

# Intensity-Based Image Registration Using Robust Correlation Coefficients

Jeongtae Kim\* and Jeffrey A. Fessler, *Senior Member, IEEE*

**Abstract**—The ordinary sample correlation coefficient is a popular similarity measure for aligning images from the same or similar modalities. However, this measure can be sensitive to the presence of “outlier” objects that appear in one image but not the other, such as surgical instruments, the patient table, etc., which can lead to biased registrations. This paper describes an intensity-based image registration technique that uses a *robust* correlation coefficient as a similarity measure. Relative to the ordinary sample correlation coefficient, the proposed similarity measure reduces the influence of outliers. We also compared the performance of the proposed method with the mutual information-based method. The robust correlation-based method should be useful for image registration in radiotherapy (KeV to MeV X-ray images) and image-guided surgery applications. We have investigated the properties of the proposed method by theoretical analysis, computer simulations, a phantom experiment, and with functional magnetic resonance imaging data.

**Index Terms**—Image registration, mutual information, outlier, robust correlation coefficient, robustness.

## I. INTRODUCTION

**I**MAGE registration is a useful technique for aiding diagnosis, performing patient set-up estimation for radiation therapy [1] and for image-guided surgery [2], [3], etc. For the set-up estimation problem, a preoperative image [usually a computed tomography (CT) volume] is transformed geometrically to align with measured radiographs. Intensity-based registration methods work by maximizing a similarity measure based on the intensity values of the two images. Therefore, designing an effective similarity measure is very important. This paper proposes a robust similarity measure for intramodality image registration.

One fundamental design criterion is that a similarity measure should be maximized at the true registered position in the absence of noise. Establishing this characteristic analytically is challenging since the behavior of the objective function depends on the nature of the images being registered. Another important criteria is the statistical *efficiency* of the registration method, i.e., the variability that would result from repeating the registration

with identical images except for noise. In addition, registration methods can differ in their *robustness* to the presence of unexpected objects in images.

Many intensity-based image registration methods implicitly treat the intensity pairs taken from corresponding spatial locations in two images as independent identically distributed (i.i.d.) samples of two random variables. With that assumption, statistical concepts such as correlation, joint entropy and mutual information (MI) are used as similarity measures by estimating those statistical properties from the i.i.d. samples.

The correlation coefficient is a particularly popular similarity measure, and is a natural choice when registering two images from the same modality [4], [5]. Although correlation is poor similarity measure for multimodality image registration, in terms of statistical efficiency and computational efficiency, the correlation coefficient is one of the best similarity measures for intramodality image registration. Since image registration for set-up estimation in radiation therapy and image-guided surgery often involves images from the same (or similar) modality, the correlation coefficient can be useful for those applications.

The sample correlation coefficient has been used widely to estimate the correlation coefficient due to its simplicity. However, a drawback of the sample correlation coefficient is its sensitivity to outliers [6, p. 199]. Even a few outliers can affect the sample correlation coefficient greatly and, thus, degrade image registration performance. A significant number of “outliers” may be present in the image-guided surgery application due to the presence of operating instruments and in the radiation therapy application due to the effect of radiotherapy table [1]. For X-ray CT images, differences in contrast agents also occur. Although a bias in estimating the correlation coefficient need not directly imply a bias in image registration, we have observed such biases empirically when outliers are present [7].

The MI similarity measure is used widely for multimodality image registration since it does not assume any functional relationship between the two image values [8]–[10]. In this sense, the MI method has an inherent degree of robustness. However, as illustrated by our empirical results in Section III and analyses in the Appendices, for intramodality image registration, the robustness of the MI method depends on the particular images being registered. Moreover, the MI method can be statistically inefficient, i.e., the registration variability due to noise can exceed that of the sample correlation coefficient.

To overcome the drawbacks of the sample correlation method and the MI method, we have investigated an image registration method that uses robust correlation coefficients [6, p. 204] as a similarity measure, thereby improving the robustness without compromising the statistical efficiency much.

Manuscript received July 2, 2003; revised January 30, 2004. This work was supported in part by the National Institutes of Health (NIH) under Grant P01-CA59827, Grant R01-CA81161, and Grant R01-CA60711. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. H. Lowe. *Asterisk indicates corresponding author.*

\*J. Kim is with the Information Electronics Department, Ewha Womans University, Seoul 120-750, Korea (e-mail: jtkim@ewha.ac.kr).

J. A. Fessler is with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: fessler@umich.edu).

Digital Object Identifier 10.1109/TMI.2004.835313

Robust estimation of mean and covariance has been studied extensively in statistics [6], [11]–[13]. The basic idea of robust estimation is to weight the measured samples in a way that reduces the effect of outlier samples, or even removes them completely.

For example, one may compute the statistical distance of each sample value from the mean, the *Mahalanobis distance*, evaluate a weighting function based on that distance, and determine a new weighted mean and covariance and iterate until convergence [12]. Alternatively, one may estimate the probability density function (pdf) after trimming out the outliers by determining the minimum volume ellipsoidal pdf [13] or minimum covering ellipsoidal pdf [11]. A robust mean and covariance may then be estimated from the estimated pdf.

There were also several investigations that applied robust statistics for image registration. For example, robust similarity measures such as weighted square error and nonquadratic error were applied for registering MRI and other modality images [14], [15]. The performance of such robust methods were compared with other methods empirically [16].

In this study, we focused on  $M$ -estimation methods for robust correlation estimation [6, p. 211]. This framework helps to explain why the sample correlation coefficient is sensitive to outliers and provides insight into how to design a method with improved robustness. Moreover, we can explain some properties of the MI method within the same framework.

Investigating the advantages and disadvantages of using alternative robust correlation estimation methods for image registration in the presence of outliers is deferred to future research. Such a study should consider the sample distributions of representative images in a particular context.

The paper is organized as follows. Section II reviews the image registration problem and describes the proposed similarity measure. Section III compares the image registration accuracies of the proposed robust correlation similarity measure, the conventional sample correlation coefficient and the MI similarity measure. Three comparisons are reported: one-dimensional simulation, two-dimensional (2-D) functional magnetic resonance imaging (fMRI) image registration, and three-dimensional (3-D)/2-D registration of an X-ray CT volume to orthogonal radiographs of an anthropomorphic chest phantom.

The appendices present analyses of the statistical properties of the correlation-based and MI-based image registration methods by approximating the mean and the variance using first-order Taylor series expansions [17]. Since image registration is highly nonlinear and the objective function is an implicit function of the images, it is challenging to obtain concise and insightful results from such approximations. Nevertheless, we summarize some theoretical arguments that complement the empirical results.

## II. THEORY: SIMILARITY MEASURES

The goal of image registration is to find a geometric transformation (rigid or nonrigid), denoted  $T$ , that aligns two given

images, denoted  $s_1(\vec{t})$  and  $s_2(\vec{t})$ , where  $\vec{t}$  denotes the spatial coordinates<sup>1</sup>. Intensity-based image registration methods achieve this goal by maximizing a similarity measure based on the image intensity values. If we parameterize the transformation  $T$  using  $\theta$  (e.g., three translation and three rotation parameters for rigid transformation), the image registration becomes a parameter estimation problem

$$\hat{\theta} = \arg \max_{\theta} \Phi(s_1(T_{\theta}(\cdot)), s_2(\cdot)) \quad (1)$$

where  $\Phi(s_1, s_2)$  is some measure of the *similarity* between the images  $s_1$  and  $s_2$ . In practice, registration is performed using finite number of samples  $X_i$  and  $Y_i$  as follows:

$$\begin{aligned} X_i &= s_1(T_{\theta}(\vec{t}_i)) \\ Y_i &= s_2(\vec{t}_i), \quad i = 1, \dots, N \end{aligned} \quad (2)$$

where  $\{\vec{t}_i\}$  denotes the sample locations, and where  $s_1(T_{\theta}(\cdot))$  denotes a spatially transformed (and interpolated) version of  $s_1(\vec{t})$ . Since  $X_i$  depends on the parameter  $\theta$ , all statistical quantities computed using  $X_i$  are functions of  $\theta$ . However, for simplicity of notation, often we leave this dependence implicit.

Considering the sampling, a more precise expression for the registration problem is

$$\hat{\theta} = \arg \max_{\theta} \Phi(\mathbf{X}(\theta), \mathbf{Y}) \quad (3)$$

where  $\mathbf{X} = (X_1, \dots, X_N)$  and  $\mathbf{Y} = (Y_1, \dots, Y_N)$ .

A variety of similarity measures  $\Phi$  have been proposed for image registration. Many of these are statistical quantities such as the correlation coefficient, joint entropy and MI. For such metrics, there is an underlying assumption that the  $(X_i, Y_i)$  pairs are i.i.d. samples of jointly distributed random variables with some (unknown) joint pdf  $f_{XY}(x, y)$ . This i.i.d. assumption is somewhat artificial, but nevertheless leads to useful similarity measures so we continue in this tradition in this paper.

Next we review the usual correlation coefficient similarity measure and contrast it with our proposed robust correlation coefficient approach.

### A. Correlation Coefficient Estimates

For two random variables  $X$  and  $Y$  with joint pdf  $f_{XY}(x, y)$ , the correlation coefficient  $\rho(X, Y)$  is defined as follows:

$$\rho(X, Y) \triangleq \frac{C(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}} \quad (4)$$

where the covariance is

$$C(X, Y) \triangleq \int (x - E[X])(y - E[Y])f_{XY}(x, y) dx dy \quad (5)$$

and where  $E[X]$  and  $E[Y]$  denote the means, and  $\sigma_X^2$  and  $\sigma_Y^2$  denote the variances of the random variables. Note that (under mild regularity conditions on  $f_{XY}(x, y)$ ), the correlation coefficient  $\rho$  is well-defined even if the pdf  $f_{XY}(x, y)$  is not parameterized in terms of it.

Given  $N$  i.i.d. sample pairs  $(X_i, Y_i)$ , there are several ways to estimate the correlation coefficient  $\rho$ . For example, one

<sup>1</sup>We focus on image-to-image or volume-to-volume registration, but the general approach applies equally to volume-to-projection registration, e.g., [1]. Also, we treat  $s_1(\vec{t})$  and  $s_2(\vec{t})$  as continuous-space functions in the presentation; in practice, sampling and interpolation are essential [9].

could first use the samples to compute an estimated joint pdf  $\hat{f}_{XY}(x, y)$ , and then substitute  $\hat{f}_{XY}(x, y)$  into (4) to estimate  $\rho$ . The following subsections describe two other approaches.

1) *Sample Correlation Coefficient*: Perhaps the most popular way to estimate  $\rho$  is the sample correlation coefficient, defined as follows:

$$\hat{\rho}_s(\mathbf{X}, \mathbf{Y}) = \frac{\hat{C}_s(\mathbf{X}, \mathbf{Y})}{\sqrt{\hat{\sigma}_X^2 \hat{\sigma}_Y^2}} \quad (6)$$

where the sample means, sample variances, and sample covariance are defined in the usual way as follows:

$$\begin{aligned} \hat{C}_s(\mathbf{X}, \mathbf{Y}) &\triangleq \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}) \\ \bar{X} &\triangleq \frac{1}{N} \sum_{i=1}^N X_i \\ \bar{Y} &\triangleq \frac{1}{N} \sum_{i=1}^N Y_i \\ \hat{\sigma}_X^2 &\triangleq \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 \\ \hat{\sigma}_Y^2 &\triangleq \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2. \end{aligned} \quad (7)$$

Relative to alternative methods for estimating the correlation coefficient, the sample correlation method has the advantage of simplicity since  $\hat{\rho}_s(\mathbf{X}, \mathbf{Y})$  is an *explicit* function of the data samples  $(\mathbf{X}, \mathbf{Y})$ . Furthermore, on the surface it appears not to require any specific model for the joint pdf  $f_{XY}(x, y)$ . A minor drawback is that  $\hat{\rho}_s(\mathbf{X}, \mathbf{Y})$  is not unbiased, even for normal distributions although it is asymptotically unbiased in that case [18, p. 90]. More importantly,  $\hat{\rho}_s(\mathbf{X}, \mathbf{Y})$  is not robust to outliers [6, p. 199], as explained intuitively in the next subsection.

