A Continuous Method for Reducing Interpolation Artifacts in Mutual Information-Based Rigid Image Registration

Lin Xu, Justin W. L. Wan, and Tiantian Bian

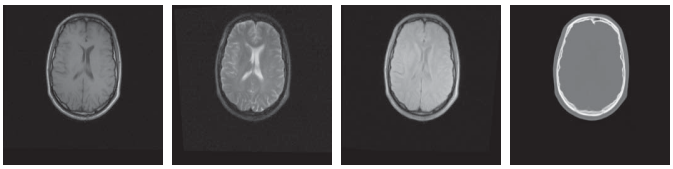
***Abstract*— We propose an approach for computing mutual information in rigid multimodality image registration. Images to be registered are modeled as functions defined on a continuous image domain. Analytic forms of the probability density functions for the images and the joint probability density function are first defined in 1D. We describe how the entropies of the images, the joint entropy, and mutual information can be computed accurately by a numerical method. We then extend the method to 2D and 3D. The mutual information function generated is smooth and does not seem to have the typical interpolation artifacts that are commonly observed in other standard models. The relationship between the proposed method and the partial volume (PV) model is described. In addition, we give a theoretical analysis to explain the nonsmoothness of the mutual information function computed by the PV model. Numerical experiments in 2D and 3D are presented to illustrate the smoothness of the mutual information function, which leads to robust and accurate numerical convergence results for solving the image registration problem.**

***Index Terms*— Continuous model, image registration, interpolation artifacts, multimodality, mutual information, rigid.**

I. Introduction

MUTUAL information [1] was introduced by Viola and Wells [2] and Collignon *et al.* [3] as a similarity measure for rigid multimodality image registration, which aligns images generated from different sensor devices; see Fig. 1. The evaluation of mutual information requires the computation of the probability distribution functions (PDFs) of the image intensity values. However, since only discrete intensity values can be acquired from raw images, PDFs are usually estimated using discrete probability distributions. Collignon *et al.* estimate the probability distributions based on histogramming, as one of the interpolation-based models (IM). Viola and Wells [4] apply the Parzen window method for density estimation. Maes *et al.* [5] introduce the partial volume model (PV).

However, it has been observed that the mutual information function computed by standard methods is not smooth, which is generally known as interpolation artifacts usually in the form of local minima, local maxima, or "elbows" typically present in perfect (or close to) grid alignment (see Fig. 2). These artifacts pose difficulties to robust and accurate image registration using optimization schemes. There have been a number of recent studies of the interpolation artifacts in the literature [6]–[17]. Chen and Varshney [7] propose a generalized partial volume model based on higher order B-spline interpolation. Other approaches include resizing the pixels [16], jittering or blurring images [16], using non-uniform interpolation function as the interpolation kernels [18], balancing the size of bins for computing probability distributions [9], [19], and quasi-random image sampling based on Halton sequences [20]. Some recent studies [14], [21], [22] attribute the interpolation artifacts with the non-constant variance of the interpolated image when the image is shifted between pixels. In [14], different solutions based on high degree B-spline interpolation, low pass filter, and stochastic integration are discussed. Salvado *et al.* [21] propose a constant variance filter for linear interpolation. Thévenaz *et al.* [22] develop a family of interpolation bases which can maintain the constant-variance property.



(a) (b) (c) (d)

Fig. 1. Brain images from the retrospective image registration evaluation (R.I.R.E.) project. (a) T1 weighted MRI. (b) T2 weighted MRI. (c) PD weighed MRI. (d) CT.

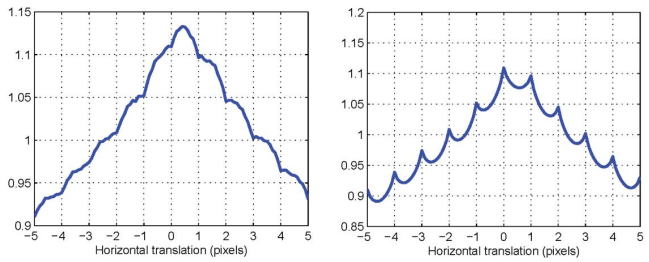
Dowson *et al.* [23] apply the nonparametric windows in order to obtain the PDF histograms more accurately than Parzen window. Bilinear interpolation is used to compute the 2D joint PDF, which leads to complex equations with more than 100 terms. Half-bilinear interpolation is then introduced to obtain a simpler implementation. Extension to 3D is not given. Joshi *et al.* [24] propose a simplified computation for probability density functions using planar interpolation for the nonparametric windows methods in 2D and 3D. However, there is no report on the smoothness of the resulting mutual information function and how it performs in rigid image registration. Rajwade *et al.* [25] derive analytic formulas for continuous marginal and joint densities in 2D and 3D based

Manuscript received April 17, 2012; revised December 24, 2012; accepted February 18, 2013. Date of publication March 7, 2013; date of current version May 24, 2013. This work was supported by the Natural Sciences and Engineering Research Council of Canada. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Charles Creusere.

The authors are with the David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: l8xu@uwaterloo.ca; jwlwan@uwaterloo.ca; tbian@uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2251644



(a) (b)

Fig. 2. (a) Mutual information computed using IM model. (b) Mutual information computed using PV model. The target and the template are T1-weighted MRI and CT from Fig. 1, respectively.

on geometric arguments. Regions of constant intensities are treated by slight blurring of the image.

In this paper, we propose a new method to avoid interpolation artifacts. In this approach, we will consider images as functions defined on a continuous image domain rather than on discrete pixels. Accordingly, we derive an analytic method for constructing PDFs in 1D for piecewise linear image functions, and then compute the mutual information based on continuous random variables. We note that the use of a continuous method has also been discussed in [10], [15], [23], [25]. We extend the idea of 1D to higher dimensions by applying the 1D technique to each dimension separately. We show in Section VI that our method is more robust in terms of success rate compared to the interpolation-based and partial volume methods when different initial guesses are used. We compare our model with the artifact-reducing methods nonparametric windows (NP) [23], third order generalized partial volume (GPVE) [7], and Parzen window (PZW) [19] on the accuracy of registering 2D images with different initial guesses. We also compare our method with the GPVE method [7] on 15 3D image datasets and our method shows better accuracy in 10 of the cases.

Interpolation artifacts have also been observed in non-rigid registration [14], [26]. A "local" mutual information model based on image gradients is developed for deformable registration, but the resulting similarity measure shows artifacts in translational misalignment [27]. In this paper, we will primarily focus on rigid multimodality registration. Also, the proposed method assumes that the images to be registered are of the same size and the same resolution.

The paper is organized as follows. Section II defines the mutual information formula and briefly describes various popular models. In Sections III and IV, we discuss the continuous model in one, two, and three dimensions. In particular, we explain how the PDF and joint PDF, entropies, and mutual information are computed. The complexity of our model is also analyzed. Section V discusses the relationship between our method and the standard partial volume model, and an analysis is given to explain the typical piecewise convex shape of the mutual information function generated by the partial volume model. In Section VI, numerical experiments are presented to illustrate the smoothness property of our model, as well as the robust and accurate convergence of numerical optimization methods. Finally, concluding remarks are made in Section VII.

