S + s, s = 1, 2, \ldots S$ and $v = 1, 2, \ldots V$ index the total number of slices in the EPI time series by acquisition order. In the presence of subject head motion, the observed intensity of an EPI voxel at coordinate $\vec{u} \in \mathbb{R}^3$ in slice n_{epi} of the time series is given by:

$$(5.8) \qquad I_{\text{obvs}}(\vec{u}, n_{\text{epi}}) \propto m^-_{z,\text{epi}}(\vec{u}, n_{\text{epi}}) \sin(\alpha(\vec{u})) e^{-\frac{TE}{T2^{*epi}}}, \quad \text{where } n_{\text{epi}} = 1, 2, \ldots VS$$

$m^-_{z,\text{epi}}(\vec{u}, n_{\text{epi}})$ is the longitudinal magnetization of the EPI voxel just before the excitation pulse for the n_{epi}th slice of the time series. Due to subject head motion, some of the S slices in each EPI volume may overlap. Consequently, spins in overlapping portions of these slices will be excited at irregular time intervals causing them to deviate from their induced steady state. This transitory response of some spin magnetizations will result in spin saturation, causing the longitudinal magnetization at a given coordinate location to vary across volumes. However, in the absence of subject motion, the magnetization

vectors in the entire brain volume can be assumed to be in an induced incoherent steady state. The intensity of an EPI voxel at location $\vec{u}$ assuming no motion is given by:

$$(5.9) \qquad I_{\text{true}}(\vec{u}) \propto m_{z,\text{epi}}^{ss}(\vec{u}) \sin(\alpha(\vec{u})) e^{-\frac{TE}{T2^{*\text{epi}}}},$$

where $m_{z,\text{epi}}^{ss}(\vec{u})$ is the incoherent steady state magnetization of the EPI voxel. In the absence of subject motion and signal noise this steady state magnetization at a given location $\vec{u}$ will not change across the time series volumes due to spin saturation.

The relation between the observed intensity of an EPI voxel with and without spin saturation artifacts is given by

$$(5.10) \qquad I_{\text{true}}(\vec{u}) \approx I_{\text{obvs}}(\vec{u}, n_{\text{epi}}) \frac{m_{z,\text{epi}}^{ss}(\vec{u})}{m_{z,\text{epi}}^{-}(\vec{u}, n_{\text{epi}})}.$$

Thus spin saturation artifacts can be detected and corrected, if we can approximate the ratio $m_{z,\text{epi}}^{ss}(.)/m_{z,\text{epi}}^{-}(.)$ at each EPI voxel location. Since such a spin saturation correction mechanism will appropriately scale observed intensity values $I_{\text{obvs}}(.)$, it will be unable to correct spin saturation at locations where the artifact is severe enough to make the observed intensity zero.

For brevity, we assume without deliberation, the existence of an SV registration scheme that can obtain reasonably accurate head motion estimates by registering each EPI slice onto a high resolution anatomical MR volume of the same subject [38]. Further these motion estimates can be used to map every large EPI voxel onto a K-neighborhood of finer anatomical volume voxels. Specific details of such a registration process are outlined in Sec. 5.4.1.

Approximating the minute GM, WM and CSF isochromats underlying each mapped EPI voxel by the K-neighborhood of small anatomical volume voxels; we can express the longitudinal magnetization of the EPI voxel at coordinate $\vec{u} \in \mathbb{R}^3$ as the average magneti-

zation over this neighborhood

$$m_{z,\text{epi}}^0(\vec{u}) \approx \frac{1}{K} \sum_{\vec{v}_k \in N(\vec{u})} m_z^0(\vec{v}_k); \quad m_{z,\text{epi}}^{ss}(\vec{u}) \approx \frac{1}{K} \sum_{\vec{v}_k \in N(\vec{u})} m_z^{ss}(\vec{v}_k)$$

(5.11) $$\text{and } m_{z,\text{epi}}^-(\vec{u}, n_{\text{epi}}) \approx \frac{1}{K} \sum_{\vec{v}_k \in N(\vec{u})} m_{z,n_k}^-(\vec{v}_k), \quad n_k \leq n_{\text{epi}}.$$

Where $N(\vec{u})$ denotes the neighborhood containing K finer voxels, $m_{z,\text{epi}}^0(\vec{u})$ is the initial (at rest) magnetization of the EPI voxel, $m_{z,\text{epi}}^{ss}(\vec{u})$ is its incoherent steady state magnetization assuming no head motion and $m_{z,\text{epi}}^-(\vec{u}, n_{\text{epi}})$ is the longitudinal magnetization just before the excitation of the n_{epi}th EPI slice at time point $t_{n_{\text{epi}}}$. The number of consecutive excitations n_k for an approximated isochromat at $\vec{v}_k$, is the number of times series slices that were mapped on to that 'isochromat' up to time $t_{n_{\text{epi}}}$. The time between each of these n_k consecutive excitations is given by the time elapsed between the excitation of the corresponding EPI slices.

We define unknown correction factors $f_{ss}^{\text{wa}}(\vec{u})$ and $f_{n_{\text{epi}}}^{\text{wa}}(\vec{u}, n_{\text{epi}})$, along the lines of (5.3) and (5.5), such that

(5.12) $$m_{z,\text{epi}}^{ss}(\vec{u}) \triangleq m_{z,\text{epi}}^0(\vec{u}) f_{ss}^{\text{wa}}(\vec{u}) \quad \text{and} \quad m_{z,\text{epi}}^-(\vec{u}, n_{\text{epi}}) \triangleq m_{z,\text{epi}}^0(\vec{u}) f_{n_{\text{epi}}}^{\text{wa}}(\vec{u}).$$

Approximating the longitudinal magnetizations $m_{z,\text{epi}}(.)$ above by the approximations in (5.11) gives:

$$f_{ss}^{\text{wa}}(\vec{u}) \approx \hat{f}_{ss}^{\text{wa}}(\vec{u}) = \sum_{\vec{v}_k \in N(\vec{u})} \frac{m_z^{ss}(\vec{v}_k)}{\sum_{\vec{v}_j \in N(\vec{u})} m_z^0(\vec{v}_j)}$$

(5.13) $$\text{and } f_{n_{\text{epi}}}^{\text{wa}}(\vec{u}) \approx \hat{f}_{n_{\text{epi}}}^{\text{wa}}(\vec{u}) = \sum_{\vec{v}_k \in N(\vec{u})} \frac{m_{z,n_k}^-(\vec{v}_k)}{\sum_{\vec{v}_j \in N(\vec{u})} m_z^0(\vec{v}_j)}.$$

Since the anatomical volume voxels are approximated as GM, WM and CSF isochromats, we can use (5.3) and (5.5) to estimate the two correction factors:

$$\hat{f}_{ss}^{\text{wa}}(\vec{u}) = \sum_{\vec{v}_k \in N(\vec{u})} \frac{m_z^0(\vec{v}_k)}{\sum_{\vec{v}_j \in N(\vec{u})} m_z^0(\vec{v}_j)} f_{ss}(\vec{v}_k)$$

(5.14) $$\text{and } \hat{f}_{n_{\text{epi}}}^{\text{wa}}(\vec{u}) = \sum_{\vec{v}_k \in N(\vec{u})} \frac{m_z^0(\vec{v}_k)}{\sum_{\vec{v}_j \in N(\vec{u})} m_z^0(\vec{v}_j)} f_{n_k}(\vec{v}_k).$$

where the incoherent steady state factors f_{ss} are given by (5.3). The spin saturation dependence of $\hat{f}^{\text{wa}}_{n_{\text{epi}}}$ is captured by the factors f_{n_k}, given by the recursion in (5.6) using time constants $T1_{\text{GM}}$, $T1_{\text{WM}}$ or $T1_{\text{CSF}}$ as appropriate.

As each anatomical volume voxel has the same size and the initial magnetization of an isochromat is proportional to its proton density ρ_0, the equilibrium magnetization m_z^0 can be replaced by ρ_0 in (5.14). Further as only relative proton density values of GM, WM and CSF up to a common factor are required, relative proton densities ρ_0 (dependent on tissue water content) $\rho_{0,\text{GM}} = 0.80$, $\rho_{0,\text{WM}} = 0.72$ and $\rho_{0,\text{CSF}} = 1.0$ can be used, i.e.

$$\hat{f}^{\text{wa}}_{ss}(\vec{u}) = \sum_{\vec{v}_k \in N(\vec{u})} \frac{\rho_0(\vec{v}_k)}{\sum_{\vec{v}_j \in N(\vec{u})} \rho_0(\vec{v}_j)} f_{ss}(\vec{v}_k)$$

(5.15)
$$\text{and } \hat{f}^{\text{wa}}_{n_{\text{epi}}}(\vec{u}) = \sum_{\vec{v}_k \in N(\vec{u})} \frac{\rho_0(\vec{v}_k)}{\sum_{\vec{v}_j \in N(\vec{u})} \rho_0(\vec{v}_j)} f_{n_k}(\vec{v}_k)$$

Finally using (5.10), (5.12) and (5.13) I_{true} can be approximated by

(5.16)
$$\hat{I}_{\text{true}}(\vec{u}, n_{\text{epi}}) = I_{\text{obvs}}(\vec{u}, n_{\text{epi}}) \frac{\hat{f}^{\text{wa}}_{ss}(\vec{u})}{\hat{f}^{\text{wa}}_{n_{\text{epi}}}(\vec{u})}.$$

In summary, the estimate $\hat{I}_{\text{true}}(\vec{u}, n_{\text{epi}})$ uses knowledge of time series acquisition parameters (TR, α, ...) and approximate tissue T1 and ρ_0 values at each anatomical volume grid location. The algorithm used to implement (5.15) and (5.16) is described in Sec. 5.4.2. In subsequent sections, for clarity, we call this approach Weighted Average Spin Saturation (WASS) correction.

5.3 fMRI Time-series Simulation

Two simulated 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) [9]. Each voxel in this volume was assumed to be a GM, WM or CSF isochromat with known T1 values of 833ms, 500ms, 2569ms, respectively. Head

motion was simulated by rotating the T2-weighted volume in three dimensional space prior to extracting each EPI slice, to form an fMRI volume. The head was assumed to be moving in the scanner's frame of reference while the coordinates at which the EPI slice was imaged by the scanner were fixed. Practically observed head rotations over successive volumes in the time series maintain continuity with previous head positions. Hence, the applied motion was designed to be smooth without being periodic.

The range of motion in an fMRI time series varies with the level of corporation from a subject as well as the tasks being studied. Typically volumetric registration motion estimates of up to ± 1 mm translation and $\pm 1°$ rotations have been reported for normal subjects [28]. In contrast, registration results for an aged subject in the same study included translations up to 5 mm and rotations up to $6°$.

Typically, fMRI tasks involving verbal responses (for e.g. picture naming) may result in larger head motion due to jaw and mouth movements. To estimate the range of realistic through plane motion in such verbal tasks, real time MRI scans were obtained from a normal volunteer without head restraints using 2D turbo field echo (TFE) in a Philips 3T System. The images were acquired while the subject was verbalizing words typically used in language tasks for fMRI studies. A hundred $140 \times 108, 10$ mm thick sagittal brain images were obtained with a dynamic scan time of 251 ms. The relative range of inter-slice motion was estimated by registering each image to the initial image as a reference. Registration results indicated rotational and translational motion of up to ± 5.4 degrees and ± 5.2 mm.