2) *Maximum Likelihood Estimates of  $\rho$* : An alternative approach to estimating  $\rho$  is the following: a) hypothesize a parametric form for the joint distribution  $f_{XY}(x, y)$  of  $X$  and  $Y$  that depends on  $\rho$ ; b) find the maximum likelihood (ML) estimate of the parameters given the data  $\mathbf{X} = \mathbf{X}(\theta)$  and  $\mathbf{Y}$ ; and c) determine  $\rho = \rho(\theta)$  from the ML parameter estimates. Usually we will have to compute the ML estimates numerically, so the estimator  $\hat{\rho}$  will be an *implicit* function of  $X$  and  $Y$ . To our knowledge, previous similarity measures used in image registration have all been *explicit* functions of the two images, so the proposed approach departs from that convention.

We propose to adapt the spirit of this ML approach for the purpose of robust image registration. However, since the joint pdf  $f_{XY}(x, y)$  is a somewhat artificial concept in this context, we do not expect to model the pdf  $f_{XY}(x, y)$  precisely. We consider the model for  $f_{XY}(x, y)$  to be a function chosen by the algorithm designer to impart desirable properties on the resulting estimates, such as robustness.

Let  $\mathbf{Z}_i = (X_i, Y_i)$  denote the pairs of corresponding image intensity values. As usual, we consider the  $\mathbf{Z}_i$ 's to be i.i.d. samples of a two-dimensional random vector  $\mathbf{Z} = (X, Y)$ . Following Huber [6, p. 211], we construct a model for the pdf of  $\mathbf{Z}$  having elliptical contours as follows. First, we choose a non-negative function  $f_0$  for which the corresponding 2-D circularly

symmetric density  $f_0(\|\mathbf{z}\|)$  integrates to unity over  $\mathbb{R}^2$ , where  $\|\mathbf{z}\| = \sqrt{x^2 + y^2}$ . Then we consider an (unknown) nondegenerate transformation  $\mathbf{z} \mapsto \mathbf{V}(\mathbf{z} - \boldsymbol{\mu})$  that leads to the following density:

$$f_Z(\mathbf{z}; \boldsymbol{\mu}, \mathbf{V}) = |\det \mathbf{V}| f_0(\|\mathbf{V}(\mathbf{z} - \boldsymbol{\mu})\|). \quad (8)$$

Under this parametric model,  $\boldsymbol{\mu}$  denotes the mean of  $\mathbf{Z}$  and the  $2 \times 2$  covariance matrix of  $\mathbf{Z}$  is  $(\mathbf{V}^T \mathbf{V})^{-1}$ . In other words,  $\mathbf{V} = (\text{Cov}\{\mathbf{Z}\})^{-1/2}$ . For example, if one were to choose

$$f_0(r) = \frac{1}{2\pi} e^{-r^2/2} \quad (9)$$

then (8) would become the bivariate normal distribution.

Having chosen some  $f_0$ , one may estimate the mean  $\boldsymbol{\mu}$  and the covariance term  $\mathbf{V}$  from the sample pairs  $\mathbf{Z}_i$  in the spirit of ML estimation as follows:

$$(\hat{\boldsymbol{\mu}}, \hat{\mathbf{V}}) = \arg \max_{(\boldsymbol{\mu}, \mathbf{V})} \prod_{i=1}^N |\det \mathbf{V}| f_0(\|\mathbf{V}(\mathbf{Z}_i - \boldsymbol{\mu})\|). \quad (10)$$

Usually, there is no closed-form expression for the estimates  $\hat{\boldsymbol{\mu}}$  and  $\hat{\mathbf{V}}$ , so (10) is an *implicit* definition.

In classical estimation theory, ideally  $f_0$  would be chosen so that the pdf  $f_Z$  in (8) agrees with the actual distribution of the  $\mathbf{Z}_i$ 's. However, since the notion that the  $\mathbf{Z}_i$ 's are i.i.d. is somewhat artificial in the context of image registration, it is more useful to think of  $f_0$  as a user-selectable function.

To help understand the "ML estimates" (10), we differentiate (10) with respect to  $\boldsymbol{\mu}$  and  $\mathbf{V}$ . Zeroing these expressions yields the following two necessary conditions [6, p. 212]:

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^N w(\|\hat{\mathbf{V}}(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}\|) \mathbf{Z}_i}{\sum_{i=1}^N w(\|\hat{\mathbf{V}}(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}\|)} \quad (11)$$

$$(\hat{\mathbf{V}}^T \hat{\mathbf{V}})^{-1} = \sum_{i=1}^N w(\|\hat{\mathbf{V}}(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}\|) (\mathbf{Z}_i - \hat{\boldsymbol{\mu}})(\mathbf{Z}_i - \hat{\boldsymbol{\mu}})^T \quad (12)$$

where  $w(\cdot)$  denotes the following *weighting function*:

$$w(r) \triangleq -\frac{f_0'(r)}{r f_0(r)}. \quad (13)$$

Huber proposed a simple iterative procedure for solving these two nonlinear equations to obtain the ML estimates [6, p. 215].

After estimating the covariance term  $\hat{\mathbf{V}}$ , we can extract an estimate  $\hat{\rho}$  of the correlation coefficient as follows:

$$\begin{aligned} (\hat{\mathbf{V}}^T \hat{\mathbf{V}})^{-1} &= \begin{bmatrix} \hat{\sigma}_X^2 & \hat{C}_{XY} \\ \hat{C}_{XY} & \hat{\sigma}_Y^2 \end{bmatrix} \\ \hat{\rho} &= \frac{\hat{C}_{XY}}{\sqrt{\hat{\sigma}_X^2 \hat{\sigma}_Y^2}}, \end{aligned} \quad (14)$$

where  $\hat{\sigma}_X^2$  and  $\hat{\sigma}_Y^2$  denote the ML estimated variances of  $X$  and  $Y$  under the assumed model (8).

After finding the estimates  $\hat{\boldsymbol{\mu}}$  and  $\hat{\mathbf{V}}$ , one can compute the weighting function  $w(\|\hat{\mathbf{V}}(\mathbf{z} - \hat{\boldsymbol{\mu}}\|)$  to examine the relative influence of different data values  $\mathbf{z}$  on the estimates. (See Fig. 3(b) for an example.) The weighting function is plotted in the joint pdf domain for illustrative purposes only. In practice, one estimates the robust correlation coefficient by solving (11) and (12) *without* estimating the joint pdf explicitly.

If we were to choose (9) for  $f_0$ , then the weighting function in (13) simplifies to  $w(r) = 1$ . In this special case, there is an explicit solution for the ML estimates:  $\hat{\boldsymbol{\mu}}$  is simply the sample mean of the  $\mathbf{Z}_i$ 's, and  $\hat{\mathbf{V}}$  is the square root of the inverse of the sample covariance of the  $\mathbf{Z}_i$ 's. This corresponds to the well-known result that the sample correlation coefficient is the ML estimator for  $\rho$  under a bivariate normal density. However, using constant weighting  $w(r) = 1$  means that *all* data points are weighted equally, even outliers. As a result, the sample mean, sample covariance, and sample correlation coefficient are all sensitive to outliers [6].

### B. Robust Correlation Coefficient

The nonrobustness of the sample correlation coefficient can be explained by considering that it is the ML estimator for an assumed normal distribution, which has “light tails” so outliers are modeled as being extremely unlikely. Conversely, if an ML estimate is based on a model distribution that has “heavy tails,” then data values that are far from the mean have less effect on the likelihood function. Thus, to design a robust estimator, we choose a model pdf  $f_Z$  that has “heavy tails.” One choice would be the Laplacian distribution, which would correspond to using  $f_0(r) = ce^{-|r|}$  for some constant  $c$ . However, that  $f_0$  is not differentiable at  $r = 0$ , so the expression (13) is inapplicable. Instead, we have chosen the following function:

$$f_0(r) = ce^{-\delta^2 [\sqrt{1+r^2/\delta^2}-1]} \quad (15)$$

where  $c$  is the constant that ensures that (8) integrates to unity. The constant  $\delta > 0$  is a design parameter. For small  $\delta$ , this model approaches the Laplacian distribution, and for large  $\delta$ , it approaches the normal distribution.

For the choice (15), the weighting function  $w(r)$  in (13) becomes the following:

$$w(r) = \frac{1}{\sqrt{1 + \frac{r^2}{\delta^2}}}. \quad (16)$$

Unlike with the normal choice (9), for this model the weighting of a given data point  $w(\|\hat{\mathbf{V}}(\mathbf{z} - \hat{\boldsymbol{\mu}})\|)$  will decrease with increasing statistical distance from the mean  $\hat{\boldsymbol{\mu}}$ . This has the desirable effect of reducing the influence of outliers that are, by definition, data points that are far from the mean.

We can make no claim of optimality of the choice (15). Indeed, the optimal choice would depend on the actual “distribution” of the  $\mathbf{Z}_i$ 's, which is unknown in practice. The function  $f_0$  is simply a design parameter for our robust estimator. For example, one could try to increase robustness relative to (16) by using the following weighting:

$$w(r) = \frac{1}{\sqrt{1 + \frac{r^4}{\delta^4}}}. \quad (17)$$

One may use many different weighting functions to compute robust correlation coefficients. Although it may be challenging to compare relative robustness among different weighting functions, any underlying pdf with heavier tails than a normal distribution should improve robustness relative to the conventional sample correlation coefficient.

Huber's algorithm for solving (11) and (12) is the following iteration:

$$\hat{\boldsymbol{\mu}}_{n+1} = \hat{\boldsymbol{\mu}}_n + \frac{\sum_{i=1}^N w(\|\hat{\mathbf{V}}_n(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}_n)\|) (\mathbf{Z}_i - \hat{\boldsymbol{\mu}}_n)}{\sum_{i=1}^N w(\|\hat{\mathbf{V}}_n(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}_n)\|)} \quad (18)$$

$$\begin{aligned} & (\hat{\mathbf{V}}_{n+1} \hat{\mathbf{V}}_{n+1})^{-1} \\ &= \sum_{i=1}^N w(\|\hat{\mathbf{V}}_n(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}_n)\|) (\mathbf{Z}_i - \hat{\boldsymbol{\mu}}_n)(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}_n)^T. \end{aligned} \quad (19)$$

We initialize  $\hat{\boldsymbol{\mu}}_0$  using the sample median and  $\hat{\mathbf{V}}_0$  using the matrix square root of the inverse of the sample covariance of the  $\mathbf{Z}_i$ 's. We then continue to iterate until two stopping criteria are satisfied:  $\|\hat{\mathbf{V}}_{n+1}^{-1} \hat{\mathbf{V}}_n - \mathbf{I}\| < \delta_V$  and  $\|\hat{\mathbf{V}}_{n+1}(\hat{\boldsymbol{\mu}}_{n+1} - \hat{\boldsymbol{\mu}}_n)\| < \delta_\mu$ . User-specified constants  $\delta_V$  and  $\delta_\mu$  control the precision of solutions.