II. Interpolation and Partial Volume Models

Given the target image G and the template image F, image registration can be formulated as a maximization problem:

$$s = \arg\max_{\tau} \ \Gamma(G, \phi(F, \tau)) \tag{1}$$

where ϕ is a transformation with parameter set τ, and Γ is chosen to be mutual information which is one of the most widely used measures for multimodality image registration [28].

The mutual information, $M(X, Y)$, of two images X and Y is defined by

$$M(X, Y) = H(X) + H(Y) - H(X, Y)$$

where $H(X)$ and $H(Y)$ are the entropies of X and Y, $H(X, Y)$ is their joint entropy, and they are defined as

$$H(X) = -\sum_{i=1}^{N^X} p_i^X \log p_i^X, \quad H(Y) = -\sum_{i=1}^{N^Y} p_i^Y \log p_i^Y,$$
$$H(X, Y) = -\sum_{i=1}^{N^X}\sum_{j=1}^{N^Y} p_{i,j}^{XY} \log p_{i,j}^{XY}. \tag{2}$$

Here, $\{p_i^X\}_{i=1}^{N^X}$ ($\{p_i^Y\}_{i=1}^{N^Y}$) is the probability distribution of the discrete random variable associated with image X (Y), assuming there are N^X (N^Y) possible observations. $\{p_{i,j}^{XY}\}$ is the joint probability distribution of X and Y.

The probability distributions are typically estimated by a histogram process which keeps a frequency count of the image intensity values defined at the pixel locations. One standard approach is to resample the intensity values of the transformed template at the pixel locations of the target by interpolation such as nearest neighbor, linear, and cubic splines [16]. However, interpolation artifacts arise in the case of perfect (or close to) grid alignment and have many local maxima and minima; see Fig. 2 (left).

The partial volume model [3], [5] addresses this issue by updating the frequency counts of several histogram bins by adding the corresponding interpolation weights. This approach improves the smoothness but still has artifacts or kinks at integer translation positions [13]; see Fig. 2 (right), which we will explain in Section V.

III. Continuous Method: 1D

To address the interpolation artifact issue, our approach is to consider intensity values defined not just at pixel locations but also between pixels. We will consider image intensities as continuous random variables and derive the corresponding formulas for computing the PDFs, entropies, and mutual information. We will describe the model in one dimension here and extend the model to higher dimensions in Sections IV-B and IV-C.

A. Probability Density Functions

Let the resolution of both the target and the template images be N, and the 1D arrays $\{G_i\}_{i=0}^{N-1}$ and $\{F_i\}_{i=0}^{N-1}$ be the intensity values of the target and template images, respectively.

We consider the intensity values as the sample values of some continuous image functions denoted by

$$F : \Omega \to \mathbb{R}, \quad G : \Omega \to \mathbb{R}$$

where Ω is a continuous domain of image pixels, $F(x)$ is the intensity of the template image at $x \in \Omega$, and $G(x)$ is the intensity of the target image. If we let $\{x_i\}_{i=0}^N$ be the pixel locations for the discrete images, then $F(x_i) = F_i$ and $G(x_i) = G_i$. For simplicity, we assume $\Omega = [0, 1]$, so the pixel size $h = 1/N$, and that F and G are periodic; i.e., $F_0 = F_N$ and $G_0 = G_N$.

Given $\{F_i\}_{i=0}^N$ and $\{G_i\}_{i=0}^N$, we define $F(x)$ and $G(x)$ to be the linear interpolants for $\{x_i, F_i\}_{i=0}^N$ and $\{x_i, G_i\}_{i=0}^N$, respectively. Higher order interpolants can also be considered, but we find that linear interpolation is sufficient to obtain a smooth mutual information function. We note that images are also treated as piecewise linear functions in [23]. However, the construction of the probability density functions are different as explained below.

1) Probability Density Functions of Images: We describe how the PDF for the piecewise linear target image $G(x)$ is defined. Consider the subinterval $[x_{i-1}, x_i)$, $i = 1, \ldots, N$. We first assume that $G_{i-1} < G_i$.

In standard discrete models, the PDF is defined as the histogram of the intensity distribution; i.e., each bin contains the normalized frequency of the pixels whose intensity values are in the range of the bin. In our model, we define the appropriate cumulative distribution and probability density function in the continuous case. For any $\alpha \in \mathbb{R}$, the cumulative distribution function, $C_i(\alpha)$, which is formally defined as the probability of intensity value being in the interval $(-\infty, \alpha]$, should be equal to the proportion of the measure of the image domain whose intensity value is in that subinterval over the length of the subinterval $[x_{i-1}, x_i)$. Since the image function $G(x)$ is linear on the subinterval $[x_{i-1}, x_i)$, $C_i(\alpha)$ should also be a linear function on the interval $[G_{i-1}, G_i)$. More precisely,

$$C_i(\alpha) = \begin{cases} 0, & \text{if } \alpha \in (-\infty, G_{i-1}) \\ \frac{\alpha - G_{i-1}}{G_i - G_{i-1}}, & \text{if } \alpha \in [G_{i-1}, G_i) \\ 1, & \text{if } \alpha \in [G_i, +\infty). \end{cases}$$

Let p_i^G be the PDF corresponding to the subinterval $[x_{i-1}, x_i)$. Then it is defined as the derivative of the cumulative distribution function:

$$p_i^G(\alpha) = \begin{cases} 0, & \text{if } \alpha \in (-\infty, G_{i-1}) \\ \frac{1}{G_i - G_{i-1}}, & \text{if } \alpha \in [G_{i-1}, G_i) \\ 0, & \text{if } \alpha \in [G_i, +\infty). \end{cases} \tag{3}$$

In the case where $G_{i-1} > G_i$, p_i^G is defined similarly as

$$p_i^G(\alpha) = \begin{cases} 0, & \text{if } \alpha \in (-\infty, G_i] \\ \frac{1}{G_{i-1} - G_i}, & \text{if } \alpha \in (G_i, G_{i-1}] \\ 0, & \text{if } \alpha \in (G_{i-1}, +\infty). \end{cases} \tag{4}$$