To consider the range of motion commonly used to evaluate most fMRI studies [23,52] two time series were simulated with $\pm 5°$ and $\pm 2°$ degree rotational head motion. The average magnitude of rotation between consecutive slice acquisitions was $0.25°$ and $0.09°$ respectively. Each simulated time series consisted of 120 volumes with 14 slices in each

volume. Rotation angles describing head position about the three coordinate axes between volume acquisitions were obtained by drawing three sets of 121 iid random numbers from a uniform distribution. To ensure smoothness, a 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 time series parameters for both simulations were TR = 3000 ms for each EPI volume, $\alpha = 90°$ and an interleaved slice acquisition sequence. Time series parameters and head trajectories constructed above determine the time instances at which every voxel in the T2 weighted MR volume was imaged by the scanner. Specifically each EPI slice, with acquisition index $n_{\text{epi}} = 1, 2, \ldots 120 \times 14$, was acquired approximately at time $t_{n_{\text{epi}}} = n_{\text{epi}} \times 3000/14$ ms. Spin saturation artifacts were introduced in the T2 volume by treating each high resolution T2 voxel as a WM, GM or CSF isochromat and using (5.4) and (5.7). The activation task was assumed to be a block design, alternating a stimulus and a control cycle every 10 volumes. Activated voxels were created by increasing intensities in manually marked regions of the high resolution T2 volume by 2% before simulating motion and spin saturation artifacts. Care was taken to ensure that the activated EPI voxels would contain mainly GM and some WM isochromats. $2 \times 2 \times 6$ mm^3 thick EPI voxels acquired at time points $t_{n_{\text{epi}}}$ were obtained by averaging voxel intensities in the corresponding $2 \times 2 \times 6$ neighborhood of the T2 volume. The ground truth activation map was obtained by downsampling the high resolution activation map to the resolution of the EPI volumes. To account for signal noise in MR magnitude images, Gaussian noise ($N(0, 4)$) and Rayleigh noise ($\sigma = 2$) were added to EPI voxels with non-zero and no signal intensities, respectively [5, 55].

5.4   Motion and Spin Saturation Artifact Correction

Motion correction approaches using volume-to-volume registration are not designed to detect subject head motion between slice acquisitions correctly. Numerous fMRI analysis techniques apply the same rigid transform to the whole volume [24, 75] allowing no inter-slice motion. Further many use L2-metrics which have been shown to result in spurious motion estimates in the presence of activation [22, 23]. Since spin saturation artifacts have magnitudes comparable to activation [52], these methods may give erroneous motion estimates for spin saturation affected time series. In contrast the Mutual Information (MI) metric has been shown to be relatively unaffected by activation related intensity fluctuations [23]. Hence we obtain motion estimates using Map-Slice-to-Volume (MSV) [38], an MI-based slice-to-volume registration algorithm briefly described in the following section.

5.4.1   Map-Slice-to-Volume motion estimation and hypothesis testing

The MSV registration algorithm has been shown to be effective in the post processing and analysis of human fMRI data [38]. The MSV algorithm models 3D motion of multi-slice EPI data by allowing each slice to have its own rigid body transform with six degrees of freedom. Each rigid transform maps the EPI slice on to a high resolution T1 volume acquired in the same fMRI session. The vector of rigid motion parameters $\phi^{n_{\mathrm{epi}}} = [t_x, t_y, t_z, \theta_x, \theta_y, \theta_z]$ for EPI slice number n_{epi}, is obtained by maximizing a plug-in approximation of MI between the EPI slice and the T1 volume. Probability density function estimates, needed to approximate MI, are computed using histograms of voxel intensities in the region of overlap between the EPI slice and T1 volume. This makes the plug-in MI metric non-differentiable; consequently MSV uses the Nelder-Mead downhill simplex optimizer [53].

Though we have restricted our motion estimates only to rigid transforms, changes in

field inhomogeneity may result in related geometric distortions of EPI data. In such cases, the subsequent WASS correction is assumed to be applied after compensating the time series for effects of field inhomogeneity [14, 51, 60, 76]. Alternatively, in some cases SV registration can be adapted to estimate non-linear warps for geometric distortion correction [39, 43].

Hypothesis testing of each candidate time series after only MSV motion correction or both motion and WASS correction (described in the following section) was performed using 2000 random permutation tests [54]. The test statistic used was Student's t test given in (4.1). Activation maps were obtained by comparing the calculated p-values with different alpha threshold values in the range of $[5 \times 10^{-4}, 1]$. Finally, ROC curves were generated by computing fractions of false positive and true positive counts relative to the known manually introduced activation pattern.

5.4.2 Weighted average spin saturation correction

The WASS correction proceeds in a sequential manner, processing each EPI slice indexed according to acquisition order by the same procedure. Similar to MSV, WASS correction uses a high resolution T1 volume acquired during the same fMRI session. Each voxel in the T1 volume is treated as an approximate GM, WM or CSF isochromat with corresponding approximate T1 and ρ_0 values. As described by the flow chart in Fig 5.1, a 'correction factor' value and a 'time elapsed' counter is associated with every T1 voxel. Each EPI slice is mapped onto the T1 volume using the corresponding MSV motion estimates. This mapping is then used to track the time elapsed since the previous excitation of each T1 voxel location, in the 'time elapsed' counters. The 'correction factor' values $f_{n_k}(\vec{v}_k), \vec{v}_k \in R$ for each T1 voxel are recursively updated by (5.6), using the 'time elapsed' counters and appropriate approximate T1 values.

For a given EPI voxel in slice n_{epi} excited at time point $t_{n_{\text{epi}}}$, the factors $\hat{f}^{\text{wa}}_{n_{\text{epi}}}$ are computed using the updated 'correction factors' f_{n_k} and approximate ρ_0 values in its T1 neighborhood using (5.15). Each f_{n_k} in the T1 neighborhood could have been updated a distinct number of times. This number of updates depends on the number of EPI slices that were mapped onto the T1 voxel location up to time $t_{n_{\text{epi}}}$. Since $\hat{f}^{\text{wa}}_{ss}$ is independent of the time between excitations; it can be computed straightforwardly using (5.3) and (5.15). Lastly, the EPI voxel intensity is corrected using (5.16). This process is sequentially repeated for every EPI slice in the time series. The T1 volume is used only as a means of approximating the percentage contributions of GM, WM and CSF in each EPI voxel. The WASS correction does not use any intensity values from the high resolution T1 volume.

For synthetic data we have access to an anatomical volume with the same accurate classification of GM, WM and CSF, as that used to induce spin saturation artifacts. In reality spin saturation artifacts can be modelled as effects of the premature excitation of a collection of infinitesimally small isochromats in the vicinity of the EPI voxel. In the real data case, due to the finite resolution of the anatomical volume, there is an inherent inaccuracy in the classification of GM, WM and CSF isochromats. To account for this error in classification in our WASS correction, a 'blurred' ICBM T1 volume was used as the anatomical volume. The blurred volume was created by using a single 'dominant' tissue to classify non-overlapping groups of 3 voxels (i.e. 3 mm) along the z-axis. This dominant single tissue classifier was obtained from a $1 \times 1 \times 3$ mm^3 resolution T1 volume downloaded from ICBM. The mis-classification was applied only along the longitudinal direction as spin saturation artifacts occur mainly due to out-of-plane motion [28]. In contrast, the time series simulation process used a high resolution volume ($1 \times 1 \times 1$ mm^3) for the tissue classification.

INPUTS

Anatomical volume (T1 and ρ-map)

MSV motion estimates $(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z)$

Spin Sat Corrupted Time Series

Time Series Parameters- **TR**, **S** (slices per EPI volume), **α** (flip angle)

Extract:
EPI slice number s from volume v.

Corresponding motion estimates

Initializations
s=1, v=1
Set all Correction factors to $f_{ss}(\vec{v}_k)$
Set all time elapsed counters to 'infinity'

Internal variables
s- EPI slice no. counter
v- EPI volume no. counter

Variables per anatomical voxel-
1) Correction factor
2) Time elapsed since last excitation

Map the EPI slice onto the Anatomical volume. Mark voxels in the anatomical neighborhood.

For each anatomical volume voxel do-

IF the anatomical voxel is marked
- Update the *factors* $f_{n_k}(\vec{v}_k)$ using recursive eq (6) and approximate GM, WM and CSF T1 values and the *time elapsed since previous excitation.*
- Reset the time elapsed since previous excitation to 0.

ELSE
- Increment the time elapsed since the last excitation by TR/S.

For each EPI slice voxel do-

Find the EPI voxel neighborhood in the anatomical volume and use the **updated** factors to calculate the correction factor:

$$\frac{\displaystyle\sum_{\vec{v}_k \in N(\vec{u})} \frac{\rho_0(\vec{v}_k)}{\sum_{\vec{v}_j \in N(\vec{u})} \rho_0(\vec{v}_j)} f_{ss}(\vec{v}_k)}{\displaystyle\sum_{\vec{v}_k \in N(\vec{u})} \frac{\rho_0(\vec{v}_k)}{\sum_{\vec{v}_j \in N(\vec{u})} \rho_0(\vec{v}_j)} f_{n_k}(\vec{v}_k)}$$

Correct the observed EPI voxel intensity.

s = s+1 ←N— All EPI slices processed? —Y→

v = v+1 ←N— All EPI volumes processed? —Y→ STOP

Figure 5.1: Flowchart describing the implementation of WASS correction.

	No Correction	Artifact-free Data	Only Motion Correction	Reduction in AUC due to Spin Sat
$\pm 5^o$	0.614	0.826	0.793	0.033 (18.4%)
$\pm 2^o$	0.763	0.868	0.856	0.012 (12.9%)

Table 5.1: Approximate AUC values quantifying the effect of spin saturation on activation detection. Corresponding ROC curves are shown in Fig. 5.2. The loss in AUC between 'Artifact-free' data and 'Only Motion Correction' can be attributed to spin saturation artifacts alone. This loss is denoted as a percentage of the improvement in AUC after 'Only Motion Correction' in parenthesis.

5.5 Results using Simulated Data

Two simulated fMRI time series with induced rotational head motion of $\pm 5°$ and $\pm 2°$ and corresponding spin saturation artifacts were treated as 'observed' time series data. For reference, two time series free of the spin saturation effect were also generated using the same motion trajectories. These time series were corrected using known true motion to establish artifact-free data. This 'artifact-free' data was treated as the reference pseudo-ground truth.

The effect of only spin saturation artifacts on activation detection was evaluated by correcting both 'observed' time series for motion only using the known true motion parameters. ROC curves following hypothesis testing are shown in Fig. 5.2. The degrading effect of spin saturation artifacts on activation detection was quantified by computing approximate area under the curve (AUC) values for both ROC curves, listed in Table 5.1.