Unfortunately, Huber's algorithm for solving (11) and (12) has been proven to converge only when estimating *one* of the two parameters, i.e.,  $\boldsymbol{\mu}$  or  $\mathbf{V}$ , but not necessarily both [6, p. 237]. However, in practice, the algorithm converged every time in our simulations and experiments.

In summary, our robust registration method works as follows. For the similarity measure  $\Phi$  described in (3), we propose to use the robust correlation coefficient

$$\Phi(\mathbf{X}(\theta), \mathbf{Y}) = \hat{\rho}(\mathbf{X}(\theta), \mathbf{Y}) \quad (20)$$

where  $\hat{\rho}$  was defined in (14) for the weighting function defined in (16) or (17). To maximize  $\Phi$  with respect to  $\theta$ , one must use some type of search algorithm such as the simplex method [19]. For each trial value of the registration parameter  $\theta$ , one must compute  $\mathbf{X}(\theta)$  by interpolation, and then compute  $\hat{\rho}$  by applying Huber's algorithm. So there are “iterations within iterations” in this approach; fortunately, the inner iteration converges quite quickly.

### C. Mutual Information

Another similarity measure that has robust characteristics is *mutual information* (MI). MI is a measure of the statistical dependence between two random variables. The MI  $I(X, Y)$  is defined in terms of marginal and joint entropies as follows:

$$\begin{aligned} h(X) &= - \int f_X(x) \log f_X(x) dx \\ h(X, Y) &= - \int f_{XY}(x, y) \log f_{XY}(x, y) dx dy \\ I(X, Y) &= h(X) + h(Y) - h(X, Y). \end{aligned} \quad (21)$$

Usually, MI is estimated by first estimating the joint pdf  $f_{XY}(x, y)$ , and then computing the MI using (21). Two popular pdf estimation methods are the kernel density approach<sup>2</sup>[20] and the histogram approach. Kernel density estimates are smooth and differentiable but can require considerable computation. Histogram estimates are usually faster to compute but yield pdfs that are discontinuous functions of the registration parameter  $\theta$ . As a practical compromise, we have used the interpolated joint histogram method in which the effect of a sample is distributed to four adjacent histogram bins using

<sup>2</sup>This is often called the *Parzen window* method.

bilinear interpolation; this approach remedies the discontinuity problem of the histogram with modest computation. Methods for estimating MI directly from the samples without first estimating a pdf are also under development [21], [22].

#### D. Analytical Comparisons

Most previous studies of image registration methods have focused on empirical comparisons. The appendices of this paper describe approximate analyses of the statistical properties of image registration methods using mean and variance approximations presented in Appendix A. We used these approximations because exact analytical expressions are unavailable since the estimator for image registration is defined implicitly as the maximizer of an objective function. Even if an analytical expression were available, finding exact expressions for the mean and variance would still remain difficult since the estimator  $\hat{\theta}$  is a nonlinear function of the images.

Our approximation method uses a first-order Taylor series expansion of the estimator about the mean data, an approach that has been used successfully for image reconstruction problems [17]. By comparison, the estimators used for image registration are more nonlinear, but we proceed with linearization nevertheless, hoping for insights.

For simplicity we focus on the asymptotic case as the number of image samples increases to infinity, i.e., the images are continuous-space functions. In addition, if there are no outliers, we assume that two images being registered are the same images except for the geometric transformation and additive white Gaussian noise.

The main points of the analyses can be summarized as follows. First, if there are no outliers, the sample correlation-based estimator is unbiased and is the most efficient (i.e., the estimator that has the smallest variance) among unbiased estimators. We argued that the sample correlation-based estimator is unbiased since the sample correlation coefficient without noise is maximized at *true* registered position<sup>3</sup>. Regarding efficiency, the sample correlation-based estimator is the most efficient one among unbiased estimators since it is the maximum-likelihood estimation (MLE) [23], if there are no outliers. The variance approximation of the sample correlation-based estimator matches the *Cramer-Rao* bound (see Appendix C). However, the sample correlation-based estimator can be significantly biased due to outliers as argued in Appendix B. In summary, the sample correlation-based estimator has good properties such as unbiasedness and efficiency, but poor robustness to outliers.

For the robust correlation-based estimator, if there are no outliers, we argue that it is also unbiased using the similar approximation as for the sample correlation-based estimator (see Appendix B). Moreover, even in the presence of outlier samples, the robust correlation-based estimator is more robust than the sample correlation-based estimator since the effects of the outlier samples are reduced by smaller weighting (see Appendix B). However, for data without outliers, the variance of the robust

correlation-based estimator is larger than the sample correlation-based estimator as argued using Cauchy-Schwarz inequality [24] (see Appendix C).

For the MI-based estimator, it is not easy to argue analytically that the estimator is unbiased even if there are no outliers. Instead, we explain qualitatively why the MI is maximized at the *true* registered position. At the *true* registered position, the estimated joint pdf is the most clustered along the  $Y = X$  line in the joint pdf domain, thereby making the values of the joint pdf around the  $Y = X$  line maxima and values elsewhere minima. Therefore, the gradient of the estimated joint pdf is zero almost everywhere, which implies that the MI is maximized since the gradient of the MI is zero. Nevertheless, due to the smoothing effect of kernel function,<sup>4</sup> it is difficult to prove the unbiasedness analytically except for some special cases [25].

We also argue qualitatively that the MI-based estimator has a certain degree of robustness. Suppose that the portion of the estimated pdf from inliers is the most clustered at the true registered position but that from outliers is not the most clustered.<sup>5</sup> In that case, the estimated MI is maximized at a position where the entire pdf is the most clustered in average sense. Since the portion of the pdf from inliers is usually much larger than that from outliers, the entire pdf is likely to be the most clustered around the true position. Nevertheless, since the robustness of the MI method is due to the behavior of the joint pdf rather than explicitly reducing the influence of outliers, the robustness can depend greatly on images being registered.

Another interesting perspective is to express MI as a generalized weighted correlation coefficient as follows:

$$I(X, Y) = \int (x - E[X])(y - E[Y]) w_I(x, y) \hat{f}_{XY}(x, y) dx dy \quad (22)$$

where  $w_I(x, y) = \log \hat{f}_{Y|X}(y|x) / (x - E[X])(y - E[Y])$  is a kind of generalized weighting. The robustness characteristics and statistical efficiency of MI can depend on the images being registered since the weighting depends on the estimated joint pdf. For example, if the estimated pdf is a normal distribution, the weighting is almost constant and the MI method is almost the same as the sample correlation method. In that case, the MI method can be very efficient, like the sample correlation method. In fact, joint entropy and the correlation coefficient have a one-to-one relationship for bivariate normal distributions [26]. For most intramodality images with nonconstant  $w_I(x, y)$ , the MI method is less efficient than the sample correlation coefficient for Gaussian noise.

In summary, both the robust correlation method and the MI method have improved robustness but larger variance compared to the sample correlation method, when noise is Gaussian. Direct analytical comparisons of the statistical properties of the robust correlation method and the MI method are difficult since the properties depend not only on design parameters (such as the underlying pdf for robust correlation method and the pdf estimation method for MI method), but also on the images being registered.

<sup>4</sup>We consider the kernel density approach since the histogram approach is not differentiable.

<sup>5</sup>Apparently, if pdf from outliers is also the most clustered at true position, there should be no bias due to outliers.

<sup>3</sup>In our approximation, if the gradient of an objective function without noise is zero at *true* registered position, it is unbiased estimator (see Appendix A).

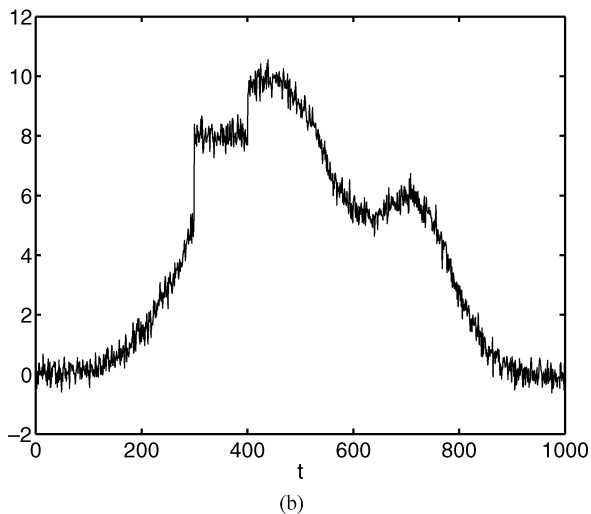
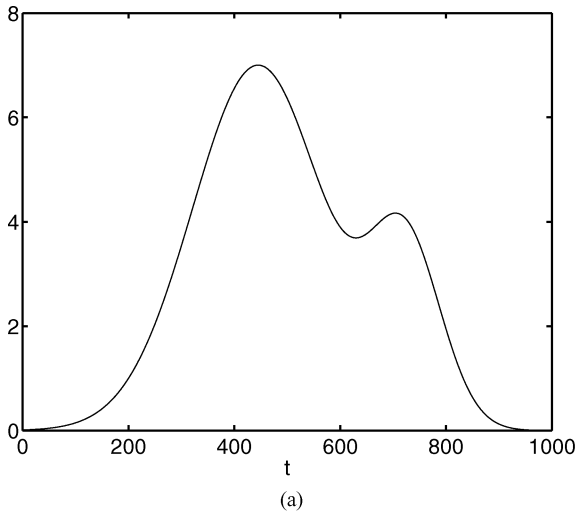


Fig. 1. Reference signal  $s_1(\tilde{t})$  (signal being translated to achieve registration) and observed signal  $s_2(\tilde{t})$  with outliers and Gaussian noise (STD = 0.3). (a)  $s_1(\tilde{t})$ . (b)  $s_2(\tilde{t})$ .

### III. EXPERIMENTAL RESULTS

To evaluate the statistical properties of the proposed image registration method, we performed three studies: simulations using a synthetic one-dimensional (1-D) signal, affine 2-D-to-2-D registration of fMRI images, and rigid 3-D-to-2-D registration of a torso phantom.

#### A. One-Dimensional Simulation

We first consider a 1-D “registration” problem where the only unknown parameter is the translation of the signal. (This is called *delay estimation* in communications.) This problem illustrates the statistical properties of the registration methods in the simplest possible setting. Fig. 1(a) shows the reference signal  $s_1(\tilde{t})$  that must be translated for registration with the signal  $s_2(\tilde{t})$  shown in Fig. 1(b) that includes both additive Gaussian noise and a segment of “outlier” signal values.

Using these signals, we computed three similarity measures as a function of the translation parameter  $\theta$ : the sample correlation coefficient, the robust correlation coefficient using weighting (17) with  $\delta = 2$  and the estimated MI using a  $32 \times 32$  interpolated joint histogram.