A special case needed to be considered is when $G_{i-1} = G_i$. By the analysis above, it is clear that $p_i^G(\alpha) = 0$ for any $\alpha \neq G_i$. On the other hand, since p_i^G is a PDF, $\int_{-\infty}^{+\infty} p_i^G(\alpha) d\alpha = 1$. Consequently, p_i^G should be a Dirac delta function in this case:

$$p_i^G(\alpha) = \delta(\alpha - G_i). \tag{5}$$

Now, define the index sets $\mathcal{J}^G = \{j \mid G_{j-1} < G_j\}$, $\mathcal{K}^G = \{k \mid G_{k-1} > G_k\}$, and $\mathcal{L}^G = \{l \mid G_{l-1} = G_l\}$. Then the PDF for the target image, p^G, is defined by combining the three cases (3)–(5):

$$\begin{aligned} p^G(\alpha) &= \frac{1}{N} \sum_{i=1}^N p_i^G(\alpha) \\ &= \frac{1}{N} \left(\sum_{j \in \mathcal{J}^G} \frac{\mathbf{1}_{[G_{j-1}, G_j)}(\alpha)}{G_j - G_{j-1}} + \sum_{k \in \mathcal{K}^G} \frac{\mathbf{1}_{(G_k, G_{k-1}]}(\alpha)}{G_{k-1} - G_k} \right. \\ &\qquad \left. + \sum_{l \in \mathcal{L}^G} \delta(\alpha - G_l) \right) \end{aligned} \tag{6}$$

where $\mathbf{1}_{[G_{j-1}, G_j)}$ denotes the characteristic function on the interval $[G_{j-1}, G_j)$. We note that the first two terms resemble the PDF defined in [23], but the last term is new here which takes into account the contribution of constant intensities by the use of Dirac delta functions.

Since the PDF construction above does not depend on the pixel locations, it is independent of any translation. Thus, the PDF for the transformed template image can be written exactly the same way as the target image:

$$\begin{aligned} p^F(\alpha) &= \frac{1}{N} \sum_{i=1}^N p_i^F(\alpha) \\ &= \frac{1}{N} \left(\sum_{j \in \mathcal{J}^F} \frac{\mathbf{1}_{[F_{j-1}, F_j)}(\alpha)}{F_j - F_{j-1}} + \sum_{k \in \mathcal{K}^F} \frac{\mathbf{1}_{(F_k, F_{k-1}]}(\alpha)}{F_{k-1} - F_k} \right. \\ &\qquad \left. + \sum_{l \in \mathcal{L}^F} \delta(\alpha - F_l) \right) \end{aligned}$$

where the index sets $\mathcal{J}^F$, $\mathcal{K}^F$ and $\mathcal{L}^F$ are defined as those for the target image. We note that not just the intensity values G_{i-1}, G_i, F_{i-1}, and F_i are used in the construction of the PDF as in the discrete case, but in fact the entire intervals between adjacent pixel nodes contribute to the probability density function. This is a distinguishing feature between the standard discrete model and our model.

2) Joint Probability Density Function: By construction, the target image function $G(x)$ is linear on the pixel interval $[x_{j-1}, x_j)$. Suppose the template image function $F(x)$ is translated by τ ($0 \leq \tau < h$), as shown in Fig. 3(a). The intensity values between pixels are no longer a linear function. One approach [23] is to resample the template image at the grid locations. The new piecewise linear function (solid line) and the true transformed template (dashed line) are shown in Fig. 3(b).

Resampling has an advantage that the interpolated transformed template and target images are both linear on the pixel interval $[x_j, x_{j+1})$ which makes the computation of the analytic PDF easier. However, the image is then altered between pixels (Fig. 3(b)), which raises the issue of computing

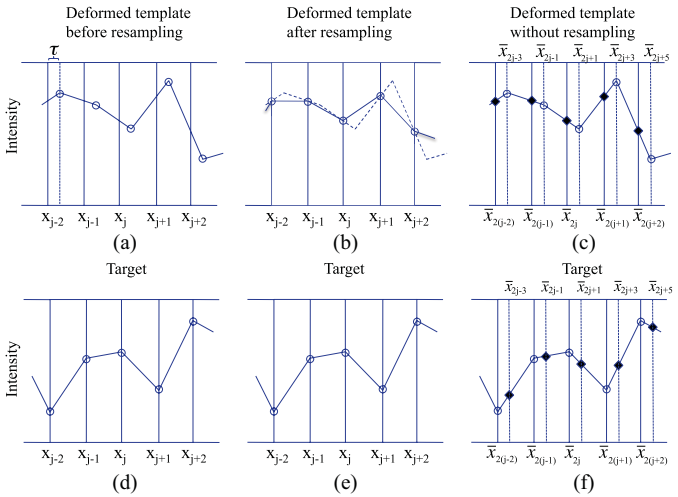


Fig. 3. (a) Transformed template images before resampling. (b) Transformed template images after resampling. (c) Transformed template images without resampling. (d)–(f) Target images. Circles and black dots: subintervals where the transformed template and target are both linear.

the continuous joint PDF that requires not just the pixel values but also the intensities between pixels. Here, we use another approach to achieve the same goal without resampling; i.e. the transformed template remains unchanged. The key is that the image function is piecewise linear and hence the intensity values are all predefined, not only at the pixel locations but also on all over the continuous image domain. Consequently, there is no need for interpolation.

More precisely, consider a new set of points $\{\bar{x}_i\}$ such that $\bar{x}_{2j} = x_j$ and $\bar{x}_{2j+1} = x_j + \tau$, $j = 0, \ldots, N-1$; see Fig. 3(c). (Note that even if $\tau \geq h$ or $\tau < 0$, the above definition still works by a slight modification.) Using this partition, the target and the transformed template image functions are *both* linear on $[\bar{x}_{i-1}, \bar{x}_i)$. Let the values of $G(x)$ at $\bar{x}_{i-1}$ and $\bar{x}_i$ be $\bar{G}_{i-1}$ and $\bar{G}_i$, respectively. Similarly, $\bar{F}_{i-1}$ and $\bar{F}_i$ are the values of the transformed template at the same end points.

We first consider the case where $\bar{G}_{i-1} \neq \bar{G}_i$ or $\bar{F}_{i-1} \neq \bar{F}_i$. Without loss of generality, assume $\bar{G}_{i-1} \leq \bar{G}_i$ and $\bar{F}_{i-1} \leq \bar{F}_i$. In the discrete case, the intensity pairs $(\bar{G}_{i-1}, \bar{F}_{i-1})$ and $(\bar{G}_i, \bar{F}_i)$ generally contribute to the frequency counts of two (or eight in the case of partial volume) histogram bins. In the continuous case, the entire line segment, $\mathcal{C}_i$, with end points at $(\bar{G}_{i-1}, \bar{F}_{i-1})$ and $(\bar{G}_i, \bar{F}_i)$, contributes to the joint PDF. Let p_i^{FG} be the joint PDF associated with the image functions on the subinterval $[\bar{x}_{i-1}, \bar{x}_i)$. We want p_i^{FG} to be nonzero on the line segment, it will give the marginal PDFs, which are defined in (3) or (4), as we integrate along one direction, and its double integral be 1. Let $\Delta\bar{G}_i = \bar{G}_i - \bar{G}_{i-1}$ and $\Delta\bar{F}_i = \bar{F}_i - \bar{F}_{i-1}$. We define p_i^{FG} as

$$p_i^{FG}(\alpha, \beta) = \delta(\Delta\bar{G}_i(\beta - \bar{F}_{i-1}) - \Delta\bar{F}_i(\alpha - \bar{G}_{i-1}))$$