As expected, the effect of spin saturation reduced almost proportionately with reduction in head motion. In particular the loss in AUC due to spin saturation alone was 0.033 for the simulated time series with $\pm 5°$ head rotations and 0.012 for the time series with $\pm 2°$ rotations. Activation maps showing the effect of spin saturation artifacts on activation detection for both simulated time series are shown in Fig. 5.3.
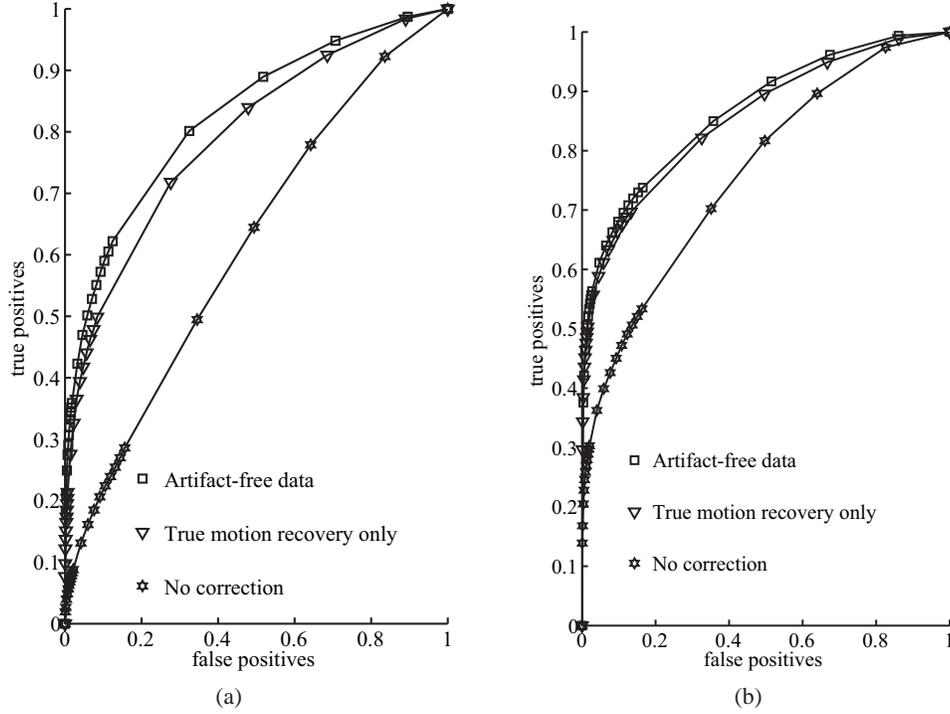
Figure 5.2: Effect of simulated spin saturation artifacts on activation detection for two time series with in-duced head rotations of (a)±5 deg and (b)±2 deg (max/min). Plots show ROC curves for simu-lated data with motion and related spin saturation artifacts following 'No Correction' and 'True motion recovery' only. 'Artifact-free' data ROC curves are included for comparison.

5.5.1 Effect of Inaccuracies in T1 Values on WASS Correction

In reality, T1 values for GM, WM and CSF vary across the subject population and are known only approximately. To study the effect of inaccuracies in T1 values on our spin saturation correction, three WASS corrections were performed on both 'observed' time series with (1) exact ICBM T1 values, $T1_{GM}$ = 833ms, $T1_{WM}$ = 500ms and $T1_{CSF}$ = 2569ms, (2) approximate (text-book) T1 values [30], $T1_{GM}$ = 900ms, $T1_{WM}$ = 600ms and $T1_{CSF}$ = 4000ms, and (3) a single T1 value, $T1_{GM}$= 833ms. In each case, the average percentage error in voxel intensities after WASS correction with respect to artifact-free intensities for each time series volume, was obtained as follows

$$(5.17) \qquad \text{Avg. \% error} = \frac{1}{N} \sum_{i=1}^{N} \frac{\left| I_{\text{true}}(\vec{u}_i) - \hat{I}_{\text{true}}(\vec{u}_i) \right|}{I_{\text{true}}(\vec{u}_i)},$$
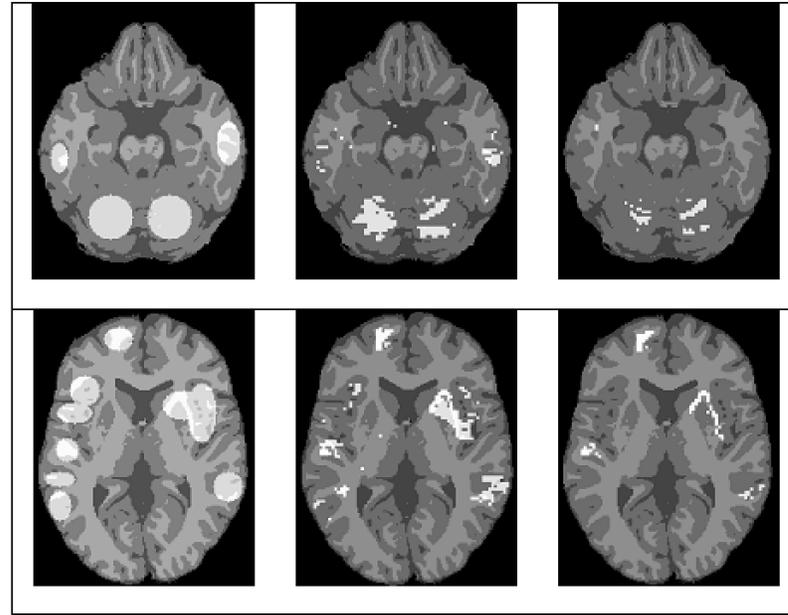
(a)
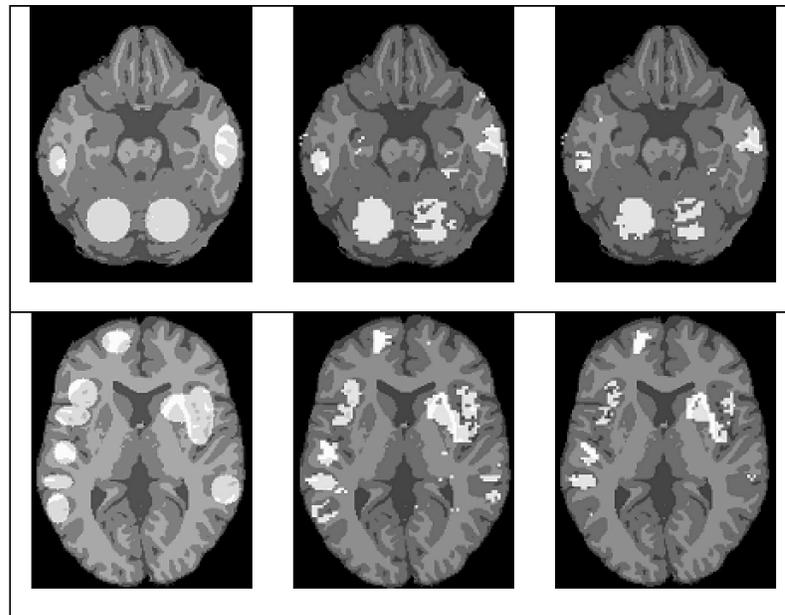
(b)

Figure 5.3: Activation maps (alpha = 0.002) for simulated time-series with (a) $\pm 5°$ and (b) $\pm 2°$ motion superimposed on the anatomical T1-weighted volume. Rows show (first column) manually applied true activation patterns for two selected slices, corresponding activation maps from (second column) artifact-free data and (third column) simulated time-series with spin saturation after only exact (true) motion correction.

where N is the number of non-zero intensity voxels in the region of interest (ROI) for each fMRI volume. $\hat{I}_{\text{true}}(\vec{u})$ is the intensity at coordinate $\vec{u}$ in the WASS corrected time series and $I_{\text{true}}(\vec{u})$ is the corresponding known true intensity obtained from the artifact-free time series.

As $T1_{\text{CSF}}$ is significantly different from $T1_{\text{GM}}$ and $T1_{\text{WM}}$, misclassifications of CSF as GM or WM will contribute strongly to the error metric in (5.17). Since WASS using a single T1 uses only $T1_{\text{GM}}$ in our comparison, inclusion of CSF in the ROI may bias this error metric toward WASS using distinct $T1_{\text{GM}}$, $T1_{\text{WM}}$ and $T1_{\text{CSF}}$ values. Hence, to account for the prevalence of activation in GM, we restrict the ROI in (5.17) to the manually introduced activation pattern only.

The standard deviation (SD) and max/min of the avg. % intensity error values are listed in Table 5.2. These results indicate a reduction in intensity errors following WASS correction. Reduction in errors after WASS correction using approximate T1 values was comparable to that using exact T1 values. Further, for the larger $\pm 5°$ motion WASS correction using approximate T1 values significantly outperformed that using a single T1 value. In contrast this improvement in performance was noticeably smaller for the time series with $\pm 2°$ motion.

	No Spin Sat Correction	WASS Correction using -		
		Exact T1s	Approx T1s	Single (exact) T1
$\pm 5^o$				
Max/Min	5.37 / 0.29	1.02 / 0.06	1.39 / 0.08	2.51 / 0.13
SD	1.20	0.19	0.29	0.52
$\pm 2^o$				
Max/Min	2.31 / 0.09	0.55 / 0.03	0.72 / 0.03	1.12 / 0.05
SD	0.50	0.11	0.15	0.22

Table 5.2: Comparison of Avg. % intensity errors following WASS correction using exact T1s, approximate T1s and a single (exact GM) T1 for simulated time series with $\pm 5°$ and $\pm 2°$ rotational head motion. Max/Min and standard deviation (SD) of Avg. % intensity errors computed using (5.17) are tabulated. The ROI was restricted to manually added activation regions in each time series volume.

5.5.2 MSV-based WASS Correction Using Approximate T1 Values

The viability of WASS correction depends on its ability to withstand the combined effect of errors in inter-slice motion estimation, approximate T1 values and approximate relative proton densities. Motion estimates for both observed time series with signal noise and $\pm 5°$ and $\pm 2°$ rotational motion were obtained using the MSV algorithm. Intensity modulation due to simulated spin saturation artifacts did not significantly affect the accuracy of MSV. RMS errors between MSV motion estimates and the known true motion are listed in Table 5.3.

	Rotation RMSE (deg)			Translation RMSE (mm)		
	Rx	Ry	Rz	Tx	Ty	Tz
$\pm 5^o$	0.3555	0.3402	0.0908	0.1627	0.1796	0.1787
$\pm 2^o$	0.2187	0.2067	0.1278	0.1257	0.1449	0.1345

Table 5.3: RMSE values between the applied ground truth motion and the motion estimates recovered by MI-based MSV for both simulated time series.

These motion estimates were then used to process both time series using WASS correction. WASS correction was performed using exact (ICBM T1s), approximate (textbook T1s) and a single (ICBM GM) T1. ROC curves following hypothesis testing for both 'observed' time series after MSV motion correction with and without WASS correction are shown in fig 5.4.

Approximate area under the ROC curve (AUC) values are tabulated in Table 5.4. The 'Approx T1s' AUC column is somewhat representative of activation detection following WASS correction in a realistic scenario i.e. when only MSV motion estimates and approximate T1 values are available. To gauge the value of performing 'MSV + WASS Correction' over 'Only MSV Motion Correction', the increase in AUC due to WASS correction alone should be considered relative to that due to MSV motion correction alone.