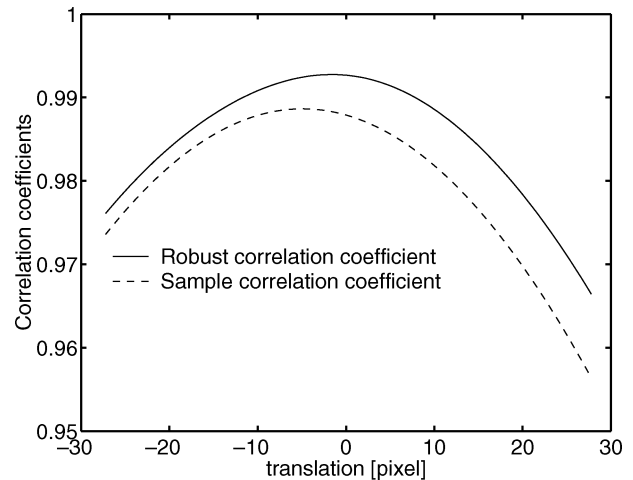


Fig. 2. Sample and robust correlation coefficients versus 1-D translation.

Fig. 2 shows the two correlation coefficients versus  $\theta$ , where the true value is  $\theta = 0$ . Due to the presence of the outliers, the sample correlation coefficient is maximized at an incorrect translation ( $\approx -4.8$  pixel). This type of systematic offset is observed for most noise realizations. In contrast, the robust correlation coefficient is maximized near the true translation ( $\approx -0.1$  pixel), illustrating the robustness of this similarity measure. This robustness can be understood by considering the joint histogram shown in Fig. 3(a) and the weighting function  $w(\|\hat{\mathbf{V}}(\mathbf{z} - \hat{\boldsymbol{\mu}})\|)$  shown in Fig. 3(b) (at the registered position where  $\theta = 0$ ). Although most of the histogram mass lies along the  $Y=X$  line, there is a group of outliers that degrade the conventional correlation coefficient estimate. Fig. 3(b) shows that the weighting function decreases the influence of those outliers, particularly those that are far from the primary linear ridge, thereby providing robustness.

Fig. 4 shows that the estimated MI is maximized at more correct translation ( $\approx -0.8$  pixel) than the sample correlation coefficient. This robustness of MI-based method is expected as argued in the preceding section.

The preceding results were for a single noise realization, so they do not fully characterize the bias and variance of the translation estimates. We performed 1000 noise realizations at each of several noise levels and computed translation estimates  $\hat{\theta}$  using each of the three similarity measures for each realization.

Fig. 5(a) shows the empirical translation estimation biases caused by the presence of the outliers for all three methods. As expected, the sample correlation coefficient method-based registration technique was the most sensitive to outliers, as argued in Appendix B.

The robust correlation-based method showed almost zero biases for low noise levels but increased biases for high noise levels. The MI-based method had the similar characteristics to the robust correlation-based method. However, the bias of the MI method was larger than the robust correlation-based method for all noise levels.

Fig. 5(b) shows the standard deviations of the three estimators. As expected, the sample correlation coefficient method had the smallest variance as argued in Appendix C. Also, shown for reference is the *Cramer-Rao* bound computed for a “no outlier”

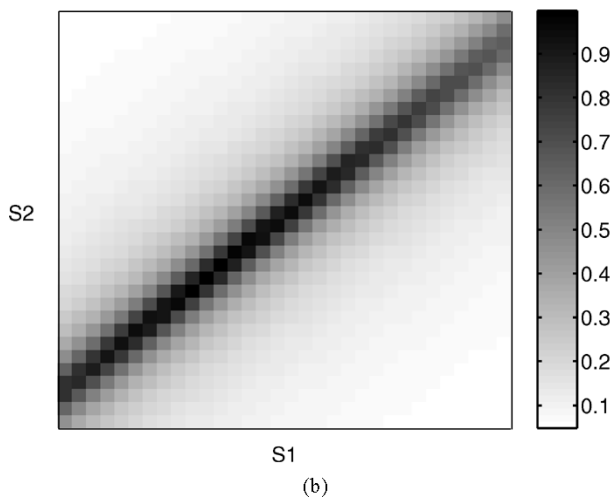
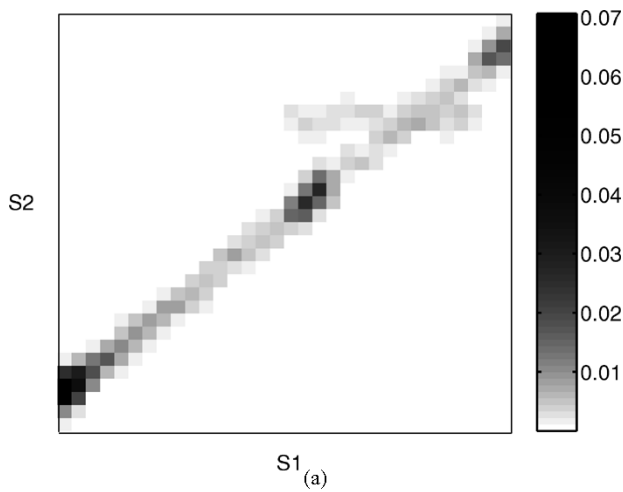


Fig. 3. Joint pdf and weighting function at registered position (evaluated on  $32 \times 32$  joint histogram). (a) Joint pdf. (b) Weighting function.

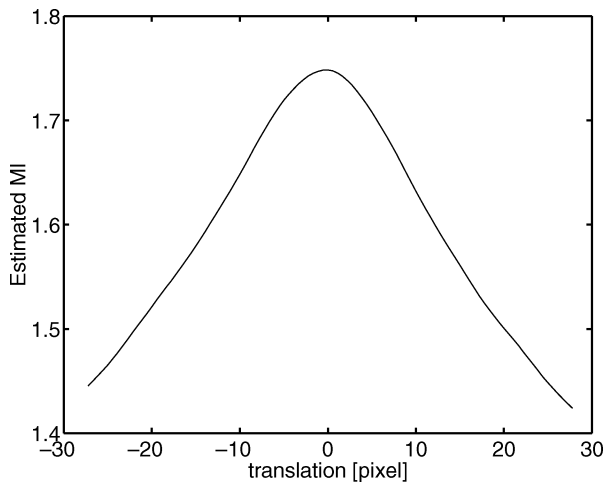


Fig. 4. MI similarity measure ( $32 \times 32$  interpolated joint histogram) versus 1-D translation.

model. In general, one would not expect the sample correlation method to match this bound for data containing outliers. However, for the particular signals in Fig. 1(a), (b), the standard deviation of the sample correlation method happened to match the *Cramer-Rao* bound. The standard deviation of the robust correlation method was almost the same as that of the MI method.

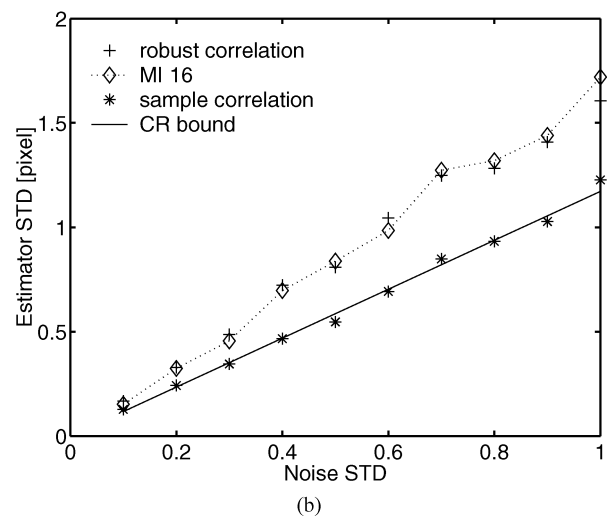
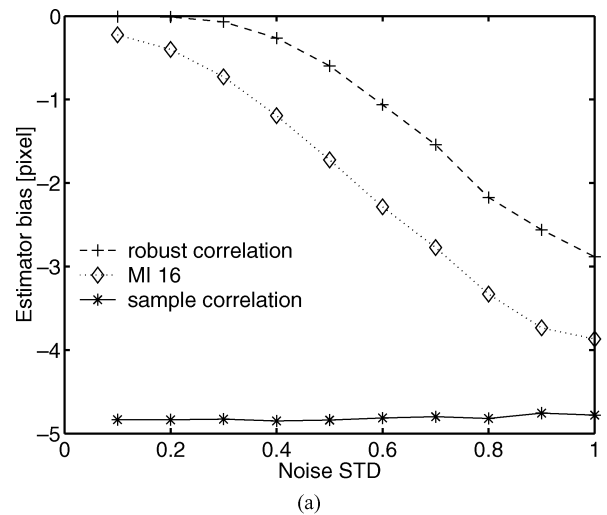


Fig. 5. Biases and standard deviations of the translation estimators versus Gaussian noise levels. (a) Biases. (b) Standard deviations.

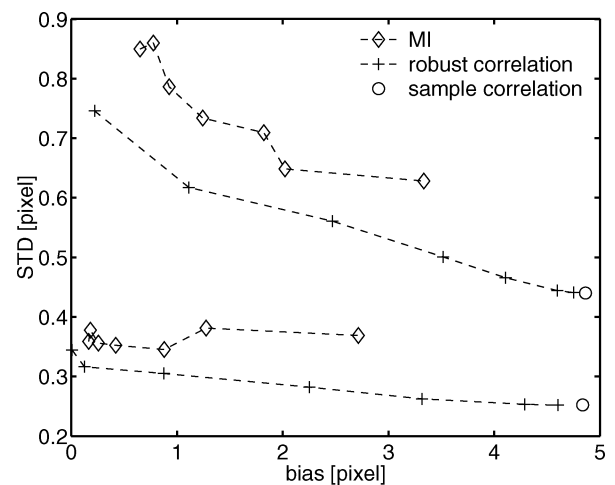


Fig. 6. Standard deviations versus bias of the translation estimators (upper part for noise STD = 0.4, lower part for noise STD = 0.2).

Since one may trade off robustness and efficiency by changing design parameter (i.e.,  $\delta$  for the robust correlation coefficient method and the number of bins for the MI method), we investigated bias-variance tradeoff of the estimators to evaluate the performance of each method.

TABLE I  
COMPUTATION TIME FOR EVALUATING EACH SIMILARITY MEASURE ONE TIME

similarity measure	Computation time	Iterations	Note
sample correlation	$3.50 \cdot 10^{-4}$		
MI( $16 \times 16$ interpolated histogram)	$1 \cdot 10^{-3}$		
robust ( $\delta = 2$ )	$1.76 \cdot 10^{-2}$	21.5	$\delta_V = 1 \cdot 10^{-3}, \delta_\mu = 1 \cdot 10^{-3}$
robust ( $\delta = 4$ )	$6.9 \cdot 10^{-3}$	7.6	$\delta_V = 1 \cdot 10^{-3}, \delta_\mu = 1 \cdot 10^{-3}$

Fig. 6 shows bias-variance plots of low noise case (lower part) with noise STD = 0.2 and high noise case (upper part) with noise STD = 0.4. Design parameters for the robust correlation method were  $\delta = 2, 3, 4, 5, 6, 8, 10$  and for the MI method were bin size  $8 \times 8, 10 \times 10, 12 \times 12, 16 \times 16, 20 \times 20, 24 \times 24, 28 \times 28$ .

For the robust correlation method, smaller  $\delta$  yielded smaller bias but larger variance. This is consistent with our analytical results, since smaller values for  $\delta$  correspond to a pdf model with “heavier tails.” The robust correlation coefficient was almost the same as the sample correlation coefficient when  $\delta$  was very large, as expected. For the MI method, there was a trend that larger number of bins (i.e., less smoothing) yielded smaller bias but larger variance. However, for low noise signal, the bias-variance characteristic of the MI method was very irregular. The performance of the robust correlation method was better than the MI method in the sense that the bias was smaller at the same variance or the variance was smaller at the same bias.