if $\alpha \in [\bar{G}_{i-1}, \bar{G}_i)$ or $\beta \in [\bar{F}_{i-1}, \bar{F}_i)$, and $p_i^{FG}(\alpha, \beta) = 0$ otherwise. Then integrating p_i^{FG} along β gives

$$\begin{aligned}\int_{-\infty}^{+\infty} p_i^{FG}(\alpha, \beta)d\beta &= \int_{-\infty}^{+\infty} \delta(\Delta\bar{G}_i(\beta - \bar{F}_{i-1}))\, d\beta \\ &= \frac{1}{\Delta\bar{G}_i}\int_{-\infty}^{+\infty} \delta(\beta - \bar{F}_{i-1})\, d\beta \\ &= \frac{1}{\Delta\bar{G}_i}\end{aligned}$$

where α is in the interval $[\bar{G}_{i-1}, \bar{G}_i)$. If α is outside of the interval, $p_i^{FG}(\alpha, \beta) = 0$. Thus the integral of p_i^{FG} yields the marginal PDF p_i^G (3) or (4). By a similar calculation, the integral of p_i^{FG} along α will yield the marginal PDF p_i^F. Consequently, the double integral of p^{FG} is one.

In the case that $\bar{G}_{i-1} = \bar{G}_i$ and $\bar{F}_{i-1} = \bar{F}_i$, the line segment degenerates to a point. In this case, p_i^{FG} is defined as

$$p_i^{FG}(\alpha, \beta) = \delta(\alpha - \bar{G}_i, \beta - \bar{F}_i)$$

which is similar to (5).

Finally, the joint PDF is defined by combining all the p_i^{FG}'s. Let the index sets $\mathcal{I}^{line} = \{i \mid \bar{G}_{i-1} \neq \bar{G}_i \text{ or } \bar{F}_{i-1} \neq \bar{F}_i\}$ and $\mathcal{I}^{point} = \{i \mid \bar{G}_{i-1} = \bar{G}_i \text{ and } \bar{F}_{i-1} = \bar{F}_i\}$. The joint PDF is then given by

$$\begin{aligned}&p^{FG}(\alpha, \beta) \\ &= \sum_{i=1}^{2N} \omega_i\, p_i^{FG}(\alpha, \beta) \\ &= \sum_{i \in \mathcal{I}^{line}} \omega_i \delta(\Delta\bar{G}_i(\beta - \bar{F}_{i-1}) - \Delta\bar{F}_i(\alpha - \bar{G}_{i-1})) \\ &\quad + \sum_{i \in \mathcal{I}^{point}} \omega_i \delta(\alpha - \bar{G}_i, \beta - \bar{F}_i) \qquad (7)\end{aligned}$$

where $\omega_i = \bar{x}_i - \bar{x}_{i-1}$. By properties of p_i^{FG}, it can be shown that the double integral of p^{FG} is 1, and integrating p^{FG} in α (β) will produce the marginal PDF for the template (target) image, which is consistent with the property of a joint PDF.

B. Entropies of the Target and Transformed Template

In our model, the image functions and PDFs are computed as continuous functions. Thus, the entropy of the target image function is given by the integral:

$$H(G) = -\int_{-\infty}^{+\infty} p^G(\alpha) \log p^G(\alpha)\, d\alpha.$$

While the PDF for the target image, p^G, can be written analytically in (6), it is in general difficult to compute the entropy exactly. Thus, we approximate the entropy numerically. Note that standard numerical integration methods would be computationally expensive to apply here since p^G is generally not continuous at $\{G_i\}$ and the discontinuities are difficult to be determined *a priori*. Other strategies need to be exploited.

Divide the range of the intensity value of the target image into N_D subintervals, or cells, $\{D_i\}$ of size h_D. We approximate p^G by a constant function on each cell D_i where the constant value is computed as

$$q_i = \frac{1}{h_D}\int_{D_i} p^G(\alpha)\, d\alpha.$$

There are two advantages to approximate the PDF in this way. First, the approximate p^G as well as $p^G \log p^G$ are both piecewise constant and hence the computation of the entropy integral is much simpler. Second, it is easy to verify that the integral of the approximate PDF over the whole domain is 1, which is consistent with the definition of a PDF.

By (6), q_i can be written as

$$q_i = \frac{1}{Nh_D}\int_{D_i}\sum_{k=1}^{N} p_k^G(\alpha)\,d\alpha = \frac{1}{Nh_D}\sum_{k=1}^{N}\int_{D_i} p_k^G(\alpha)\,d\alpha. \quad (8)$$

Note that $\int_{D_i} p_k^G(\alpha)d\alpha$ is easy to compute since $p_k^G(\alpha)$ is either a constant function or a Dirac delta function. Finally, we approximate the entropy of the target image by

$$\begin{aligned} H(G) &= -\sum_{i=1}^{N_D}\int_{D_i} p^G(\alpha)\log p^G(\alpha)\,d\alpha \\ &\approx -\sum_{i=1}^{N_D}\int_{D_i} q_i \log q_i\,d\alpha \\ &= -h_D\sum_{i=1}^{N_D} q_i \log q_i. \end{aligned} \quad (9)$$

The entropy of the transformed template image function, $H(F)$, can be computed similarly.

In practice, a small h_D (e.g. $h_D = 1/16$) is sufficient to yield smooth mutual information; see Section VI. In general, we do not have a formula for choosing h_D. One possibility is based on the image gradient which will be a topic for future work. We also remark that the formula (9) resembles the standard discrete entropy formula (2). However, the computation of q_i is derived from the analytical form of the PDF p^G, which is different from the usual p_i derived from histogramming. Also, the h_D factor is there to account for the size of the interval D_i for the continuous entropy.

C. Joint Entropy and Mutual Information

The joint entropy in the continuous model is given by the double integral:

$$H(F,G) = -\iint_{\mathbb{R}^2} p^{FG}(\alpha,\beta)\log p^{FG}(\alpha,\beta)\,d\alpha\,d\beta.$$

It can be computed using a similar idea as above. Divide the two-dimensional domain of the joint PDF into cells $\{D_{i,j}\} = D_i \times D_j$ (see Section III-B). Similar to the 1D case, on each cell, we approximate the joint PDF by a constant function where the constant value is

$$q_{i,j} = \frac{1}{h_D^2}\iint_{D_{i,j}} p^{FG}(\alpha,\beta)\,d\alpha\,d\beta. \quad (10)$$

Since the joint PDF is composed of Dirac delta functions, the double integral is easy to compute. Finally the joint entropy is approximated as

$$\begin{aligned} H(F,G) &= -\sum_{i,j=1}^{N_D}\iint_{D_{i,j}} p^{FG}(\alpha,\beta)\log p^{FG}(\alpha,\beta)\,d\alpha\,d\beta \\ &\approx -\sum_{i,j=1}^{N_D}\iint_{D_{i,j}} q_{i,j}\log q_{i,j}\,d\alpha\,d\beta \\ &= -h_D^2\sum_{i,j=1}^{N_D} q_{i,j}\log q_{i,j}. \end{aligned} \quad (11)$$

We note again that although the joint entropy formula (11) appears similar to the standard discrete joint entropy formula (2), the way to compute $q_{i,j}$ is fundamentally different. Finally, the mutual information can easily be calculated as

$$M(F,G) = H(F) + H(G) - H(F,G).$$

Notice that $M(F,G)$ is a function of the translation parameter τ since the joint PDF p^{FG} and hence the joint entropy $H(F,G)$ depend on τ. The entropies of the transformed template image, $H(F)$, and the target image, $H(G)$, are independent of τ as the images are assumed periodic (thus F and G completely overlap) and of the same size.