Percentage values in the 'Approx T1s' column in Table 5.4 were given by

$$\% \text{ increase in AUC due to WASS alone} = \frac{\text{AUC}_{\text{MSV+WASS}} - \text{AUC}_{\text{MSV}}}{\text{AUC}_{\text{MSV}} - \text{AUC}_{\text{NC}}} \times 100.$$

Where $\text{AUC}_{\text{MSV+WASS}}$ is the AUC in the 'Approx T1s' column, AUC_{MSV} is the AUC following only MSV motion correction and AUC_{NC} is the initial 'No Correction' AUC from Table 5.1, obtained without any time series correction.
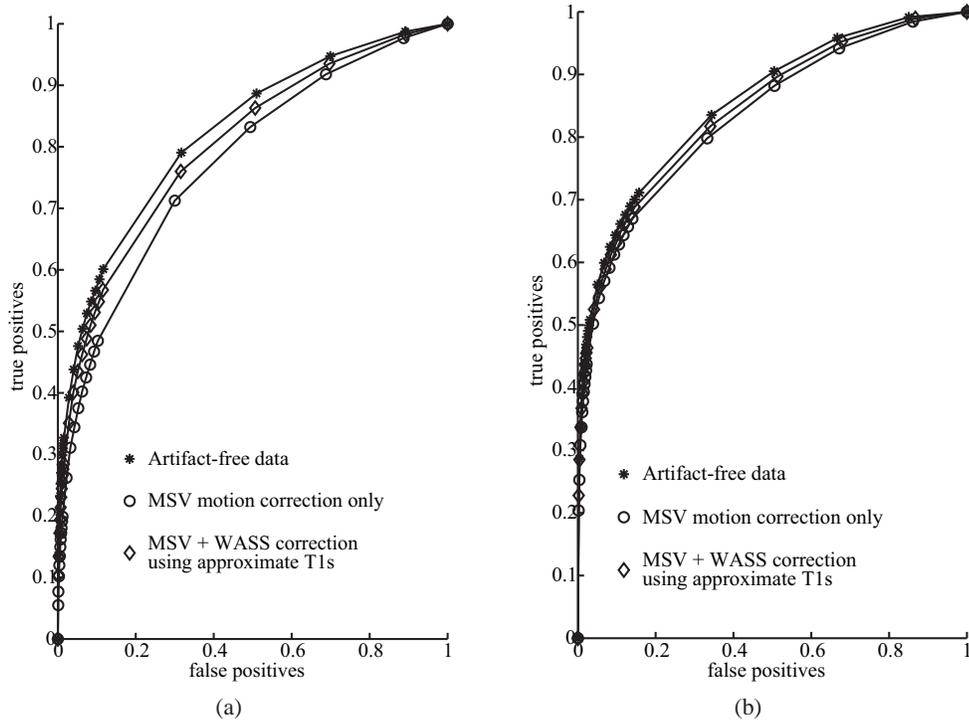
Figure 5.4: ROC curves following hypothesis testing for two simulated time series with induced head rotations of (a)$\pm 5°$ and (b)$\pm 2°$. ROC curves after MSV + WASS correction ($\diamond$) show an improvement in activation detection over MSV only correction ($\circ$). Artifact-free data ROC curves ($*$) are included for comparison. The WASS correction used MSV motion estimates and approximate tissue T1 values.

	Artifact-free (noisy)	Only MSV Motion Correction	MSV + WASS Correction		
			Exact T1s	Approx T1s	Single T1
$\pm 5^o$	0.822	0.774	0.800	0.802 (17.4%)	0.803
$\pm 2^o$	0.855	0.834	0.845	0.845 (12.2%)	0.844

Table 5.4: Approximate AUC values representative of activation detection for the noisy simulated fMRI time-series following MSV motion and WASS correction. Corresponding ROC curves are shown in Fig. 5.4.

ROC curves describe the effect of WASS correction on activation detection over the

entire time series. Given the small volume of the manually introduced activation pattern, these curves (and their approximate AUC values) may not be sensitive enough to capture the subtle effects of spin saturation artifacts and their correction. Hence, activation detection over the brain volume may be better depicted by activation maps.

Fig. 5.5 shows activation maps (alpha = 0.001) for two representative slices from both observed time series before and after WASS correction. For the time series with $\pm 5°$ head motion, Fig. 5.5(a) indicates that WASS correction using approximate $T1_{GM}$, $T1_{WM}$ and $T1_{CSF}$ values improved activation detection over that using a single (exact) $T1_{GM}$ value for some activation patterns. In contrast, there appears to be little perceptible difference in activation detection following WASS correction with approximate T1s or a single T1 for the time series with smaller motion (Fig. 5.5(b)). This is consistent with our earlier observation concerning residual average % intensity errors after WASS correction in Sec. 5.5.1.

5.6 Discussion

We have developed a spin saturation artifact correction method for fMRI time series, dubbed the WASS correction. Its performance was evaluated using ROC curves, activation maps and simulated fMRI time series data. In contrast to commonly used volume-to-volume fMRI registration, we employ an MI based slice-to-volume registration algorithm viz. MSV. MSV rigid motion estimates are used to map each EPI slice onto a high resolution T1-weighted volume obtained from the same subject. The EPI to T1 mapping is used to obtain information about the tissue composition of each low resolution EPI voxel. The WASS correction uses this knowledge of percentage contributions of WM, GM and CSF to the EPI voxel intensity, to compensate it for possible spin saturation effects.

While a full-blown validation of WASS correction using real fMRI data would have

(a)

(b)
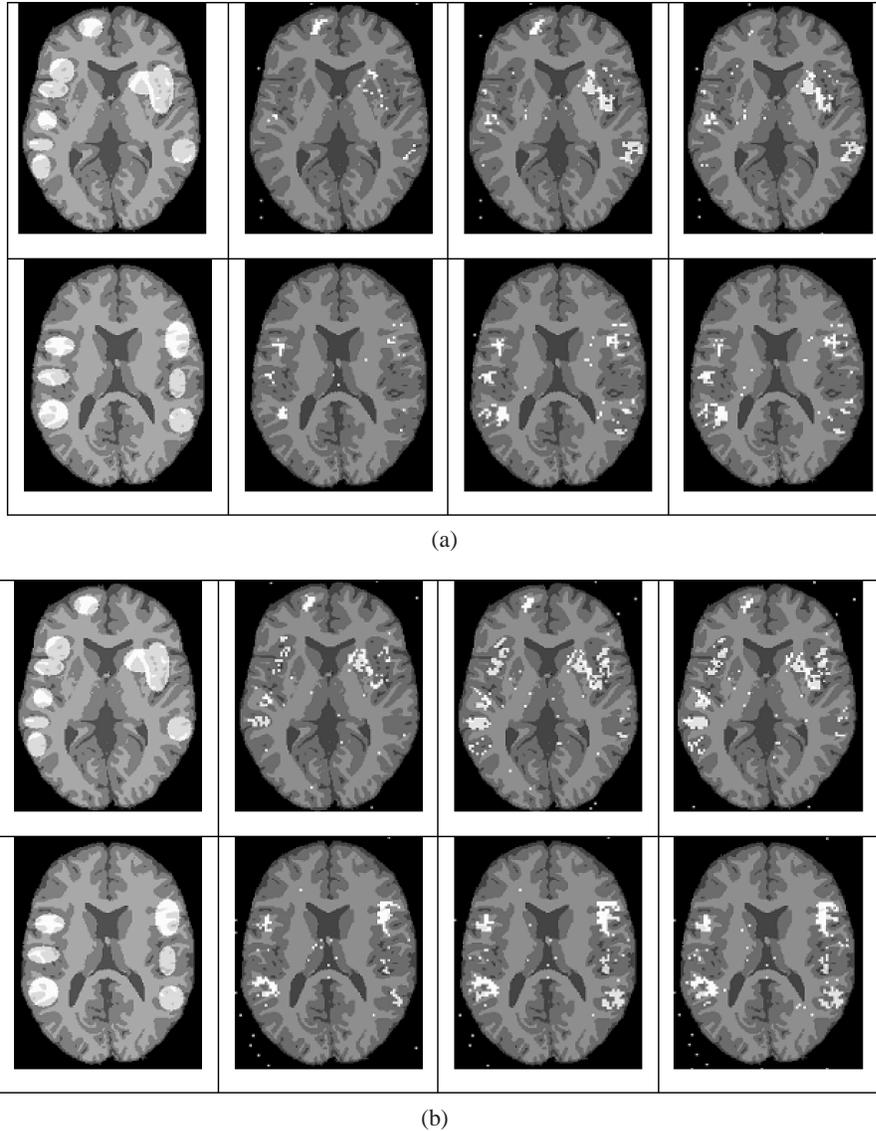
Figure 5.5: Sample activation maps (alpha = 0.001) for two simulated time series with (a)$\pm5°$ and (b)$\pm2°$ motion before and after WASS Corrections. For comparison, two sample slices are shown from (first column) the manually applied true activation pattern, corresponding activation maps after (second column) Only MSV Motion Correction, (third column) MSV + WASS Correction with approximate T1 values and (fourth column) MSV + WASS Correction with a single (exact) T1 value.

been ideal, establishing ground truth for a real fMRI time series can be a challenging task. This evaluation used simulated time series with mathematically induced head motion and related spin saturation artifacts, for which corresponding artifact-free data was readily obtainable. To loosely simulate a real world scenario, inherent GM, WM and CSF mis-

classifications, approximate T1 values and approximate proton density values were used in the MSV-based WASS correction. These approximations and inaccuracies were representative of clinical human data to some degree. However, the slice profile in both the simulation and the WASS correction was assumed to be rectangular. As modelled by (5.3) and (5.6), spin saturation artifacts depend on the cosine of the effective flip angle α. Thus, errors in approximations of the actual slice profile may further affect the performance of WASS correction.

Lastly, apart from its dependence on MSV motion estimates, the WASS correction for EPI slice n_{epi}, is largely a function of the approximate factors $\hat{f}_{ss}^{wa}(.)$ and $\hat{f}_{n_{\text{epi}}}^{wa}(.)$ in (5.15). These approximations improve with an increase in the resolution of the anatomical volume with respect to the EPI volumes.

CHAPTER 6

Improved fMRI Time-series Registration Using Joint Probability Density Priors

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. [21, 23] 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 [38]. 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 [63, 72, 77]. 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.

6.1 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 [50] and MSV [38] algorithms respectively. Our implementation used a GD optimizer, instead of the Nelder-Mead simplex optimizer used in [38, 50], 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 [70].

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

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

(6.1)

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 [16, 70]. 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. (6.1).

6.1.1 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) [9]. 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 [5,55]. 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.

6.1.2   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:

$$(6.2) \qquad \text{RMS error} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left\|T_{\theta_s^*}(x_i^s) - T_{\theta_s}(x_i^s)\right\|^2},$$

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. 6.1 (slow motion) and Fig. 6.2 (fast motion).

(a) Center-slices

(b) End-slices
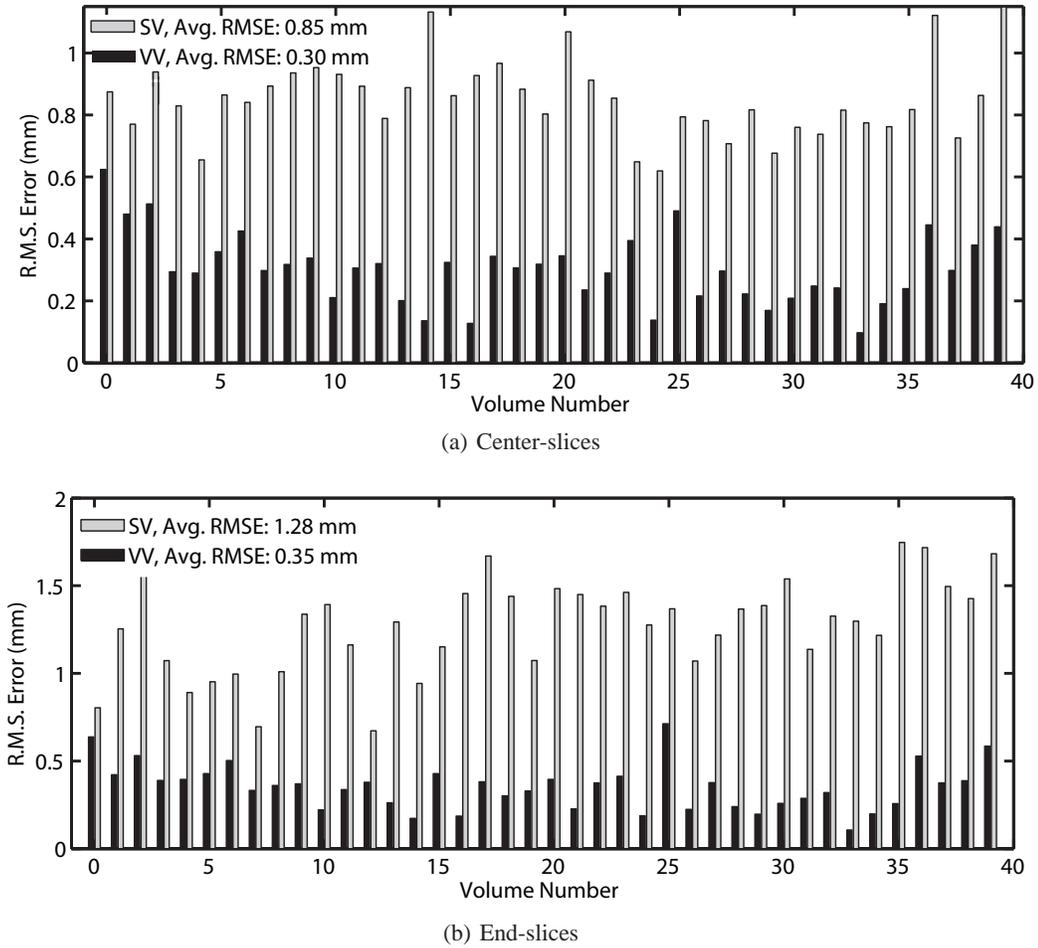
Figure 6.1: Comparison of RMS errors in time-series rigid motion estimates for simulated slow head motion.