The better performance of the robust correlation method was achieved at the expense of computation time. Table I summarizes computation time requirement for evaluating each similarity measure one time on 3 GHz Intel Pentium 4 machine and the average number of iterations required for computing robust correlation coefficient one time. The computation time for the sample correlation coefficient was the smallest while the computation time requirement for the robust correlation coefficient was the most intensive. The computation time for the robust correlation coefficient depends greatly on the design parameter  $\delta$  and the stopping criteria. Note that the robust correlation coefficient for larger  $\delta$  was computed more quickly.

For the signals in Fig. 1(a), (b), the MI method showed a certain degree of robustness to outliers. However, in that example, the “outliers” in the observed signal had a constant intensity value. To investigate the behavior of each similarity measure for a different outlier distribution, we also generated the reference and observed signal shown in Fig. 7. In this case, the outliers corresponded to shifting a small portion of the reference signal.

Fig. 8 shows the three similarity measures as a function of the translation parameter. The sample correlation coefficient as well as the MI were maximized at incorrect positions due to outliers. However, the robust correlation coefficient was maximized at the correct translation. This example partly supports the argument that the robustness of the MI-based method depends on the particular images being registered.

### B. Two-Dimensional MRI Image Registration

We investigated the three similarity measures for the problem of registering two functional MRI images acquired with a spiral  $k$ -space trajectory. Both images were reconstructed from the

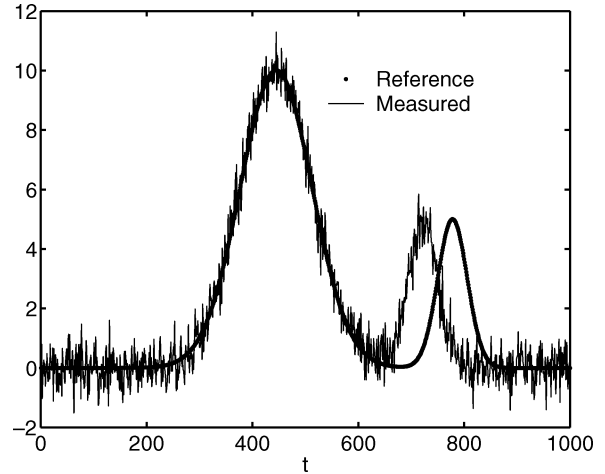


Fig. 7. Reference signal and measured signal with outliers and additive noise.

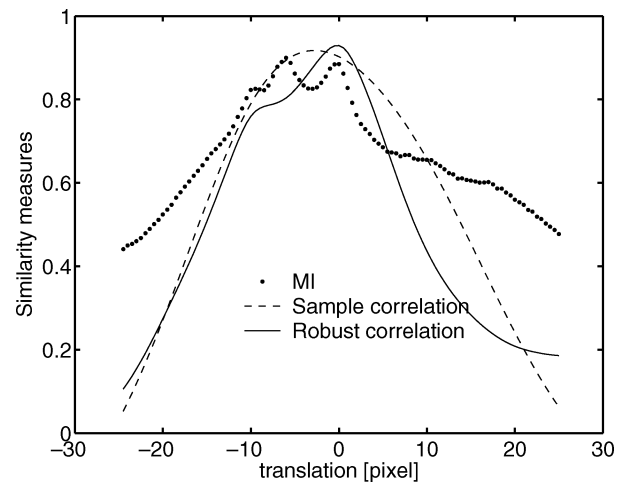


Fig. 8. Changes of the three similarity measures versus 1-D translation.

same raw data but one image was reconstructed with field inhomogeneity correction [27] while the other was without correction, so the true registered geometric transformation is identity.

Because of field inhomogeneity, there are signal voids in portions of the uncorrected images. We chose these images to test the registration methods since the true registered position (i.e., identity transformation) is known, yet the voids act as outliers. The robustness of the MI method for occluded images has been investigated previously [9].

Fig. 9(a), (b) shows the reference image and the target image. The anterior of the brain shows signal void in target image that is corrected in the reference image.

We plotted the change of each similarity measure as a function of vertical translation and scale. Fig. 10(a) shows that the sample correlation coefficient is maximized at incorrect position since if we translate the reference image down, brighter pixels



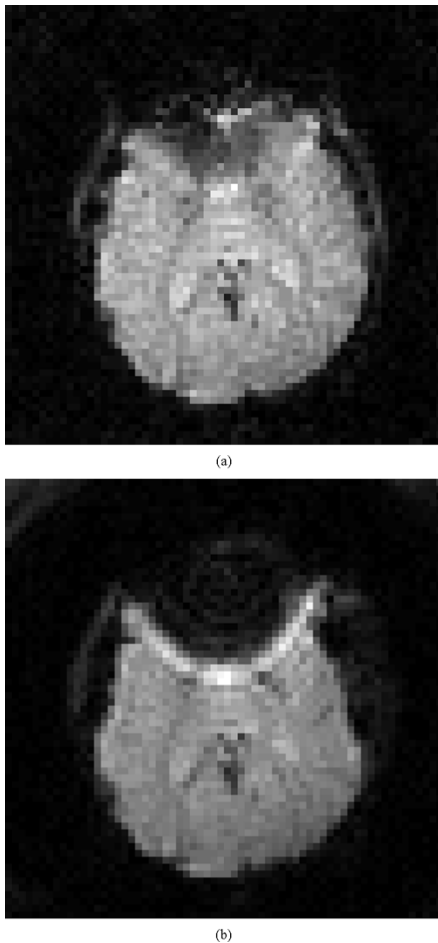


Fig. 9. Reference fMRI image with field inhomogeneity correction and target fMRI image without field inhomogeneity correction. (a) Reference image. (b) Target image.

in the reference image correspond to brighter pixels in target image better. Moreover, if we shrink the reference image, the correlation coefficient increases more since the brighter pixel region in Fig. 9(a) is larger than Fig. 9(b). As a result, the sample correlation coefficient is maximized at around 5% scaling down and  $-0.9$  pixel translation. As expected from the analysis, the sample correlation-based method was biased due to the outliers.

Fig. 10(b) and (c) shows that the registration errors of both MI and robust correlation-based estimators are smaller than the sample correlation-based estimator.

We tested the bias and variance of each estimator using 100 noise realizations. We added Rician noise since noise in reconstructed magnitude images is Rician [28], whereas noise in MRI raw data is complex Gaussian. Table II shows the empirical means and standard deviations based on 100 registration trials of the three cost functions, for the case case of a horizontal ( $t_x$ ) and vertical ( $t_y$ ) translation and vertical scaling ( $k_y$ ). The sample correlation method had the smallest variance but the largest bias due to the outliers. The robust correlation method had the best robustness (i.e., the smallest bias due to the outliers) and had smaller variance than the MI method. The MI method was more robust than the sample correlation method. The robustness of the MI method was improved as the number of histogram bins increased while the statistical efficiency was degraded. These tradeoffs are consistent with the 1-D simulations.

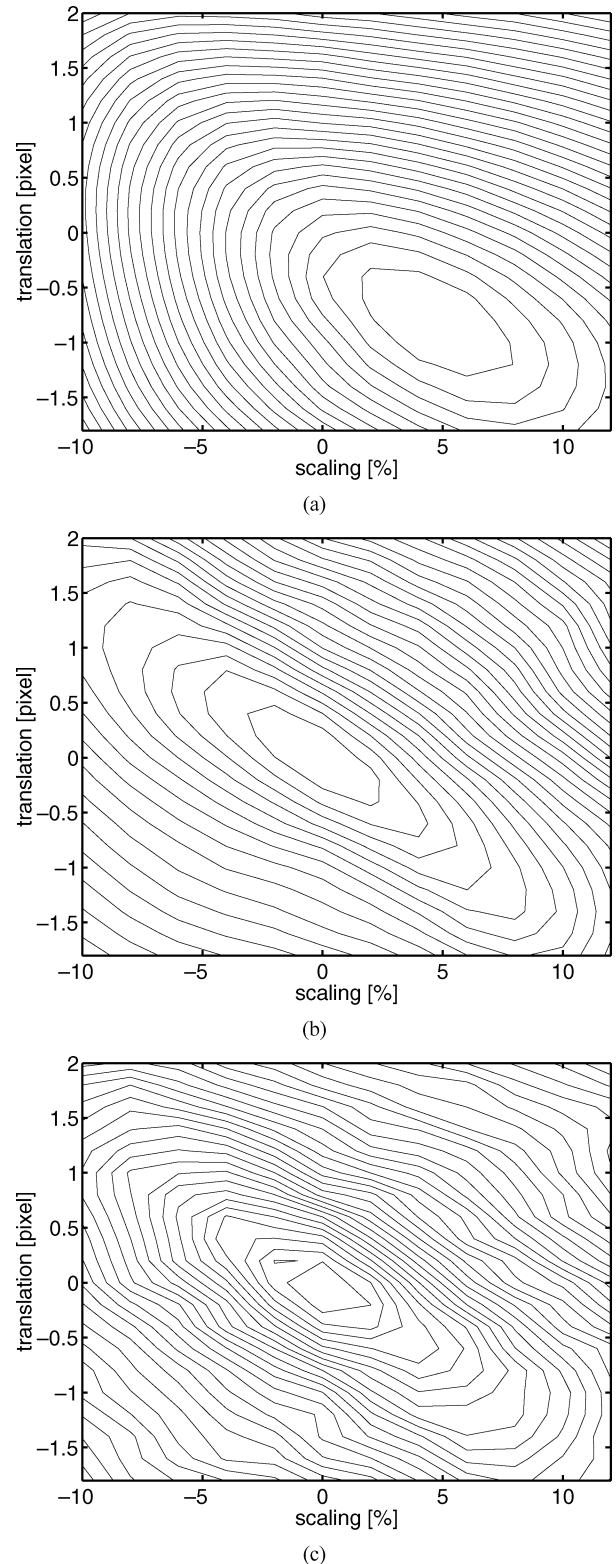


Fig. 10. Contour plots of the three similarity measures versus vertical scaling and translation. (a) Sample correlation coefficient. (b) Robust correlation coefficient. (c) MI.