IV. Continuous Method: 2D and 3D

We first generalize the 1D model to two dimensions. Let the 2D arrays $\{G_{i,j}\}$ and $\{F_{i,j}\}$ be the intensity values of the target and template images, respectively, and the resolution of the images be $N_x \times N_y$ with pixel size $h = h_x \times h_y$. Denote the image functions defined on a continuous domain by

$$F(x,y) : \Omega \to \mathbb{R}, \quad G(x,y) : \Omega \to \mathbb{R}$$

where Ω is a continuous domain of image pixels, $F(x,y)$ is the intensity of the template image at $(x,y) \in \Omega$, and $G(x,y)$ the target image. If $\{(x_i, y_j)\}$ are the pixel locations of the discrete images, then $F(x_i, y_j) = F_{i,j}$ and $G(x_i, y_j) = G_{i,j}$. As in 1D, $F(x,y)$ and $G(x,y)$ are the bilinear interpolants of the transformed template and target images, and they are assumed to be periodic.

In principle, analytic PDFs for 2D image functions can be constructed similarly as in 1D. However, the construction and resulting PDF formulas can be very complex. As shown in [23], the computation of the 2D marginal PDFs requires separate implementation for 6 special cases. For joint PDFs, the numerous geometric configurations lead to a final equation containing more than 100 terms. The authors simplify the method by half-bilinear interpolation which involves standard triangle rendering techniques from computer graphics.

Instead of computing the PDFs analytically, our approach is to extend the 1D method to higher dimensions by piecewise continuous approximation, which preserves the advantages of the continuous construction. Specifically, we solve the image registration problem (1) in 2D and 3D by searching the optimal solution in one dimension each time. Thus, we reduce the registration problem for 2D and 3D images to a problem for 1D images. In this way, we only need to compute mutual information sufficiently accurately for the corresponding 1D problem without constructing the full 2D or 3D analytic PDFs.

A. Optimization Scheme

Our approach is to optimize the translation parameters in the direction parallel to each axis separately. More precisely, the mutual information function is first maximized in the x-direction, using a standard optimization scheme such as Powell's [3] (i.e. Brent's method in 1D) or Nelder-Mead [29] methods. Then, taking the local optimal translation in x, the transformed template image is resampled on the image grid. We note that this resampling procedure is not part of the

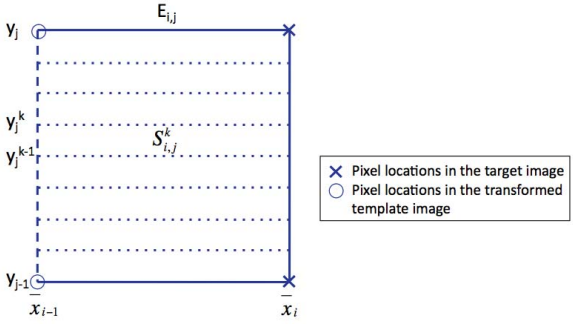


Fig. 4. Each subdomain $E_{i,j}$ is divided into stripes with equal width.

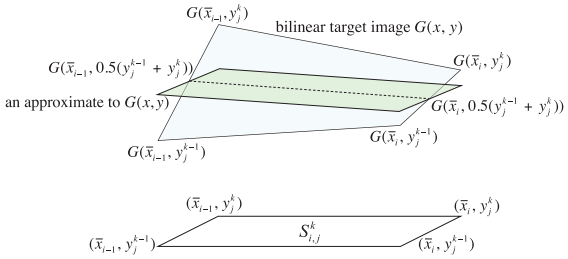


Fig. 5. Bilinear target image function $G(x, y)$ is approximated by function which is linear in x and constant in y on stripe $S^k_{i,j}$.

PDF calculation and hence it has no effect on the smoothness of mutual information. The maximization is then carried out in the y-direction, followed by a similar procedure in the z-direction. Afterwards, the 3 rotation angles are optimized at once. One sweep in each direction and 3 rotations would not normally have obtained the final result. The procedure is repeated iteratively until it converges. We note that the 1D optimizations in the x, y, z direction is similar to the coordinate optimization methods [30]. Instead of using the search directions picked by a line search method such as Powell, we fix the search directions to be in the x, y, z directions. The step of optimizing the 3 rotation angles is not exactly a line search, but one can view it as a block coordinate.

We observe that the mutual information function has only one maximum and it appears concave near the maximum using our method; see Fig. 7–10. The maximization in each principal direction would only improve the solution. Thus, the iteration will eventually converge to the exact solution. In our experience, it typically converges in less than 10 iterations (each iteration consists of three 1D optimization problems and one 3D optimization problem).

There are two major advantages of this approach. One is that only a 1D optimization is solved in each translation direction (3D optimization for rotations). In our experiments, it usually converges in a few number of iterations. But more importantly, the other advantage is that the translation is optimized in one direction each time. As such, we can relax the accuracy of the PDF computation in the other directions without compromising the smoothness of the mutual information function.

B. Piecewise Continuous Approximation of 2D Images

Without loss of generality, we only discuss the case of spatial translation in horizontal direction. Suppose the template image has been translated by τ_x. As in the construction of the joint PDF in 1D, we consider the nodal points $\{\bar{x}_i\}$ derived from the translation τ_x, and $\{y_j\}$. The set of nodal points $\{(\bar{x}_i, y_j)\}$ divides the image domain into subdomains:

$$E_{i,j} = \left\{(x, y) | x \in [\bar{x}_{i-1}, \bar{x}_i), y \in [y_{j-1}, y_j)\right\}.$$

We then divide each $E_{i,j}$ into several stripes with equal width as shown in Fig. 4. Let

$$y_j^k = y_{j-1} + k\, h_S \quad k = 0, \ldots, N_S$$

where N_S is the number of stripes and h_S is the thickness of the stripe. Each stripe $S^k_{i,j}$, $k = 1, \ldots, N_S$, is defined as

$$S^k_{i,j} = \left\{(x, y) \,|\, x \in [\bar{x}_{i-1}, \bar{x}_i),\ y \in [y_j^{k-1}, y_j^k)\right\}.$$