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
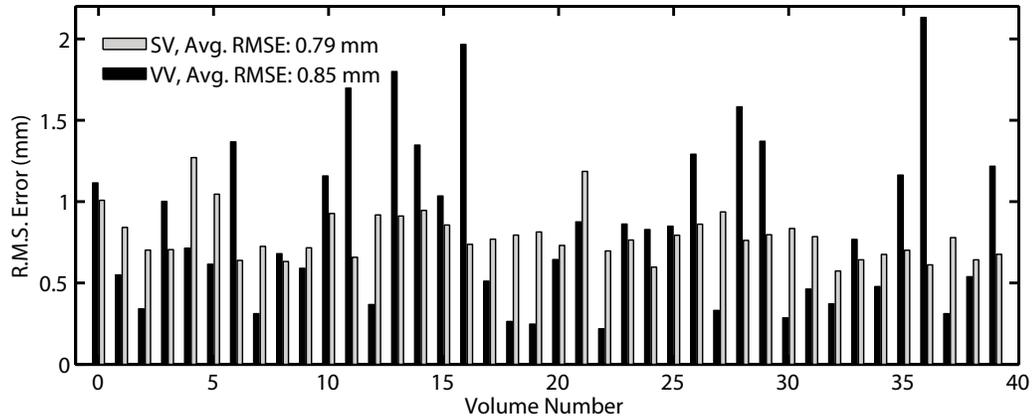
(a) Center-slices
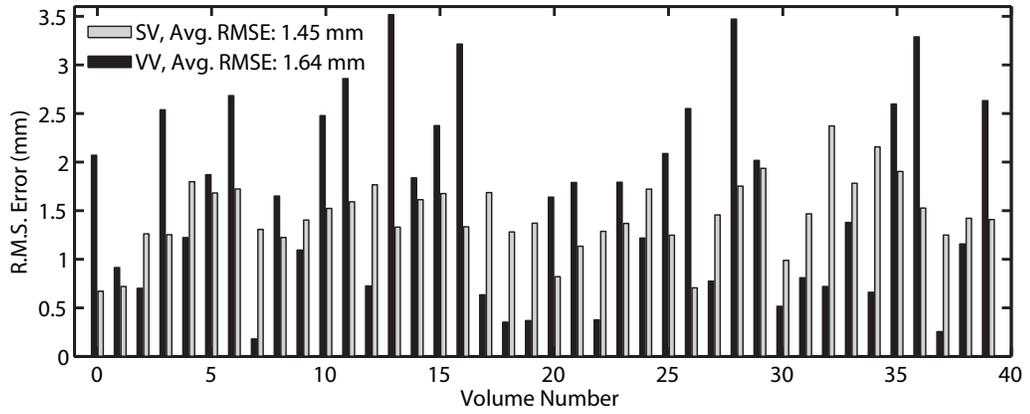
(b) End-slices

Figure 6.2: Comparison of RMS errors in time-series rigid motion estimates for simulated fast head motion.

slow rate of change of head position. This coupled with the reliability of the MI approx-imation 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 os-cillate 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.

6.2 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.

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 [56]. 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 [6] 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.

6.2.1 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 datasets to improve the accuracy of registration of new images from similar modalities have recently been proposed [62, 72, 77].

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:

$$(6.3) \qquad P(\{d_{kl}\}_{k=1,l=1}^{K,L}) = M! \prod_{k=1,l=1}^{K,L} \frac{(P_{uv}^{kl})^{d_{kl}}}{d_{kl}!},$$

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. [72] 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. [77] 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

$$(6.4) \qquad P\big(\{P_{uv}^{kl}\}_{k=1,l=1}^{K,L}; \{\alpha_{kl}\}_{k=1,l=1}^{K,L}\big) = \Gamma\bigg(\sum_{k,l} \alpha_{kl}\bigg) \prod_{k=1,l=1}^{K,L} \frac{(P_{uv}^{kl})^{\alpha_{kl}-1}}{\Gamma(\alpha_{kl})}$$

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 (6.3) and (6.4). 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.

6.2.2 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. 6.1.2 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. 6.3 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.

<div align="center">(a) Center-slice   (b) End-slice</div>
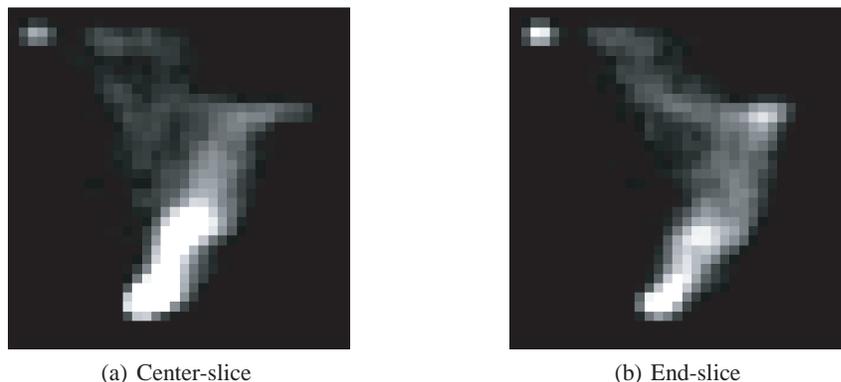
Figure 6.3: Comparison of the estimated joint pdf using intensity counts from a center-slice and an end-slice at registration.

For SV-JP, we propose to replace the joint pdf $\hat{P}_{uv}(g_k, h_l; \theta_s)$ in (6.1) by:

$$(6.5) \qquad \tilde{P}_{uv}(g_k, h_l; \theta_s, \beta) = (1 - \beta)\hat{P}_{uv}(g_k, h_l; \theta_s) + \beta P_{uv}^*(g_k, h_l);$$

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.

In contrast to registration algorithms that jointly estimate constrained motion parameters for all (or a sub-set of) time-series slices, such as JMSV [56], 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.

6.3 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. 6.1.2. 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 $P^*_{uv}(g_k, h_l)$; $g_k \in \{g_m\}^K_{m=1}$, $h_l \in \{h_n\}^L_{n=1}$ 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 $\{T_{\theta_S}(x^S_i)\}^M_{i=1}$ with ground truth coordinates $\{T_{\theta^*_S}(x^S_i)\}^M_{i=1}$. 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 6.1 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 6.1: 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.

6.3.1 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 [77]. The estimate of rigid motion parameters obtained using SV-JP registration for fMRI slice s, $\tilde{\theta}_s$ is given by:

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

where, $\tilde{H}_u$ was dropped as it is independent of the registration parameter θ_s. Using (6.5) to split $\tilde{P}_{uv}(g_k, h_l; \theta_s, \beta)$ the entropy terms above can be written as

(6.7)

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

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

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

Using (6.7) 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

$$
\begin{aligned}
\tilde{\theta}_s &= \arg\max_{\theta_s} \left\{ (1-\beta)\hat{\Phi}_{\text{MI}}(\theta_s) - \beta\sum_{l=1}^{L} P_v^*(h_l) D_{\text{KL}}\left(P_{u|v}^*(.|h_l)\|\tilde{P}_{u|v}(.|h_l; \theta_s, \beta)\right) \right. \\
(6.8) \qquad &\left. -(1-\beta)\sum_{l=1}^{L} \hat{P}_v(h_l; \theta_s) D_{\text{KL}}\left(\hat{P}_{u|v}(.|h_l; \theta_s)\|\tilde{P}_{u|v}(.|h_l; \theta_s, \beta)\right) \right\};
\end{aligned}
$$

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 $P^*_{u|v}$. That is, we expect SV-JP registration to penalize local minima solutions that result in large MI values but yield an undesired conditional pdf estimate.

Results from the 30 repeated registration runs for each time-series end-slice were also used to empirically study the occurrence of local minima in the SV and SV-JP similarity metrics. Figure 6.4 shows statistics of the RMS errors associated with rigid motion estimates corresponding to the ten largest similarity metric values for both methods. For ease of comparison, time-series volumes are indexed in order of increasing median RMS errors obtained using SV-JP. The box-plots show that for most datasets SV-JP registration results in better rigid motion estimates more often than SV registration; indicating that, as hypothesized, SV-JP is less likely to get caught in local minima than SV.

6.4   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
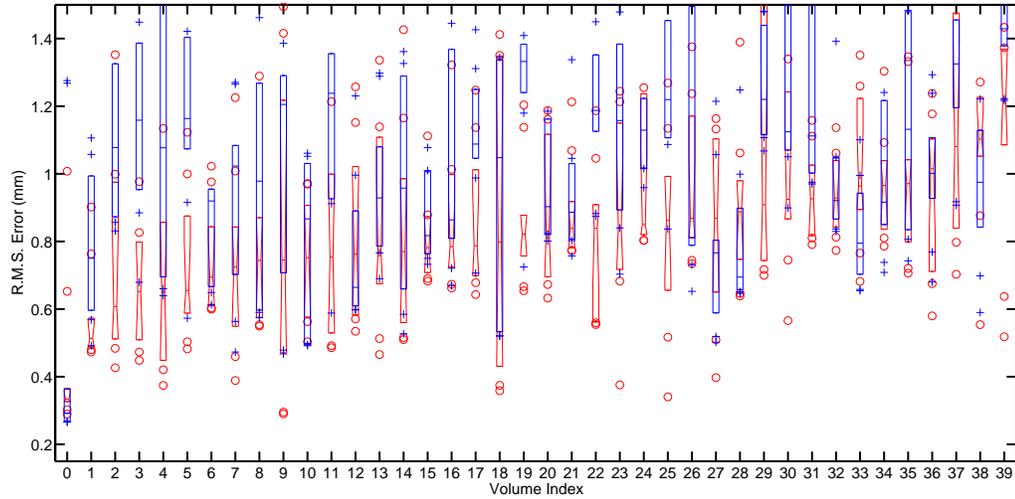
(a) Slow head motion
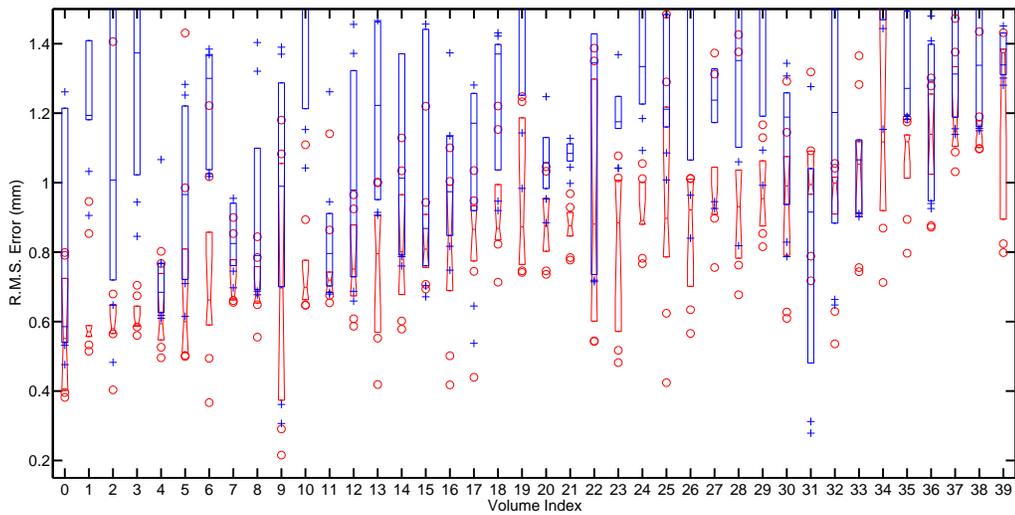
(b) Fast head motion

Figure 6.4: Comparison of the RMS errors of the 10 best rigid motion estimates for time-series end-slices obtained using SV-JP (red/notched) and SV (blue/plain) registration. The line at the center of each boxplot shows the median RMS error value and top and bottom edges are the 75 and 25 percent quantile RMS errors. The smallest and the largest RMS errors are shown by (o) for SV-JP and by (+) for SV. In general SV-JP results in lower RMS errors more often than SV. RMS errors for both time-series with (a) slow motion and (b) fast motion are shown.