### C. The 3-D/2-D Phantom Study

Previously we conducted an anthropomorphic phantom experiment to evaluate the performance of the set-up estimators

TABLE II  
MEAN (AND STD) OF ESTIMATED REGISTRATION PARAMETERS FOR 2-D-2-D MRI REGISTRATION

similarity measure	$t_x$	$t_y$	$k_y$	Computation time	Iterations	SNR
sample	0.41 (0.07)	-2.82 (0.10)	0.05 (0.01)	$5.99 \cdot 10^{-4}$		
MI (16×16 interpolated histogram)	0.14 (0.14)	-1.10 (0.17)	0.01 (0.01)	$2.3 \cdot 10^{-3}$		27.7 dB
MI (24×24 interpolated histogram)	0.07 (0.21)	-0.10 (0.24)	0.01 (0.01)	$2.4 \cdot 10^{-3}$		
robust ( $\delta = 2$ )	0.03 (0.10)	0.03 (0.12)	0.01 (0.01)	$1.48 \cdot 10^{-1}$	35.9	
sample	0.12 (0.05)	-2.89 (0.28)	0.05 (0.01)	$5.99 \cdot 10^{-4}$		
MI (16×16 interpolated histogram)	0.41 (0.21)	-1.90 (0.69)	0.02 (0.01)	$2.3 \cdot 10^{-3}$		13.8 dB
MI (24×24 interpolated histogram)	0.07 (0.34)	-1.41 (0.90)	0.02 (0.02)	$2.4 \cdot 10^{-3}$		
robust ( $\delta = 2$ )	0.10 (0.21)	-0.10 (0.52)	0.01 (0.01)	$1.05 \cdot 10^{-1}$	49.9	

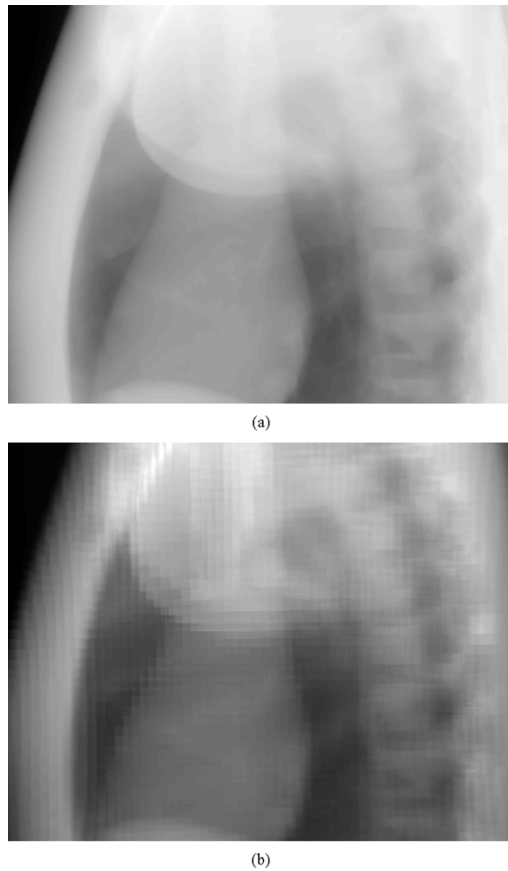


Fig. 11. Measured lateral radiograph and computed DRR from 3-D CT image for 3-D/2-D registration. (a) Measured lateral radiograph. (b) Lateral DRR computed from 3-D CT volume.

using 3-D/2-D image registration [1]. We estimated six parameters, rotations and translations along the X, Y, Z axes, from two orthogonal cone-beam projection views. For this study we used only one lateral view to estimate one rotation parameter and two translation parameters that are associated with in-plane motion. We chose this approach to better illustrate the effects of outliers since only the lateral image had noticeable outliers caused by the radiotherapy table. For this study, we held the other three parameters fixed at the “ground truth” values that were established by the most accurate marker-based method using eleven 1 mm diameter lead markers attached to the phantom’s surface [1].

One could estimate the three parameters using 2-D/2-D registration of digitally reconstructed radiograph (DRR) and the radiograph that geometrically transforms the DRR in 2-D plane to achieve registration. We did not follow this approach. Instead,

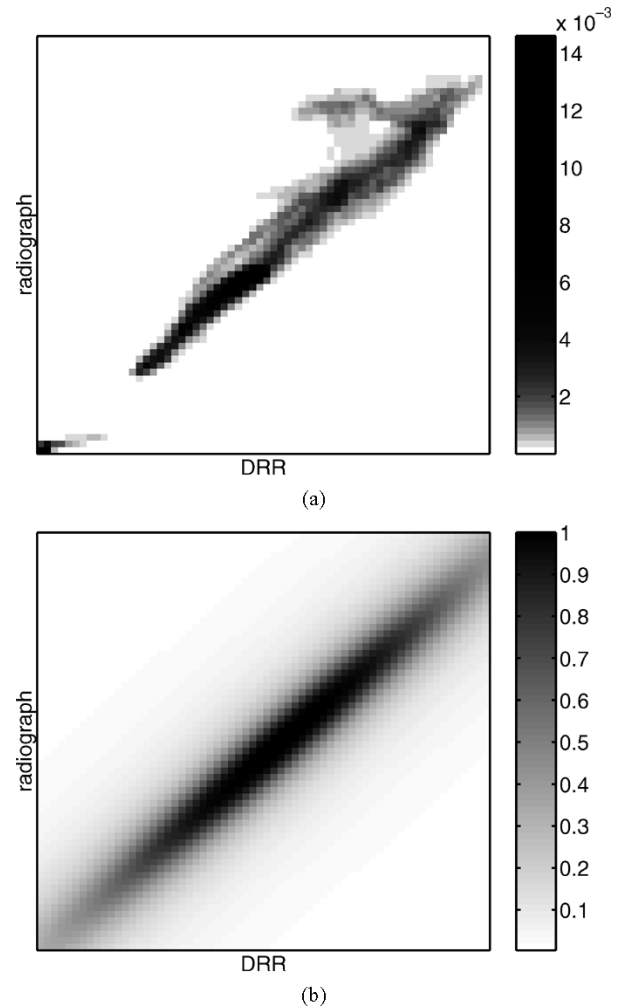


Fig. 12. Estimated joint pdf using  $64 \times 64$  interpolated joint histogram and weighting function corresponding to the proposed robust correlation coefficient, displayed over the domain of the histogram shown in Fig. 12(a). (a) Estimated joint pdf at the registered position. (b) Weighting function at the registered position.

we chose 3-D/2-D registration that computes DRR at each optimization step from transformed 3-D CT image. We chose this approach to demonstrate that general principles explained in this investigation are also applicable for volume-to-projection registration.

A  $512 \times 512 \times 398$  voxel CT image with  $0.9375 \times 0.9375 \times 1$  mm spacing was acquired on a GE CT/i scanner with a 140 kv x-ray source. Tattoos were drawn on the phantom where three alignment laser planes crossed the phantom to facilitate consistent set-up in a

TABLE III  
MEAN (AND STD) OF ESTIMATED SET-p PARAMETERS FOR 3-D-2-D REGISTRATION

Similarity measure	$\phi_x$	$t_y$	$t_z$	average TRE	Computation time	Iterations
sample	1.39 (0.02)	-2.06 (0.03)	2.11 (0.03)	5.61	0.014	
robust ( $\delta = 2$ )	0.95 (0.08)	-0.09 (0.09)	0.58 (0.23)	2.54	1.36	15.1
robust ( $\delta = 4$ )	1.19 (0.10)	-1.52 (0.07)	1.74 (0.13)	4.65	0.43	4.12
MI (32×32 interpolated histogram)	0.98 (0.08)	-0.55 (0.09)	0.81 (0.14)	3.05	0.043	
MI (64×64 interpolated histogram)	0.86 (0.07)	-0.44 (0.07)	0.73 (0.13)	2.65	0.045	
MI (64×64 histogram w/o interpolation)	0.90 (0.36)	-0.37 (0.16)	0.63 (0.38)	2.66	0.024	

treatment room. Next, the phantom was moved to the treatment room and it was set up at the isocenter by manually aligning tattoos to three laser planes in the treatment room. Four radiographs were obtained from different angles by rotating the x-ray source and Varian Portal Vision amorphous silicon active matrix flat panel image detector in 30° increments. For 90° view, we acquired 10 repeated radiographs without realignment for evaluating the effect of noise on the estimator. The x-ray source voltage was 6 MV and the detector size was 512 × 384 pixels with 0.78 mm × 0.78 mm spacing. We used only the 90° radiograph (i.e., the lateral image) for the correlation-based methods and the MI-based method. However, to enhance the accuracy of the “ground truth,” we used all four radiographs for the fiducial marker-based method. For all other methods except for the marker-based method, the planning CT image was down-sampled by four along each axis to reduce computation time and memory usage.

For image registration, while geometrically transforming the CT image, we computed DRR of the transformed CT from the same angle as the radiograph. The registration is achieved by maximizing the similarity measure between such DRR and radiograph. We used only the central 400 × 300 subimage of the DRR and the radiograph to avoid the effect of the markers which are not usually used in clinical practice. We have established the geometry of the EPID imaging systems by determining radiation field edges using simple thresholding method [29].

Fig. 11(a) shows the radiograph and Fig. 11(b) shows the DRR at the registered position. The radiotherapy table causes pixels along the rightmost part of the radiograph to be brighter than the corresponding DRR pixels.

Fig. 12(a) shows the estimated joint histogram from the registered DRR and radiograph. The histogram has a dominant distribution along the  $Y=X$  line and an outlier distribution caused by the radiotherapy table. Fig. 12(b) shows the weighting function from the robust correlation estimate. This weighting clearly reduces the influence of the outliers.

We repeated 10 estimations using 10 acquisitions of the radiograph. Table III summarizes the experimental results. Since a small rotation error may result in a large registration error depending on the location of the rotation center, we evaluated target registration error (TRE) values [30] in addition to parameter estimation error values. We computed TRE values at the spatial locations along the bottom row of in Fig. 11(a), locations that are far from the rotation center. The average TRE values are shown in Table III for each method. The experimental results were consistent with the previous simulation results. The robust correlation coefficient using (17) with  $\delta = 2$  had the smallest bias due to outliers (i.e., the most robust to outliers). The sample correlation-based method was the worst in terms of the robust-

ness but the best in terms of the variance. Interestingly, the MI method showed small variance and small bias as well. We think that this was because the shape of the estimated joint pdf was close to a normal distribution and the number of samples from outliers was small.

#### IV. DISCUSSION

Statistical properties such as bias, robustness, efficiency are very important in designing image registration methods. In previous investigations, the bias of the sample correlation method for intramodality image registration and that of the MI method for multimodality registration have been studied extensively empirically<sup>6</sup> [1], [4], [5], [8], [9]. Some authors also reported that the MI method is more robust to outliers than the sample correlation coefficient method [1], [31].

The sample correlation is a natural similarity measure for intramodality image registration [4], [5], [31] and the MI method performs well for both intramodality [28], [31] and multimodality image registrations [8]–[10]. Considering only intramodality image registration, we point out that those two most well known similarity measures have drawbacks such as nonrobustness to outliers of the sample correlation method and statistical inefficiency of the MI method. Moreover, even though the MI method has a certain degree of robustness for many cases, it may not be very robust for some cases since the MI method does not reject nor reduce the influence of outliers directly.

We have proposed a robust correlation coefficient-based image registration method to improve the robustness of the sample correlation method without compromising statistical efficiency very much. We showed analytically and experimentally that the robust correlation method has improved robustness but larger variance compared with the sample correlation method.

The statistical properties of the robust correlation method are controlled by underlying pdf model. More specifically, one may improve the robustness by using “heavier tail” underlying pdf model  $f_0$  in (10) at the expense of the efficiency.

For the MI-based method, it is difficult to relate the design parameters (i.e., joint pdf estimation method) to its statistical properties analytically. Instead, if we summarize empirical observations in our simulations and experiments, the MI method become more robust but less efficient when more bins were used. Roughly speaking, increasing the number of bins is equivalent to narrowing the Parzen window, i.e., to less smoothing. We also observed that using joint histograms without interpolation increased variance, and increasing the number of bins excessively

<sup>6</sup>Since many previous studies ignored the effect of noise, bias was called *registration error*.

led to many local maxima. More analyses are needed to relate the statistical properties of the MI-based method to the joint pdf estimation method. One method for designing kernel functions was proposed in a different point of view [8].

Even when qualitative relationships between design parameters and statistical properties are available, automatic determination of the design parameters remains challenging. In our simulations and experiments, we determined  $\delta$  for the robust correlation-based method and the number of bins for the MI-based method empirically. More investigations are required to automate this selection. Such studies should consider image sample distributions and noise characteristics in a particular context.