When h_S is small, then the image functions on each stripe can be approximated as functions which are linear horizontally and constant vertically; see Fig. 5. For instance, the target image on the stripe $S^k_{i,j}$ can be approximated by

$$G(x, y) \approx \frac{\bar{x}_i - x}{\bar{x}_i - \bar{x}_{i-1}} G(\bar{x}_{i-1}, \bar{y}_j^k) + \frac{x - \bar{x}_{i-1}}{\bar{x}_i - \bar{x}_{i-1}} G(\bar{x}_i, \bar{y}_j^k)$$

where $\bar{y}_j^k = (y_j^{k-1} + y_j^k)/2$. Since the approximate image on $S^k_{i,j}$ is constant in y and linear in x, its corresponding PDF, $p^{G,k}_{i,j}$, can be defined as in the 1D case; see Section III-A. The PDF for the subdomain $E_{i,j}$ is defined by combining the approximate PDFs on each stripe,

$$p^G_{i,j} \approx \frac{1}{N_S} \sum_{k=1}^{N_S} p^{G,k}_{i,j}. \tag{12}$$

Finally the PDF for the target image is approximated by combining the PDFs for each subdomain with weight proportional to the subdomain area,

$$p^G = \frac{(\bar{x}_i - \bar{x}_{i-1})(y_j - y_{j-1})}{N_x h_x N_y h_y} \sum_{i=1}^{2N_x} \sum_{j=1}^{2N_y} p^G_{i,j}. \tag{13}$$

Hence, we can compute the 2D PDF by computing a sequence of 1D PDFs defined on the stripes.

The PDF for the transformed template image can be computed similarly as p^G. Following a similar procedure, the 2D joint PDF comes down to computing the joint PDF on each stripe for which we again apply the 1D construction. Once the PDFs are defined, the entropies and mutual information can be computed as in the 1D case, and the details are omitted.

When the 2D template image is translated horizontally, it is similar to the translation of 1D images where mutual information can be computed accurately by the 1D method. Since there is no change in the vertical direction, there is no need of resampling and hence no danger of interpolation artifacts. Ideally, a large N_S should be used for accuracy, but it will also result in high computational costs. In practice, we find that small N_S (e.g. $N_S = 1$) is sufficient to obtain a smooth mutual information function; see Section VI-A.

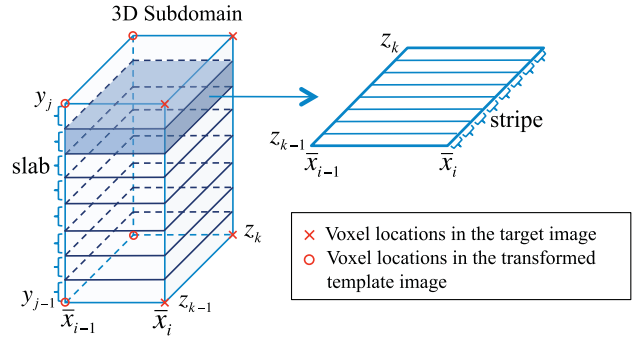


Fig. 6. Each 3D subdomain is divided into slabs, which are considered as 2D planes. Each plane is further divided into stripes.

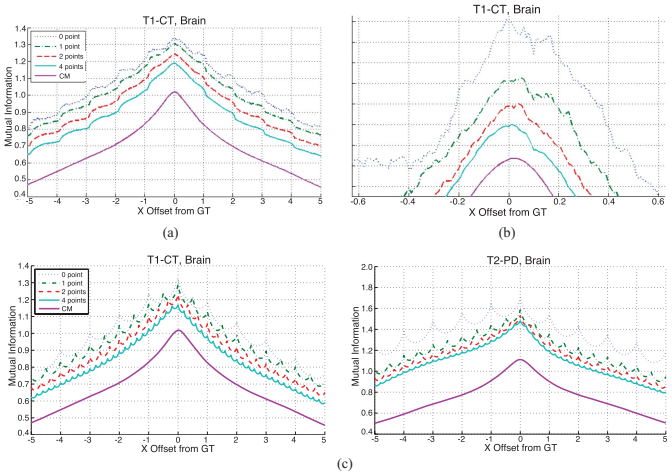


Fig. 7. (a) Trace plots of mutual information computed by interpolation-based method and our method (CM) as x translation parameter (in pixel) is varied for T1-CT image pair. (b) Magnified plot near ground truth. (c) Trace plots of mutual information computed by PV method and our method for T1-CT and T2-PD image pairs. Zero to four points are inserted between pixels.

Compared to computing the PDFs analytically, our approach is much simpler and yet no interpolation artifacts occur.

In 2D, the transformation parameters also include rotation in addition to translations. In our experience, MI as a function of rotation behaves smooth locally when the rotation angle is away from 0° and less than 90°. Similar observations have also been reported in the literature (e.g. [14], [16]). To simplify the algorithm, we separate the optimization of the translation and rotation parameters. For translation, we apply the method described above. For rotation, we apply a standard interpolation-based method. Although nonsmoothness of MI may occur near 0°, it does not seem to affect the convergence of our optimization methods.

C. Piecewise Continuous Approximation of 3D Images

The extension to 3D is as follows. Consider translation in the x-direction by τ_x. As in the construction of the joint PDFs in 1D and 2D, we consider the nodal points $\{\bar{x}_i\}$ derived from the translations τ_x, $\{y_j\}$, and $\{z_k\}$. The set of nodal points $\{(\bar{x}_i, y_j, z_k)\}$ divides the image domain into subdomains.

As in 2D, we divide a subdomain into slabs (stripes in 2D) along the y direction and approximate the image functions as piecewise constant functions in that direction. These slabs can

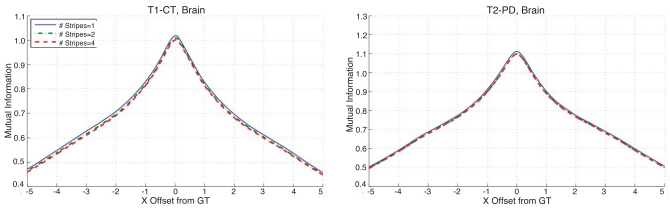


Fig. 8. Trace plots of mutual information as x translation parameter (in pixel) is varied for different image pairs with different numbers of stripes.

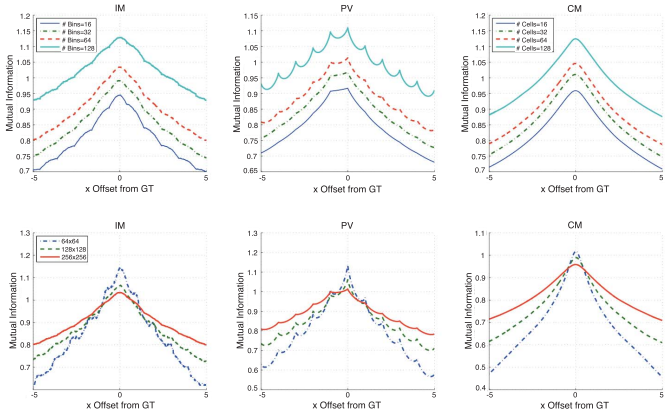


Fig. 9. Trace plots of mutual information as x translation parameter (in pixel) is varied with different number of bins (or cells) (first row) and image resolutions (second row) for IM, PV, and CM.