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 6.1, 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.

CHAPTER 7

Summary and Future Work

7.1 Summary

This dissertation dealt with various ways of enhancing intensity-based registration methods. Intensity-based similarity metrics (such as, mutual information (MI) [10, 74], (negative of) sum of squared differences (SSD)) and gradient optimization methods are commonly used in nonrigid registration algorithms [49, 70]. Computing the gradient of these similarity metrics with respect to the large number of warp parameters is often the bottleneck of the algorithm. One approach to reduce this computational cost is to use a small random subset of image voxels to approximate the gradient [40, 49]. In Chapter 3 we developed a novel framework based on importance sampling (IS) to accelerate such nonrigid intensity-based registration schemes, by efficiently improving the accuracy and reducing the variance of these gradient approximations. This work leveraged the influence of image edges on intensity-based similarity metrics, to design a sampling strategy that encouraged sampling from these regions. We also investigated the use of two types of stochastic approximation (SA) methods, viz. sample-size controlled SA and step-size controlled SA, in conjunction with the importance sampling strategy. Results using B-spline warps to register simulated brain data and real CT lung data show that our framework can accelerate nonrigid registration while preserving accuracy. However, for real brain data, both IS-

SA and US-SA registration strategies showed comparable speed and accuracy. This may have been due to the larger number of edges in the brain volumes or because the initial deformations in the brain datasets were smaller than those in the CT lung data.

In Chapter 6 we turned to a special class of MI-based registration algorithms that register a 2D slice to a 3D volume, namely slice-to-volume (SV) registration. These algorithms are used to estimate motion in fMRI time-series when significant inter-slice subject head motion is expected [38]. However the MI approximation in SV registration is based on intensity counts from a single time-series slice. Hence, the registration is less accurate at time-series end-slices (i.e., slices near the top of the head scan), where the MI approximation is noisy due to low image complexity. We developed an improved SV registration algorithm, dubbed SV-JP, by using joint pdf priors derived from successfully registered time-series center-slices (i.e., slices near the middle of the head scan) to bolster the noisy MI approximation. We compared SV, SV-JP and volume-to-volume (VV) registration head motion estimates for end-slices from two simulated time-series with 'fast' (avg. speed = 1.35 mm/sec) and 'slow' (avg. speed = 0.14 mm/sec) head motion. SV-JP was more accurate then SV for both time-series. For slow head motion VV registration performed better than both SV and SV-JP registration while for fast head motion SV-JP registration was the most accurate.

Given the popularity of MI-based registration, in Appendix A we developed an efficient linear approximation of the covariance of registration motion estimates obtained by completely maximizing a differentiable plug-in estimate of MI. This approximation was based on results for M-estimates in [19, 65]. The performance of MI-based registration algorithms may be strongly affected by the choice of tuning parameters (e.g., width of the kernel in kernel density pdf estimates) used to implement the plug-in MI estimate. The covariance approximation, if satisfactory, could be used to help find suitable tuning param-

eter values to improve registration performance. However our approximation performed well only for a simple 2D mono-modality registration used to find a single translation parameter. For more realistic multi-modality registration the MI estimate is strongly non-linear, greatly degrading the accuracy of the linear covariance approximation.

Lastly, in Chapter 5 we digressed from motion estimation algorithms to address extraneous motion-dependant intensity-modulation, i.e. spin saturation artifacts, in fMRI time-series data. These artifacts reduce fSNR and can hamper brain-activation detection using fMRI time-series. We described spin saturation artifacts using mathematical expressions and developed a weighted-average spin saturation (WASS) correction technique. An algorithm to identify and correct fMRI voxels affected by spin saturation was outlined. In contrast to existing spin saturation correction methods [28,52], WASS correction takes into account the approximate tissue composition (percentage contributions of white-matter, gray-matter and CSF) of each low resolution time-series voxel. Results on simulated data showed that WASS correction can improve brain-activation detection using fMRI time-series.

7.2 Future Work

- The data used to evaluate the performance of IS-SA registration in Chapter 3 has few or sparse edges. For data with more edges, it may be beneficial for IS-SA to use a more stringent criterion to include a smaller percentage of image edges in the sampling distribution in (3.9). For a population of images, it may be possible to empirically determine the percentage of edges that need to be retained to obtain a particular level of registration accuracy.

- Highly uniform point-sets (HUPS) were used in [69] to improve the performance of uniform sampling based registration. It would be useful to investigate whether a

similar strategy, i.e., transforming HUPS to obtain samples following the distribution designed in (3.9), can improve the performance of IS-SA.

- Another possibility is to use domain specific knowledge to choose 'important regions' - for example, in subject-to-atlas registrations, structures with larger shape variability may be harder to align. Hence, in such cases, drawing a larger percentage of samples from these regions may benefit IS-SA.

- In Chapter 3 importance sampling (IS) was used to improve the gradient of differentiable intensity-based similarity metrics, such as MI, SSD. Since mutual information (MI) is an expectation, IS could also be used to improve the estimate of MI itself. Such an approach would also facilitate the use of IS with non-gradient optimization approaches, e.g., Nelder-Mead optimization.

- In our MI metric implementation (Sec. 3.3.1), the number of uniformly spaced intensity levels K and L at which the joint and marginal kernel density pdf estimates were evaluated was proportional to the number of intensity pairs used to approximate the (gradient of) MI. That is, K and L were larger for the deterministic GD registration scheme (based on all available intensity pairs) and smaller for the US-SA and IS-SA registration schemes (based on a small random subset of intensity pairs). However, this empirical approach may not yield the best values of K and L. Investigating an approach to choose the optimal number of possibly non-uniformly spaced levels K and L to estimate the pdfs will be valuable.

- At IS-SA iteration k, our gradient approximation uses i.i.d. random samples drawn from an adaptive sampling distribution $P_s^{\theta_k}$, given by (3.9). That is, $P_s^{\theta_k}$ changes slowly over the iterations with variations in θ_k. In our implementation, the random samples drawn at iteration $k+1$ are completely different from those used at iteration k. By treating the registration process as a slowly varying dynamic system, it may

be possible to adapt the procedures in [46] to 'reuse' random samples obtained at iteration k to construct random samples at iteration $k+1$, making the IS-SA algorithm more efficient.

- The performance of the weighted-average spin saturation (WASS) correction algorithm in Chapter 5 was evaluated on simulated data. Though our simulations tried to incorporate inaccuracies representative of clinical data (such as white-matter, gray-matter and CSF mis-classifications, approximate T1 values), it will be desirable to test the algorithm using controlled phantom studies and real data.

- As the WASS correction is a function of the approximate factors $\hat{f}_{ss}^{\text{wa}}(.)$ and $\hat{f}_{n_{\text{epi}}}^{\text{wa}}(.)$ in (5.15), its accuracy should increase with an increase in the ratio of the resolution of the anatomical volume to that of the EPI volumes. Hence it will be useful to the study the effect of changes in this ratio on the performance of the WASS correction. Specifically, establishing a lower limit on this ratio (below which WASS correction is ineffective) may facilitate the development of acquisition protocols required to enable successful post-processing of EPI time-series data.

- In Chapter 5 we use motion estimates obtained from EPI time-series to T1 anatomical volume registration to estimate the percentage contribution of white-matter, gray-matter and CSF in each EPI voxel. This makes the method susceptible to registration errors. Source separation methods (such as Independent Component Analysis) may be another *registration-independent* approach to estimate the tissue composition of low resolution EPI voxels.

- For SV-JP registration in Chapter 6, β was chosen such that the pdf estimate $\tilde{P}_{uv}$ was obtained by pooling in all the (averaged) intensity counts used to build the pdf prior and the intensity counts from the to-be-registered time-series slice. However, in the formulation in (6.5), β is a tuning parameter of the SV-JP method. Hence it

may be instructive to study the effect of varying β on SV-JP registration accuracy for different head motion speeds.

- Our evaluation of SV-JP used simulated time-series data. A full validation using real EPI time-series from a variety of fMRI stimulus studies (e.g., motor tasks, verbal tasks) will be valuable.

- Results on simulated data in Sec. 6.3 indicated that VV registration is more accurate than SV or SV-JP registration for slow motion, while SV-JP registration is well-suited to estimate faster head motion. In block stimulus fMRI studies it may be reasonable to expect slow head motion during the OFF or no stimulus periods and faster head motion during the ON periods. Hence, a scheme that incorporates such prior knowledge to estimate head motion using some combination of VV and SV or SV-JP registration may show improved accuracy.

- SV and SV-JP registration both estimate a completely independent rigid transform for each EPI time-series slice. However, in reality, head motion at adjacent slice-acquisition time-points may show some level of correlation. Hence, for a given type of fMRI study, it may be useful to characterize and study the (possible) correlation in true head motion for a population of subjects. Such a motion model can then be used to incorporate motion priors in the registration process.

- As discussed in Appendix A, our linear approximation of the (co)variance of MI-based registration estimates was unsatisfactory for realistic multi-modality registration, where, the MI metric becomes strongly nonlinear. Developing or extending such covariance approximations to handle nonlinear similarity metrics can aid in the analysis and improvement of popular registration methods.

APPENDIX

APPENDIX A

Approximate Covariance of MI-based Image Registration

Many registration methods that obtain estimates by completely maximizing a differentiable similarity metric are in effect finding an M-estimate. This appendix draws on the theory of M-estimates [34, 65] and on results in [19] to present a more theoretical method of approximating the (co)variance of image registration estimates obtained by maximizing a differentiable plug-in estimate of mutual information.

Consider registration between a reference and homologous image with noisy intensities $\tilde{u}_i = u(x_i) + n_i, i = 1, 2 \ldots N$ and $\tilde{v}_j = v(y_j) + n_j, j = 1, 2 \ldots M$, respectively. The estimate of parameters $\hat{\theta} \in \mathbb{R}^p$, defining a mapping $T_{\hat{\theta}}$ between the two images, is implicitly given by

(A.1) $$\hat{\theta} = \arg\max_{\theta} \Psi(\theta, \tilde{Z}), \quad \text{where } \tilde{Z} = \{\tilde{u}_i\}_1^N \bigcup \{\tilde{v}_j\}_1^M.$$

Implicitly defined estimates, such as $\hat{\theta}$ are called M-estimates in statistics. In some cases the implementation of $\Psi(.)$ employs user defined tuning parameters which affect the bias and variance properties of these M-estimates.