It is challenging to compare the performance of one image registration method with another since the performances depend both on design parameters and on the images being registered. Despite the difficulty, if we compare the performance of the robust correlation method with the MI method based on our simulations and experiments, the robust correlation method performed better than the MI method in our 1-D simulation and 2-D MRI simulation in the sense that the variance of the robust correlation method was smaller at the same bias or the bias was smaller at the same variance. Interestingly, the MI method was very efficient in the 3-D/2-D phantom experiments. We suspect that this was because the estimated pdf shape was close to the normal distribution as argued in Section II-D.

We believe that there exist several advantages of the robust correlation method over the MI method. First of all, the robust correlation method always has robustness to outliers since it reduces or rejects the effects of the outlier samples based on the statistical distances. In contrast, the robustness of the MI method depends on the behavior of estimated pdf without explicitly excluding the effects of outlier samples. Therefore, its robustness can be very image dependent. We suspect that the MI method may not be very robust for some images, as illustrated in Fig. 7. Another advantage is that the design parameters of the robust correlation method relate directly to the statistical properties, whereas the relationship of the MI method is less clear analytically.

The disadvantage of the robust correlation method is increased computation time. In our simulations and experiments, computation time for the robust correlation coefficient increased when “heavier tail” underlying pdfs were assumed. Since the robust correlation coefficient by  $M$ -estimation method requires numerical optimization, the computation time depends not only on the images being registered and the underlying pdf but also on the stopping criteria. However, for nonrigid registration problems, computing the warping interpolations is likely to outweigh the cost of evaluating the similarity measure.

Conceivably the performance of the MI method could be improved by designing different MI estimation methods and/or using more generalized Rényi entropy [9], [26]. Similarly, one might improve the robust correlation method using different robust techniques for estimating the correlation coefficient [11]–[13].

The sample correlation-based estimator is the most efficient estimator among unbiased estimators when noise is i.i.d. Gaussian and there are no outliers (see Appendix C). Even though the noise was not Gaussian in practice (Rician for MRI

simulation, Poisson for 3-D/2-D experiment) and outliers were present, the sample correlation method was the most efficient in all of our simulations and experiments. We suspect that this is because Rician and Poisson noise are approximately Gaussian, although the variance at each pixel is not the same.

Analysis using mean and variance approximations provided qualitative arguments about the statistical properties of the intensity-based image registration methods. By choosing the  $M$ -estimation method for robust correlation estimation, we were able to analyze the robust correlation and the sample correlation within the same framework. In addition, we were able to analyze the MI-based method by representing estimated MI as a type of weighted correlation. Beyond the analysis carried out in this investigation, it would be desirable if the statistical properties of estimators can be related analytically to design parameters. For example, if one can approximate the bias and variance of the MI-based estimator as a function of Parzen window width, that approximation can be very useful. Since the statistical properties of the image registration methods have not been analyzed much, we expect that our analysis can be a first step for further investigations.

## V. CONCLUSION

We have introduced a novel intensity-based similarity measure, a robust correlation coefficient, to design an image registration method that is robust to outliers. Using the proposed image registration technique, we achieved improved robustness relative to the sample correlation-coefficient-based method without compromising statistical efficiency much. Moreover, in 1-D simulation, 2-D and 3-D experiments, the proposed method performed better than the well-known MI-based image registration method. We believe that the robust correlation coefficient is an effective similarity measure for intramodality image registration task where the presence of the outliers is unavoidable such as set-up estimation for radiotherapy and image-guided surgery.

## APPENDIX A

### MEAN AND COVARIANCE APPROXIMATION

This appendix presents approximations for the mean and variance of implicitly defined estimators such as (1). Let  $\Phi(\theta, Z)$  be an objective function depending on unknown parameter  $\theta$  and noisy measurement  $Z$ . We assume that the objective function has a unique maximum for any  $Z$  and that one can find the maximum by zeroing the partial derivatives of  $\Phi(\cdot, Z)$  to determine the estimates  $\hat{\theta}$ :

$$0 = \frac{\partial}{\partial \theta_j} \Phi(\theta, Z) \Big|_{\theta=\hat{\theta}}, j = 1, \dots, p. \quad (23)$$

Since  $\hat{\theta}$  is an implicit function of  $Z$ , it is difficult to determine its exact mean and variance. To approximate its mean and variance, we linearize the gradient of  $\Phi$ .

Define the column gradient of the objective function as follows:

$$\Psi(\theta, Z) \triangleq \nabla^{10} \Phi(\theta, Z) \quad (24)$$

where the  $j$ th element of  $p \times 1$  operator  $\nabla^{10}$  is  $\partial/\partial\theta_j$ . Linearizing  $\Psi(\hat{\theta}, Z)$  around the *true* parameter  $\tilde{\theta}$  by the first-order Taylor series expansion yields

$$\Psi(\hat{\theta}, Z) \approx \Psi(\tilde{\theta}, Z) + \nabla^{20}\Phi(\tilde{\theta}, Z)(\hat{\theta} - \tilde{\theta}) \quad (25)$$

where the  $(j, k)$ th element of  $p \times p$  operator  $\nabla^{20}$  is  $\partial^2/\partial\theta_j\partial\theta_k$ . We assume that  $[-\nabla^{20}\Phi(\tilde{\theta}, Z)]$  is a positive definite symmetric matrix so that its inverse is well defined. Since  $\Psi(\tilde{\theta}, Z) = 0$  by (23), the estimator  $\hat{\theta}$  can be approximated as follows:

$$\hat{\theta} \approx \tilde{\theta} + [-\nabla^{20}\Phi(\tilde{\theta}, Z)]^{-1}\Psi(\tilde{\theta}, Z). \quad (26)$$

Rearranging (26) yields the following bias approximation:

$$E[\hat{\theta}] - \tilde{\theta} \approx E\left[[-\nabla^{20}\Phi(\tilde{\theta}, Z)]^{-1}\Psi(\tilde{\theta}, Z)\right] \approx H^{-1}\Psi(\tilde{\theta}, E[Z]), \quad (27)$$

where possible approximations for  $H$  matrix include

$$H = -\nabla^{20}\Phi(\tilde{\theta}, E[Z]). \quad (28)$$

The covariance of the estimator is approximated as follows:

$$\text{Cov}\{\hat{\theta}\} \approx \text{Cov}\left\{[-\nabla^{20}\Phi(\tilde{\theta}, Z)]^{-1}\Psi(\tilde{\theta}, Z)\right\} \approx H^{-1}\text{Cov}\{\Psi(\tilde{\theta}, Z)\}H^{-1}. \quad (29)$$

One may approximate the mean and covariance further by linearizing  $\Psi(\theta, Z)$  with respect to the measurements  $Z$  as developed previously [17].

## APPENDIX B

### BIAS AND ROBUSTNESS OF CORRELATION BASED METHODS

This appendix uses the general results from Appendix A to analyze the bias and robustness of correlation-based registration methods. Define the noisy measurement  $Z = [s_1(\cdot), s_2(\cdot)]$ . We represent  $s_1(T_\theta(\vec{t})) = s_1(\theta, \vec{t})$  for notational convenience and  $\mathbf{z}_\theta(\vec{t}) = [s_1(\theta, \vec{t}), s_2(\vec{t})]^T$ . Without loss of generality, we assume that the volume of the field of view is unity, and the image empirical means are zero:  $\int s_1(\theta, \vec{t})d\vec{t} = 0$  and  $\int s_2(\vec{t})d\vec{t} = 0$ .

The empirical variances are, thus,  $\sigma_X^2(\theta) = \int w(\|V_\theta \mathbf{z}_\theta(\vec{t})\|)s_1^2(\theta, \vec{t})d\vec{t}$  and  $\sigma_Y^2(\theta) = \int w(\|V_\theta \mathbf{z}_\theta(\vec{t})\|)s_2^2(\vec{t})d\vec{t}$ .

We analyze the robust correlation method having fixed zero mean and the following objective function:

$$\Phi(\theta, Z) = \frac{C_{XY}(\theta)}{\sigma_X(\theta)\sigma_Y(\theta)} \quad (30)$$

where  $C_{XY}(\theta) = \int w(\|V_\theta \mathbf{z}_\theta(\vec{t})\|)s_1(\theta, \vec{t})s_2(\vec{t})d\vec{t}$ . The gradient of the objective function is evaluated as follows:

$$\Psi(\theta, Z) = \frac{1}{\sigma_X^2(\theta)\sigma_Y^2(\theta)} \left[ C'_{XY}(\theta)\sigma_X(\theta)\sigma_Y(\theta) - C_{XY}(\theta)(\sigma_X(\theta)\sigma'_Y(\theta) + \sigma'_X(\theta)\sigma_Y(\theta)) \right] \quad (31)$$

where  $C'_{XY}(\theta) = \nabla_\theta C_{XY}(\theta)$ ,  $\sigma'_X(\theta) = \nabla_\theta \sigma_X(\theta)$  and  $\sigma'_Y(\theta) = \nabla_\theta \sigma_Y(\theta)$ .

If  $s_1(\tilde{\theta}, \vec{t}) = ks_2(\vec{t})$  with a constant  $k$ , (i.e., the two images are linearly related at true registered position), then  $\Psi(\tilde{\theta}, E[Z]) = 0$  since  $\sigma_X(\tilde{\theta})\sigma_Y(\tilde{\theta}) = C_{XY}(\tilde{\theta})$  and  $C'_{XY}(\tilde{\theta}) = \sigma'_X(\tilde{\theta})\sigma_Y(\tilde{\theta}) + \sigma_X(\tilde{\theta})\sigma'_Y(\tilde{\theta})$ . Therefore, the bias approximation using (27) is zero. This holds not only for the constant weighting function but also for any weighting function.

Next, we consider bias due to outliers. Suppose that there are additive outlier components  $o(\vec{t})$  and noise  $n(\vec{t})$  such that  $s_2(\vec{t}) = ks_1(\tilde{\theta}, \vec{t}) + o(\vec{t}) + n(\vec{t})$ . In this case,  $\Psi(\tilde{\theta}, E[Z])$  does not equal zero due to the outliers, in general. For simplicity, we assume<sup>7</sup> that  $\sigma'_X(\theta) = \sigma'_Y(\theta) = 0$  for all  $\theta$ . With this assumption, we approximate the Hessian matrix at the true registered position as follows:

$$H \approx \nabla^{20}\Phi(\tilde{\theta}, E[Z]) = \frac{1}{\sigma_X^2(\tilde{\theta})\sigma_Y^2(\tilde{\theta})} [C''_{XY}(\tilde{\theta})\sigma_X(\tilde{\theta})\sigma_Y(\tilde{\theta})]. \quad (32)$$

Plugging (32) and  $\Psi(\tilde{\theta}, E[Z])$  into (27) yields the bias approximation shown in (33) at the bottom of the page.

For constant weighting, such as in the sample correlation method, the estimator can be biased by outliers since  $\int [\nabla^{10}s_1(\tilde{\theta}, \vec{t})]o(\vec{t})d\vec{t}$  is nonzero in general. For the robust correlation method, the bias can be reduced significantly by the weighting in the final bracketed term in (33).