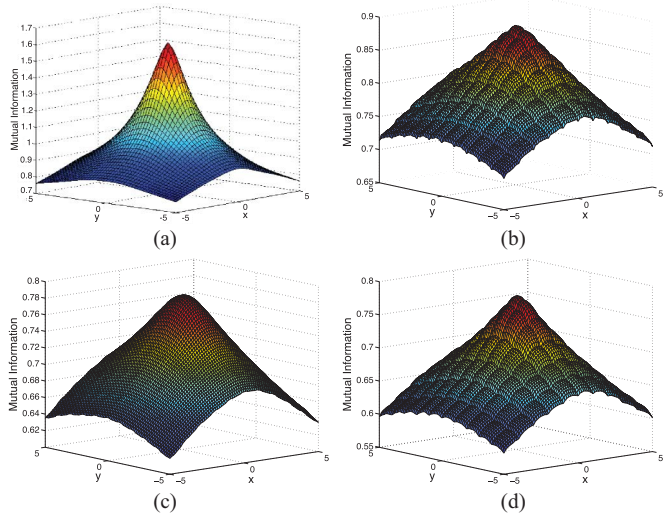


Fig. 10. Trace plots of mutual information computed by different models as x and y translations (in pixel) are varied. (a) CM. (b) NP. (c) GPVE. (d) PZW.

be considered as 2D planes and hence the problem is now reduced to the 2D case. On each plane, we further divide it into stripes along the z direction, as shown in Fig. 6. Thus, a subdomain is divided into planes of stripes and on each stripe, the image function is approximated by a function which is linear in x and constant in y and z. We may again apply the 1D formulas for computing the PDFs on each stripe. The final PDFs are computed by adding all the stripes together as in (12) and (13). Finally, the entropies and mutual information are computed as in 1D, and the details are omitted.

The accuracy in the y and z directions depends on the number of slabs and stripes used. Since in each optimization,

only one translation is considered, there is no translations in the other directions. We found that a small number of slabs and stripes (1 is used for all our tests) are sufficient to yield a smooth mutual information function; see Fig. 8.

D. Complexity

We first consider the 1D case. The PDF for a target image is approximated by $\{q_i\}$ which are computed using (8). Note that we do not actually compute each q_i one by one. Instead, we go through each of the p_k^G's and assign the appropriate portion $\int_{D_i} p_k^G(\alpha)\, d\alpha$ to the q_i's. Each p_k^G is obtained from a pixel subinterval. Thus, one sweep of all the pixel intervals will give all the $\{q_i\}$. The resulting complexity is $O(cN)$ where c denotes the average number of cells where p_k^G is nonzero. Typically, $c = 2$ or 3 if the image is not too noisy. Then by (9), the complexity of computing $H(G)$ is $O(N_D)$. N_D is usually chosen to be 8, which is much smaller than N. Thus the total complexity for computing the target (as well as template) image entropy is $O(cN)$.

By a similar argument, the complexity of computing $\{q_{i,j}\}$ for the joint entropy is $O(cN)$. To compute the sum for $H(F, G)$ takes $O(N_D^2)$ work. Considering that $N_D^2 \ll N$, the complexity for computing the joint entropy and hence the mutual function is also $O(cN)$.

Now in 3D (and also similarly in 2D), we compute the PDF of an image by combining the PDFs on all the stripes. Since we maximize the mutual information function in one of the directions at a time, the subdomain is generally divided into two in our approach. Hence, the complexity is $O(2cN_S^2 N_x N_y N_z)$, where N_S is the number of slabs as well as the number of stripes, and N_x, N_y, and N_z are the resolutions in different dimensions of the 3D image. N_S is usually a small number; see Figure 8. In our experiments, we used $N_S = 1$. The computation of the entropies and mutual information will require a further complexity of $O(N_D^3)$. Since $N_D^3 \ll N_x N_y N_z$, the total complexity is $O(2cN_x N_y N_z)$.

As for comparison, the complexity for the interpolation and partial volume models is $O(8N_x N_y N_z + N_B^3)$, where N_B is the number of bins in each dimension. The factor 8 originates from either the interpolation of the transformed template image or the weighted updates in the joint probability distribution. Considering that $N_x N_y N_z \sim N_B^3$ in practice, the complexity of the proposed method is comparable to the standard models.

In the partial volume model, the weights for updating the probabilities is determined by the linear kernel. Chen *et al.* [7] generalize the linear kernel to higher order B-spline kernels. Although the smoothness of the final mutual information function has been improved, the complexities get higher: $O(27N_x N_y N_z + N_B^3)$ for the second order B-spline and $O(64N_x N_y N_z + N_B^3)$ for the third order B-spline.

V. Relationship With Partial Volume Model

The partial volume model can be viewed as a modified version of the continuous method in which the nearest-neighbor interpolation is used to construct a piecewise constant image function. The constant value on each pixel (voxel) cell will then lead to a Dirac delta function in the PDF; see (5), and it contributes the same weight as the constant functions on the other pixel (voxel) cells. This is equivalent to the case in the partial volume model where each pixel (voxel) contributes to a bin with equal weight when the probability distribution is constructed. By a similar argument, the joint PDF computed using the piecewise constant function gives equivalent computation as in the partial volume model.

Now, consider q_i in (8) for computing the entropies of the images. Since p_k^G are all Dirac delta functions, the integral on the right-hand side is either 0 or 1, depending on whether the point where the delta function is nonzero is located inside D_i. Thus, if we choose D_i to be the bins in the histogram used by the partial volume model, then the summation on the right-hand side essentially counts the number of pixel/voxel intensity values in bin D_i. Hence, q_i is exactly the same as p_i in the standard formula (2), except for a constant factor $1/h_D$. More precisely, $h_D q_i = p_i$. This constant factor accounts for the fact that $\{q_i\}$ are constant values of a continuous PDF and $\{p_i\}$ are values of a discrete probability distribution. Similarly, $q_{i,j}$ in (10) is also the same as $p_{i,j}$ in (2), except for a constant factor $1/h_D^2$. Thus, the modified version of the continuous method with the piecewise constant image functions would give rise to piecewise constant marginal and joint PDFs whose values are equal to (modulo a constant scaling factor) the marginal and joint probability distributions of the partial volume model.