Information theoretic similarity metrics such as mutual information (MI) are widely used in intensity-based image registration. These metrics treat $\tilde{Z}$ as observations of pseudo-random variables U and V and estimate their probability density functions (pdfs). A pop-

ular differentiable approximation of MI between the two images is given by,

$$\Psi(\theta, \tilde{Z}) = -\sum_{k=1}^{K} \hat{P}_u(f_k; \tilde{Z}) \log(\hat{P}_u(f_k; \tilde{Z})) - \sum_{l=1}^{L} \hat{P}_v(g_l; \theta, \tilde{Z}) \log(\hat{P}_v(g_l; \theta, \tilde{Z}))$$

$$(A.2) \qquad + \sum_{l=1}^{L} \sum_{k=1}^{K} \hat{P}_{uv}(f_k, g_l; \theta, \tilde{Z}) \log(\hat{P}_{uv}(f_k, g_l; \theta, \tilde{Z})).$$

$\hat{P}_u$, $\hat{P}_v$ and $\hat{P}_{uv}$ are pdfs estimated using a differentiable density kernel $\kappa(.)$ [16, 57], i.e.

$$(A.3) \qquad \hat{P}_{uv}(f_k, g_l; \theta, \tilde{Z}) = \frac{1}{|\Omega_\theta| h^2} \sum_{i: \tilde{u}_i \in \Omega_\theta} \kappa\Big(\frac{\tilde{u}_i - f_k}{h}\Big) \kappa\Big(\frac{\hat{v}_i^\theta - g_l}{h}\Big).$$

Where, $\kappa(.)$ integrates to unity and has finite support $h, h > 0$, $\hat{v}_i^\theta \approx v(T_\theta(x_i))$ is some

differentiable function of $\{\tilde{v}_j\}_1^M$, as in (3.5) and Ω_θ is the region of overlap between the two

images at parameter guess θ. The sets $\{f_k\}_{k=1}^{K}$ and $\{g_l\}_{l=1}^{L}$ are fixed grayscale intensity

levels chosen to sample the joint pdf finely enough. The marginal pdfs are given by

$$(A.4) \qquad \hat{P}_v(g_l; \theta, \tilde{Z}) = \sum_{k=1}^{K} \hat{P}_{uv}(f_k, g_l; \theta, \tilde{Z}) \text{ and } \hat{P}_u(f_k; \tilde{Z}) = \sum_{l=1}^{L} \hat{P}_{uv}(f_k, g_l; \theta, \tilde{Z}).$$

The constant h in (A.3), is the user defined width of the kernel and strongly influences

the pdf estimates. A large h will result in very smooth pdfs and will possibly reduce the

sensitivity of $\Psi(\theta, \tilde{Z})$ to changes in θ. On the other hand a very small h may make the

metric unnecessarily sharp, introducing many local maxima. The 'optimal' choice of h

will depend on the level of noise in the observations $\tilde{Z}$, i.e. h should be large enough to

suppress noise in the pdf estimates but small enough to retain their significant features.

Thus the quality of the estimate $\hat{\theta}$, depends on the value of tuning parameter h.

For a given pair of imaging modalities, an obvious choice for h is to use values which

result in the most accurate registration at a given noise level. However obtaining a direct

measure of registration accuracy is not trivial. Usually registration accuracy is cited only

with respect to a large number of controlled phantom studies and other registration meth-

ods. Such empirical evaluations of image registration accuracy can be very time consum-

ing. However a reasonable choice of h can be efficiently estimated, if we can approximate

the variance (and bias) of the M-estimate $\hat{\theta}$.

We draw on the theory of M-estimates [19, 34, 65] to obtain a computationally efficient approximation for $\text{Cov}(\hat{\theta})$. For completeness we reproduce some of the results from [19] here. Let M-estimate $\hat{\theta} = w(\tilde{Z}) \in \mathbb{R}^p$, where $w(.)$ is some non-linear function. Expanding $w(\tilde{Z})$ about the mean $\bar{Z}$ of the noisy observations using the first order Taylor series yields,

$$(A.5) \qquad w(\tilde{Z}) \approx w(\bar{Z}) + \nabla w(\bar{Z}) \times (\tilde{Z} - \bar{Z}).$$

Where $\tilde{Z}$ is a column vector of $N + M$ noisy observations and $\nabla = [\frac{\partial}{\partial \tilde{Z}_1}, \ldots, \frac{\partial}{\partial \tilde{Z}_{N+M}}]$ is a row gradient operator. Taking the covariance of the above expression with respect to the noisy observations $\tilde{Z}$ gives

$$\text{Cov}(w(\tilde{Z})) \approx [\nabla w(\bar{Z})]\text{Cov}(\tilde{Z})[\nabla w(\bar{Z})]'.$$

However $w(.)$ is unknown, hence we need to estimate $\nabla w(\bar{Z})$. Considering only M-estimates obtained by completely maximizing differentiable similarity metrics, $\hat{\theta} \in \mathbb{R}^p$ must satisfy

$$\frac{\partial \Psi(\theta, \tilde{Z})}{\partial \theta_j}\Big|_{\theta=\hat{\theta}} = 0; \quad j = 1, 2 \ldots p.$$

Using $\hat{\theta} = w(\tilde{Z})$ and differentiating the above expression w.r.t. $\tilde{Z}$ gives

$$\sum_k \frac{\partial^2}{\partial \theta_j \partial \theta_k} \Psi(\theta, \tilde{Z}) \frac{\partial}{\partial \tilde{Z}_i} w_k(\tilde{Z}) + \frac{\partial^2}{\partial \theta_j \partial \tilde{Z}_i} \Psi(\theta, \tilde{Z}) = 0.$$

Putting $\tilde{Z} = \bar{Z}$ and expressing the above equation in matrix form,

$$\nabla^{20} \Psi(w(\bar{Z}), \bar{Z}) \nabla w(\bar{Z}) + \nabla^{11} \Psi(w(\bar{Z}), \bar{Z}) = 0.$$

Where ∇^{20} is the $p \times p$ Hessian of operator w.r.t θ, whose (j, k)th component is $\frac{\partial^2}{\partial \theta_j \partial \theta_k}$ and ∇^{11} is a $p \times (N + M)$ operator whose (j, i)th component is $\frac{\partial^2}{\partial \theta_j \partial \tilde{Z}_i}$. Assuming that $\nabla^{20} \Psi(w(\bar{Z}), \bar{Z})$ is invertible, $\nabla w(\bar{Z})$ is given by,

$$\nabla w(\bar{Z}) = [-\nabla^{20} \Psi(w(\bar{Z}), \bar{Z})]^{-1}[\nabla^{11} \Psi(w(\bar{Z}), \bar{Z})].$$

Letting $\check{\theta} = w(\bar{Z})$, the approximate covariance of our M-estimate $\hat{\theta}$ is

$$\text{(A.6)} \quad \text{Cov}(\hat{\theta}) \approx [-\nabla^{20}\Psi(\check{\theta},\bar{Z})]^{-1}[\nabla^{11}\Psi(\check{\theta},\bar{Z})]\text{Cov}(\tilde{Z})[\nabla^{11}\Psi(\check{\theta},\bar{Z})]'[-\nabla^{20}\Psi(\check{\theta},\bar{Z})]^{-1}.$$

To apply the covariance approximation to MI-based registration, (A.2) should be twice differentiable. However the region of overlap Ω_θ used in (A.3) changes abruptly with variations in θ, either including or excluding an integer number of voxels. Hence we introduce an smooth approximation to $|\Omega_\theta|$,

$$|\hat{\Omega}_\theta| = \sum_{i=1}^{N} \beta_\theta(i) \quad \text{and} \quad \beta_\theta(i) = \sum_{j=1}^{M} B(T_\theta(x_i)/\alpha - y_j/\alpha);$$

where α is chosen such that $y_j/\alpha \in I$. The function $B(.)$ is twice differentiable and satisfies the partition of unity property, i.e, $\sum_{k\in I} B(\eta - k) = 1, \forall \eta \in R$; we choose $B(.)$ to be the cubic B-spline. Using this approximation (A.3) is modified as,

$$\text{(A.7)} \qquad \hat{P}_{uv}(f_k, g_l; \theta, \tilde{Z}) = \frac{1}{h^2 \sum_{j=1}^{N} \beta_\theta(j)} \sum_{i=1}^{N} \beta_\theta(i)\kappa\Big(\frac{\tilde{u}_i - f_k}{h}\Big)\kappa\Big(\frac{\hat{v}_i^\theta - g_l}{h}\Big).$$

The approximate MI is computed by using these modified pdf estimates in (A.2).

Preliminary experiments used (A.6) to approximate the variance of 2D mono-modality registration. Registration was performed using two 256×256 T1 weighted MR images with 1×1 mm voxels. Both images were initially in registration. The homologous image was created by applying local cubic B-spline translations along (a) only the x-axis (max. 4.2 mm) and (b) only the y-axis (max. 3.2 mm). The local deformation was spread over an approximately 40×40 mm image region. Gaussian noise ($N(0, \sigma^2)$) with increasing σ from 0.5 to 3 in steps of 0.5 was added to both images. Translation estimates were obtained by maximizing the differentiable MI approximation using the Conjugate Gradient

(CG) optimizer. At each noise variance σ^2 level, sample variances of the estimated translations were computed using estimates from 75 registration runs. Each run registered a pair of noisy images, simulated using independent noise realizations drawn from $N(0, \sigma^2)$.

To compute the approximate variance in (A.6), $\breve{\theta} = \arg\max_\theta \Psi(\theta, \bar{Z})$ was estimated using $\bar{Z} = E(\tilde{Z})$. As noisy images were simulated using additive zero mean Gaussian noise, $\bar{Z}$ was simply the column vector of noiseless image intensities. Thus $\breve{\theta}$ was the M-estimate computed by MI-based registration between the pair of noiseless images using CG. Finally the matrices $[-\nabla^{20}\Psi(\breve{\theta}, \bar{Z})]^{-1}$ and $\nabla^{11}\Psi(\breve{\theta}, \bar{Z})$ were constructed by differentiating eqs. A.2, A.7 and A.4 and using the chain rule. This approximate (co)variance of registration M-estimates varies linearly with the covariance of noisy image intensities, $\text{Cov}(\tilde{Z})$.
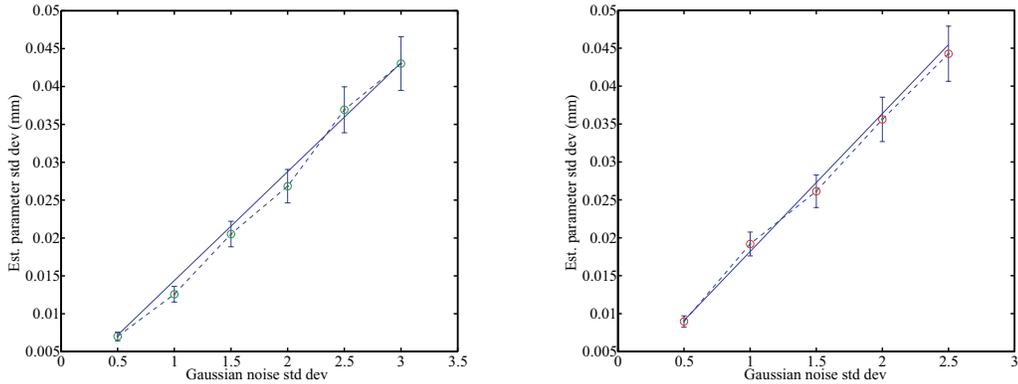
(a) B-spline based local translation along the x-axis, (max. 4.2 mm).

(b) B-spline based local translation along the y-axis, (max. 3.2 mm).