## APPENDIX C

### EFFICIENCY

The sample correlation coefficient-based method is the MLE if the two images have constant backgrounds and are identical

<sup>7</sup>If the backgrounds of the images have constant values, then this assumption holds for constant weighting (i.e., the sample correlation coefficient).

$$E[\hat{\theta}] - \tilde{\theta} \approx [C''_{XY}(\tilde{\theta})]^{-1}C'_{XY}(\tilde{\theta}) = \left[ \int w(\|V_{\tilde{\theta}} \mathbf{z}_{\tilde{\theta}}(\vec{t})\|) \left( \nabla^{10}s_1(\tilde{\theta}, \vec{t})[\nabla^{10}s_1(\tilde{\theta}, \vec{t})]^T - \nabla^{20}s_1(\tilde{\theta}, \vec{t})o(\vec{t}) \right) d\vec{t} \right]^{-1} \cdot \left[ \int w(\|V_{\tilde{\theta}} \mathbf{z}_{\tilde{\theta}}(\vec{t})\|) [\nabla^{10}s_1(\tilde{\theta}, \vec{t})]o(\vec{t}) d\vec{t} \right] \quad (33)$$

except for geometric transformation and additive i.i.d. Gaussian noise [32]. Thus, the covariance of the sample correlation-based estimator asymptotically equals the inverse of the following Fisher information matrix [23]:

$$\begin{aligned} I_{\tilde{\theta}} &= \frac{1}{2\sigma_n^2} E \left[ -\nabla_{\tilde{\theta}}^2 \int (s_2(\vec{t}) - s_1(\tilde{\theta}, \vec{t}))^2 d\vec{t} \right] \\ &= \frac{1}{\sigma_n^2} \int \left[ \nabla^{10} s_1(\tilde{\theta}, \vec{t}) \right] \left[ \nabla^{10} s_1(\tilde{\theta}, \vec{t}) \right]^T d\vec{t} \end{aligned}$$

where  $\sigma_n^2$  is the noise power.

For robust correlation-based estimators, we use the covariance approximation developed in Appendix A. First, we approximate the covariance matrix of  $\Psi(\tilde{\theta}, Z)$  as follows:

$$\begin{aligned} \text{Cov} \left\{ \Psi(\tilde{\theta}, Z) \right\} &\approx \frac{1}{\sigma^2(\tilde{\theta})} \text{Cov} \left\{ C'_{XY}(\tilde{\theta}) - \left( \frac{\sigma'_X(\tilde{\theta})}{\sigma_X(\tilde{\theta})} + \frac{\sigma'_Y(\tilde{\theta})}{\sigma_Y(\tilde{\theta})} \right) \right\} \\ &\approx \frac{1}{\sigma^2(\tilde{\theta})} E \left\{ \int w(\|V_{\tilde{\theta}} z_{\tilde{\theta}}(\vec{t})\|) n(\vec{t}) [\nabla^{10} s_1(\tilde{\theta}, \vec{t})] d\vec{t} \right. \\ &\quad \cdot \left. \int w(\|V_{\tilde{\theta}} z_{\tilde{\theta}}(\vec{\tau})\|) n(\vec{\tau}) [\nabla^{10} s_1(\tilde{\theta}, \vec{\tau})]^T d\vec{\tau} \right\} \\ &= \frac{\sigma_n^2}{\sigma^2(\tilde{\theta})} \int w(\|V_{\tilde{\theta}} z_{\tilde{\theta}}(\vec{t})\|)^2 [\nabla^{10} s_1(\tilde{\theta}, \vec{t})] [\nabla^{10} s_1(\tilde{\theta}, \vec{t})]^T d\vec{t} \quad (34) \end{aligned}$$

where  $\sigma^2(\tilde{\theta}) \triangleq \sigma_X(\tilde{\theta})\sigma_Y(\tilde{\theta})$ . We approximate the  $H$  matrix without outliers using (32)

$$H \approx \frac{1}{\sigma^2(\tilde{\theta})} \int w(\|V_{\tilde{\theta}} z_{\tilde{\theta}}(\vec{t})\|) [\nabla^{10} s_1(\tilde{\theta}, \vec{t})] [\nabla^{10} s_1(\tilde{\theta}, \vec{t})]^T d\vec{t} \quad (35)$$

Finally, the covariance is approximated by plugging (34) and (35) into (29)

$$\text{Cov} \left\{ \hat{\theta} \right\} \approx \sigma_n^2 C_1^{-1} C_2 C_1^{-1} \quad (36)$$

where  $C_1 = \int w(\|V_{\tilde{\theta}} z_{\tilde{\theta}}(\vec{t})\|) [\nabla^{10} s_1(\tilde{\theta}, \vec{t})] [\nabla^{10} s_1(\tilde{\theta}, \vec{t})]^T d\vec{t}$  and  $C_2 = \int w^2(\|V_{\tilde{\theta}} z_{\tilde{\theta}}(\vec{t})\|) [\nabla^{10} s_1(\tilde{\theta}, \vec{t})] [\nabla^{10} s_1(\tilde{\theta}, \vec{t})]^T d\vec{t}$ .

By the vector Cauchy-Schwarz inequality [24], one can prove the following inequality:

$$\text{Cov} \left\{ \hat{\theta} \right\} \succeq I_{\tilde{\theta}}^{-1} \quad (37)$$

where  $A \succeq B$  means the positive semi-definiteness of matrix  $A - B$ , and equality holds iff  $w(\cdot) [\nabla^{10} s_1(\tilde{\theta}, \vec{t})] = \alpha [\nabla^{10} s_1(\tilde{\theta}, \vec{t})]$  for a constant  $\alpha$ . Thus, any nonconstant weighting function will yield larger covariance than the ordinary sample correlation-based estimator, whereas the sample correlation coefficient method achieves (asymptotically) the Cramer-Rao bound (the inverse of the Fisher information matrix). However, this analysis assumed  $s_2(\vec{t}) = k s_1(\tilde{\theta}, \vec{t}) + n(\vec{t})$ , which is unrealistic in the presence of outliers. When outliers occur, the noise variance is not uniform. One may also approximate the covariance matrix of each estimator in the presence of outliers. However, it is challenging to find an inequality such as (37) for this case.

## ACKNOWLEDGMENT

The authors would like to thank J. Balter, K. Lam and R. Ten Haken from the Radiation Oncology Department in the University of Michigan, Ann Arbor, for providing the experimental data and bringing image registration problem for radiation therapy to us. They also thank B. Sutton and D. Noll from the Biomedical Engineering Department in the University of Michigan, Ann Arbor, for providing functional MRI images.

## REFERENCES

- [1] J. Kim, J. A. Fessler, K. L. Lam, J. M. Balter, and R. K. T. Haken, "A feasibility study of mutual information based set-up error estimator for radiotherapy," *Med. Phys.*, vol. 28, no. 12, pp. 2391–2593, 2001.
- [2] M. J. Murphy, "An automatic six-degree-of-freedom image registration algorithm for image-guided frameless stereotactic radiosurgery," *Med. Phys.*, vol. 24, no. 6, p. 857, June 1997.
- [3] J. Weese, G. P. Penney, P. Desmedt, T. M. Buzug, D. L. G. Hill, and D. J. Hawkes, "Voxel-based 2-D/3-D registration of fluoroscopy images and CT scans for image-guided surgery," *IEEE Trans. Inform. Technol. Biomed.*, vol. 1, pp. 284–293, Dec. 1997.
- [4] L. Lemieux, R. Jagoe, D. R. Fish, N. D. Kitchen, and D. G. T. Thomas, "A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs," *Med. Phys.*, vol. 21, no. 11, pp. 1749–1760, 1994.
- [5] L. Dong and A. L. Boyer, "An image correlation procedure for digitally reconstructed radiographs and electronic portal images," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 33, no. 5, pp. 1053–1060, 1995.
- [6] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
- [7] J. Kim and J. A. Fessler, "Image registration using robust correlation," in *Proc. IEEE Int. Symp. Biomedical Imaging*, 2002, pp. 353–356.
- [8] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Imag. Anal.*, vol. 1, no. 1, pp. 35–51, Mar. 1996.
- [9] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, pp. 187–98, Apr. 1997.
- [10] C. R. Meyer *et al.*, "Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin plate spline warped geometric deformations," *Med. Imag. Anal.*, vol. 1, no. 3, pp. 195–206, 1997.
- [11] D. M. Titterton, "Estimation of correlation coefficients by ellipsoidal trimming," *Appl. Statist.*, vol. 23, no. 3, pp. 227–234, 1978.
- [12] N. A. Campbell, "Robust procedure in multivariate analysis 1: Robust covariance estimation," *Appl. Statist.*, vol. 29, no. 3, pp. 231–237, 1980.
- [13] P. J. Rousseeuw and B. C. van Zomeren, "Unmasking multivariate outliers and leverage points," *J. Amer. Statist. Assoc.*, vol. 85, no. 411, pp. 633–639, Sept. 1990.
- [14] C. Nikou, F. Heitz, J.-P. Armspach, I.-J. Namer, and D. Grucker, "Registration of MR/MR and MR/SPECT brain images by fast stochastic optimization of robust voxel similarity measures," *NeuroImage*, vol. 8, no. 1, pp. 30–43, July 1998.
- [15] O. Nestares and D. J. Heeger, "Robust multiresolution alignment of MRI brain volumes," *Magn. Reson. Med.*, vol. 43, no. 5, pp. 705–15, May 2000.
- [16] L. Freire, A. Roche, and J.-F. Mangin, "What is the best similarity measure for motion correction in fMRI time series?," *IEEE Trans. Med. Imag.*, vol. 21, pp. 470–84, May 2002.
- [17] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Processing*, vol. 5, pp. 493–506, Mar. 1996.
- [18] E. L. Lehmann, *Theory of Point Estimation*. New York: Wiley, 1983.
- [19] F. Maes, D. Vandermeulen, and P. Suetens, "Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information," *Med. Imag. Anal.*, vol. 3, no. 4, pp. 373–86, 1999.
- [20] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. New York: Chapman & Hall, 1986.
- [21] B. Ma, A. Hero, J. Gorman, and O. Michel, "Image registration with minimum spanning tree algorithm," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2000, pp. 481–484.

- [22] H. F. Neemuchwala, A. O. Hero, and P. L. Carson, "Image registration using entropic graph-matching criteria," in *Proc. 36th Asilomar Conf. Signals Systems and Computers*, Pacific Grove, CA, Nov. 2002, pp. 134–138.
- [23] M. Cramer, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.
- [24] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimators with constraints," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1285–1301, Nov. 1990.
- [25] J. Kim, "Intensity based image registration using robust similarity measure and constrained optimization : Applications for radiation therapy," Ph.D. dissertation, Univ. Michigan, Ann Arbor, 2004.
- [26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [27] D. C. Noll, C. H. Meyer, J. M. Pauly, D. G. Nishimura, and A. Macovski, "A homogeneity correction method for magnetic resonance imaging with time-varying gradients," *IEEE Trans. Med. Imag.*, vol. 10, p. 629, Dec. 1991.
- [28] M. Holden, D. L. G. Hill, E. R. E. Denton, J. M. Jarosz, T. C. S. Cox, T. Rohlfing, J. Goodey, and D. J. Hawkes, "Voxel similarity measures for 3-D serial MR brain image registration," *IEEE Trans. Med. Imag.*, vol. 19, pp. 94–102, Feb. 2000.
- [29] L. Dong and A. L. Boyer, "A portal image alignment and patient setup verification procedure using moments and correlation techniques," *Phys. Med. Biol.*, vol. 41, pp. 697–723, 1996.
- [30] J. M. Fitzpatrick, J. M. West, and C. R. Maurer, Jr., "Predicting error in rigid body point based registration," *IEEE Trans. Med. Imag.*, vol. 17, pp. 694–702, Oct. 1998.
- [31] P. A. Viola, "Alignment by maximization of mutual information," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, 1995.
- [32] M. D. Srinath, P. K. Rajasekaran, and R. Viswanathan, *Introduction to Statistical Signal Processing with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1996.