Figure A.1: Comparison of theoretical approximate standard deviation (SD) and empirical sample SD values. For each image noise level σ, the estimator sample SD ($\circ$) was calculated from 75 independent registration runs. The approximate SD (solid line) required a single registration run using noiseless data. Each mono-modality registration M-estimate was obtained by maximizing approximate MI using CG.

Fig. A.1 shows how the variance approximation compares with empirically calculated sample variance values. For the simple case of mono-modality registration with local deformations and Gaussian noise, the variance approximation provides a good line fit to the

sample variance values. The approximate variance method requires a single registration to estimate $\breve{\theta}$. This is very efficient when compared to the completely empirical method which requires a large number of registration runs to obtain a satisfactory estimate of $\text{cov}(\hat{\theta})$ at each noise variance level.

However for larger noise variances and more realistic multi-modal image registration, the MI similarity metric becomes strongly non-linear. In such cases the performance of the proposed linear variance approximation degrades significantly. It may be possible to remedy this by extending the Taylor series in A.5 to include second order terms, making the covariance approximation quadratic. However this requires estimation of higher order derivatives of the similarity metric which may not be very reliable.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] V. Arsigny, X. Pennec, and N. Ayache. Polyrigid and polyaffine transformations: A new class of diffeomorphisms for locally rigid or affine registration. In *Medical Image Computing and Computer-Assisted Intervention*, volume LNCS 2879, pages 829–37, 2003.

[2] R. Bhagalia, J. A. Fessler, and B. Kim. Gradient based image registration using importance sampling. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 446–9, 2006.

[3] R. Bhagalia and B. Kim. Spin saturation artifact correction using slice-to-volume registration motion estimates for fMRI time series. *Med. Phys.*, 35(2):424–34, February 2008.

[4] F. L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Patt. Anal. Mach. Int.*, 11(6):567–87, June 1989.

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

[6] A. G. Chandler, T. Netsch, C. A. Cocosco, J. A. Schnabel, and D. J. Hawkes. Slice-to-volume registration using mutual information between probabilistic image classifications. In J. M. Fitzpatrick and M. Sonka, editors, *Medical Imaging 2004: Image Processing. Edited by Fitzpatrick, J. Michael; Sonka, Milan. Proceedings of the SPIE, Volume 5370, pp. 1120-1129 (2004).*, volume 5370 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 1120–1129, May 2004.

[7] Y. Choi and . S. Lee. Local injectivity conditions of 2D and 3D uniform cubic B-spline functions. In *Proc Computer Graphics and Applications*, pages 302–11, 1999.

[8] S. Clare. *Functional MRI : Methods and Applications*. PhD thesis, University of Nottingham, 1997.

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

[10] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration using information theory. In Y. Bizais, C. Barillot, and R. D. Paola, editors, *Information Processing in Medical Im.*, pages 263–74. Kluwer, Dordrect, 1995.

[11] M. M. Coselmon, J. M. Balter, D. L. McShan, and M. L. Kessler. Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines. *Med. Phys.*, 31(11):2942–8, November 2004.

[12] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.

[13] W. Crum, T. Hartkens, and D. Hill. Non-rigid image registration: theory and practice. *The British Journal of Radiology*, 77:140–153, 2004.

[14] R. Cusack, M. Brett, and K. Osswald. An evaluation of the use of magnetic field-maps to undistort echo-planar images. *NeuroImage*, 18(1):127–42, January 2003.

[15] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley-Iterscience, 2001.

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

[17] P. Dupis and R. Simha. On sampling-controlled stochastic approximation. *IEEE Trans. Automated Control*, 36(8):915–24, August 1991.

[18] S. Faro and F. Mohamed, editors. *Functional MRI: Basic Principles and Clinical Applications*. Springer Science, 2006.

[19] J. A. Fessler. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography. *IEEE Trans. Im. Proc.*, 5(3):493–506, March 1996.

[20] J. A. Fessler. On NUFFT-based gridding for non-Cartesian MRI. *J. Mag. Res.*, 188(2):191–5, October 2007.

[21] L. Freire and J-F. Mangin. Motion correction algorithms of the brain mapping community create spurious functional activations. In R M Leahy M F Insana, editor, *Information Processing in Medical Im.*, pages 246–48. Springer, Berlin, 2001.

[22] L. Freire and J.-F. Mangin. Two stage alignment of fMRI time series using the experiment profile to discard activation-related bias. *Proc. 5th MICCAI*, pages 663–670, 2002.

[23] 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.*, 21(5):470–84, May 2002.

[24] K. Friston, J. Ashburner, C. Frith, J. Poline, J. Heather, and R. Frackowiak. Spatial registration and normalization of images. *Human Brain Mapping*, 2:165–189, 1995.

[25] K. Friston, A. Holmes, J. Poline, J. Grasby, S. Williams, R. Frackowiak, and R. Turner. Analysis of fmri time-series revisited. *Neuroimage*, 2:45–53, 1995.

[26] K. Friston, A. Holmes, K. Worsley, J. Poline, C. Frith, and R. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210, 1995.

[27] K. Friston, P. Jezzard, and R. Turner. Analysis of functional mri time-series. *Human Brain Map*, 1:153–171, 1994.

[28] K. Friston, S. Williams, R. Howard, R. Frackowiak, and R. Turner. Movement-related effects in fmri time-series. *Magnetic Resonance in Medicine*, 35:346–355, 1996.

[29] K.J. Friston. Introduction: Experimental design and statistical parametric mapping. In R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, K.J. Friston, C.J. Price, S. Zeki, J. Ashburner, and W.D. Penny, editors, *Human Brain Function*. Academic Press, 2nd edition, 2003.

[30] E. M. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan. *Magnetic resonance imaging: Physical principles and sequence design*. Wiley, New York, 1999.

[31] A. Hero, B. Ma, O. Michel, and J.D. Gorman. Applications of entropic spanning graphs. *IEEE Signal Proc. Mag.*, 19(5):85–95, 2002.

[32] W. S. Hinshaw and A. H. Lent. An introduction to NMR imaging: From the Bloch equation to the imaging equation. *proc-ieee*, 71(3):338–50, March 1983.

[33] A. P. Holmes, R. C. Blair, J. D. G. Watson, and I. Ford. Nonparametric analysis of statistic images from functional mapping experiments. *J. Cerebral Blood Flow and Metabolism*, 16(1):7–22, January 1996.

[34] P. J. Huber. *Robust statistics*. Wiley, New York, 1981.

[35] S. Huettel, A. Song, and G. McCarthy. *Functional Magnetic Resonance Imaging*. Sinauer Associates, Inc., 2004.

[36] H. Kesten. Accelerated stochastic approximation. *Ann. Math. Stat.*, 29(1):41–59, March 1958.

[37] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Ann. Math. Stat.*, 23(3):462–6, September 1952.

[38] B. Kim, J. L. Boes, P. H. Bland, T. L. Chenevert, and C. R. Meyer. Motion correction in fMRI via registration of individual slices into an anatomical volume. *Mag. Res. Med.*, 41(4):964–72, May 1999.

[39] B. Kim, J. L. Boes, K. A. Frey, and C. R. Meyer. Mutual information for automated multimodal image warping. *Visualization in Biomedical Computing in Lecture Notes in Computer Science*, 1131:349–354, 1996.

[40] S. Klein, M. Staring, and J. P. Pluim. Comparison of gradient approximation techniques for optimisation of mutual information in nonrigid registration. In *Proc. SPIE 5747, Medical Imaging 2005: Image Proc.*, pages 192–203, 2005.

[41] S. Koonin. *Computational Physics*. Addison-Wesley, California, 1986.

[42] H. Kushner and T. Gavin. Extensions of Kesten's adaptive stochastic approximation method. *The Annals. of Statistics*, 1(5):851–61, September 1973.

[43] J. Kybic, P. Thevenaz, A. Nirkko, and M. Unser. Unwarping of unidirectionally distorted EPI images. *IEEE Trans. Med. Imag.*, 19(2):80–93, February 2000.

[44] H. Lester and S. R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–49, January 1999.

[45] X. Liang, W. Zeng, Z. Dong, Z. Zhang, and L. Luo. Non-uniform mr image reconstruction based on non-uniform fft. *27th International Congress on High-Speed Photography and Photonics*, 6279(1):62793D–1–62793D–2, 2007.

[46] J. S. Liu and R. Chen. Sequential Monte Carlo Methods for Dynamic Systems. *Journal of the American Statistical Association*, 93(443):1032–44, September 1998.

[47] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.*, 16(2):187–98, April 1997.

[48] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Med. Im. Anal.*, 2(1):1–36, March 1998.

[49] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank. PET-CT image registration in the chest using free-form deformations. *IEEE Trans. Med. Imag.*, 22(1):120–8, January 2003.

[50] C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, and R. L. Wahl. 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(3):195–206, April 1997.

[51] P. Munger, G. R. Crelier, T. M. Peters, and G. B. Pike. An inverse problem approach to the correction of distortion in EPI images. *IEEE Trans. Med. Imag.*, 19(7):681–9, July 2000.

[52] L. Muresan, R. Renken, J. BT. M. Roerdink, and H. Duifhuis. Automated correction of spin-history related motion artefacts in fMRI: simulated and phantom data. *IEEE Trans. Biomed. Engin.*, 52(8):1450–60, August 2005.

[53] J. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[54] T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Hum. Brain Map.*, 15(1):1–25, 2002.

[55] D. G. Nishimura. Principles of magnetic resonance imaging, 1996. Unpublished textbook.

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

[57] E. Parzen. On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

[58] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Trans. Med. Imag.*, 22(8):986–1004, August 2003.

[59] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Stat.*, 22(3):400–7, September 1951.

[60] V. Roopchansingh, R. W. Cox, A. Jesmanowicz, B. D. Ward, and J. S. Hyde. Single-shot magnetic field mapping embedded in echo-planar time-course imaging. *Mag. Res. Med.*, 50(4):839–43, October 2003.

[61] D. Ruan, J. A. Fessler, M. Roberson, J. Balter, and M. Kessler. Nonrigid registration using regularization that accommodates local tissue rigidity. In *Proc. SPIE 6144, Medical Imaging 2006: Image Proc.*, page 614412, 2006.

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

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

[64] D. W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.

[65] R. J. Serfling. *Approximation theorems of mathematical statistics*. Wiley, New York, 1980.

[66] J. C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Auto. Control*, 37(3):332–41, March 1992.

[67] M. Staring, S. Klein, and J. P. Pluim. Nonrigid registration with adaptive, content-based filtering of the deformation field. In *Proc. SPIE 5747, Medical Imaging 2005: Image Proc.*, pages 212–21, 2005.

[68] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, January 1999.

[69] P. Thevenaz, M. Bierlaire, and M. Unser. Halton sampling for image registration based on mutual information, 2006. Sampling Theory in Signal and Image Processing (in press).

[70] P. Thevenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Trans. Im. Proc.*, 9(12):2083–99, December 2000.

[71] K. Thulborn, J. Waterton, P. Matthews, and G. Radda. Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochimica et biophysica acta*, 714(2):265–70, February 1982.

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

[73] M. Unser, A. Aldroubi, and M. Eden. B-spline signal processing: Part II—efficient design and applications. *IEEE Trans. Sig. Proc.*, 41(2):834–48, February 1993.

[74] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Med. Im. Anal.*, 1(1):35–51, March 1996.

[75] R. P. Woods, S. R. Cherry, and J. C. Mazziotta. Rapid automated algorithm for aligning and reslicing PET images. *J. Comp. Assisted Tomo.*, 16(4):620–33, July 1992.

[76] D. A. Yoder, Y. Zhao, C. B. Paschal, and J. M. Fitzpatrick. MRI simulator with object-specific field map calculations. *Mag. Res. Im.*, 22(3):315–28, April 2004.

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