A. Analysis

We analyze the special shape of the mutual information function as shown in Fig. 2(b). Suppose a translation τ changes from $\tau_0 = n_0 h$ to $\tau_1 = (n_0 + 1)h$ for some integer n_0. By (2), the joint entropy can be written as

$$H(F, G) = -\sum_{i,j} \gamma\left(p_{i,j}^{FG}(\tau)\right)$$

where $\gamma(\alpha) = \alpha \log \alpha$. By construction of the partial volume model, it can be shown that $p_{i,j}^{FG}(\tau)$ is a linear function in τ (we omit the details here). Let $p_{i,j}^0 = p_{i,j}^{FG}(\tau_0)$ and $p_{i,j}^1 = p_{i,j}^{FG}(\tau_1)$. Then $p_{i,j}^{FG}(\tau)$ can be written as

$$p_{i,j}^{FG}(\tau) = \frac{\tau_1 - \tau}{h} p_{i,j}^0 + \frac{\tau - \tau_0}{h} p_{i,j}^1 \qquad \tau \in [\tau_0, \tau_1].$$

The first derivative of $\gamma(p_{i,j}^{FG}(\tau))$ with respect to τ is

$$\frac{d\gamma(p_{i,j}^{FG}(\tau))}{d\tau} = \frac{1}{h}\left(\log p_{i,j}^{FG}(\tau) + \frac{1}{\ln 2}\right)(p_{i,j}^1 - p_{i,j}^0).$$

Hence, the first derivative of $H(F, G)$ is

$$\frac{dH(F, G)}{d\tau} = -\frac{1}{h}\sum_{i,j}\left(\log p_{i,j}^{FG}(\tau) + \frac{1}{\ln 2}\right)(p_{i,j}^1 - p_{i,j}^0).$$

Let $\tau_2 = (n_0 + 2)h$ and $p_{i,j}^2 = p_{i,j}^{FG}(\tau_2)$. We have

$$\lim_{\tau \to \tau_1^-} \frac{dH(F, G)}{d\tau} = -\frac{1}{h}\sum_{i,j}\left(\log p_{i,j}^1 + \frac{1}{\ln 2}\right)(p_{i,j}^1 - p_{i,j}^0)$$

$$\lim_{\tau \to \tau_1^+} \frac{dH(F, G)}{d\tau} = -\frac{1}{h}\sum_{i,j}\left(\log p_{i,j}^1 + \frac{1}{\ln 2}\right)(p_{i,j}^2 - p_{i,j}^1).$$

In general, $p^1_{i,j} - p^0_{i,j} \neq p^2_{i,j} - p^1_{i,j}$ and hence

$$\lim_{\tau \to \tau_1^-} \frac{dH(F,G)(\tau)}{d\tau} \neq \lim_{\tau \to \tau_1^+} \frac{dH(F,G)(\tau)}{d\tau}.$$

Thus, we have shown that $H(F,G)$ is not smooth at $\tau = \tau_1$.

We further examine the second derivative of $\gamma(p^{FG}_{i,j})$ on the interval (τ_0, τ_1). Suppose $p^0_{i,j} \neq 0$ or $p^1_{i,j} \neq 0$. Then

$$p^{FG}_{i,j}(\tau) \neq 0 \quad \forall \tau \in (\tau_0, \tau_1).$$

We have

$$\frac{d^2\gamma(p^{FG}_{i,j}(\tau))}{d\tau^2} = \frac{1}{\ln 2} \cdot \frac{(p^1_{i,j} - p^0_{i,j})^2}{h^2 p^{FG}_{i,j}(\tau)} \geq 0.$$

Hence, $\gamma(p^{FG}_{i,j})$ is a convex function. If $p^0_{i,j} = p^1_{i,j} = 0$, then $p^{FG}_{i,j}(\tau) = 0$. Thus, $\gamma(p^{FG}_{i,j}(\tau)) \equiv 0$, which is clearly a convex function. Therefore, $\gamma(p^{FG}_{i,j}(\tau))$ is a convex function on (τ_0, τ_1). As a result,

$$\begin{aligned}\gamma(p^{FG}_{i,j}(\tau)) &= \gamma\left(\frac{\tau_1 - \tau}{h} p^0_{i,j} + \frac{\tau - \tau_0}{h} p^1_{i,j}\right) \\ &\leq \frac{\tau_1 - \tau}{h}\gamma(p^0_{i,j}) + \frac{\tau - \tau_0}{h}\gamma(p^1_{i,j})\end{aligned}$$

for any $\tau \in [\tau_0, \tau_1]$. Consequently,

$$\begin{aligned}&H(F,G)(\tau) \\ &= -\sum_{i,j} \gamma(p^{FG}_{i,j}(\tau)) \\ &\geq \frac{\tau_1 - \tau}{h}\left(-\sum_{i,j}\gamma(p^0_{i,j})\right) + \frac{\tau - \tau_0}{h}\left(-\sum_{i,j}\gamma(p^1_{i,j})\right) \\ &= \frac{\tau_1 - \tau}{h} H(F,G)(\tau_0) + \frac{\tau - \tau_0}{h} H(F,G)(\tau_1).\end{aligned}$$

Thus the joint entropy is a concave function on the interval (τ_0, τ_1). As a result, the mutual information function is piecewise convex and it has kinks when the pixel grids of the two images coincide.

VI. Numerical Results

Numerical experiments are performed to show the smoothness of mutual information using different models and with different parameters. We also compare the robustness and accuracy of our continuous model with other standard and artifact-reducing models in 2D and 3D. The implementation of nonparametric windows (NP) [23], third order generalized partial volume (GPVE) [7], and Parzen window (PZW) [19] are given by the publicly available code.[1]

The images for the 2D experiments are CT and MRI brain images from R.I.R.E. project.[2] The images for the 3D experiments are MRI images obtained from Brainweb[3] as well as from the R.I.R.E project. The images are initially aligned, thus the "ground truth" is zero unless it is otherwise specified. The voxel size for R.I.R.E. images is $1.25 \times 1.25 \times 4$ mm and BrainWeb images is $1.00 \times 1.00 \times 1.00$ mm. The pixel resolutions are 256×256 for 2D and 3D experiments with 64 slices for BrainWeb images and 20 to 26 slices for R.I.R.E images.

[1]Available at URL http://www.ict.csiro.au/staff/nicholas.dowson/

[2]Available at URL http://insight-journal.org/rire/

[3]Available at URL http://www.bic.mni.mcgill.ca/brainweb/

A. Smoothness

In this section, we illustrate the smoothness of the mutual information computed by the proposed continuous method. The continuous method differs from the interpolation-based model in that we use the intensity values at all subpixel locations when computing PDFs whereas the interpolation-based model only uses the intensity values at pixel locations. One may improve the interpolation-based model by using more intensity values. For instance, sample multiple points between two pixel locations. When a large number of sampling points are used, one would expect to obtain smoother mutual information [15]. Fig. 7(a) shows plots of mutual information for the T1-CT image pair when 0 to 4 samples points are used as well as by our method (CM). Mutual information becomes smoother as the number of sample points per pixel increases. However, we should point out that as the number of sample points goes to infinity, the computed mutual information will actually converge to the one given by [23], not the mutual information by our method; see Fig. 3.

A drawback of this multiple sampling approach is high computational cost. As shown by the magnified plot in Fig. 7(b), even with 4 sample points, the mutual information function is still not as smooth as our method. If we increase the inserted points, we also increase the complexity. In general, the complexity increases quadratically in 2D and cubically in 3D with the number of sample points.

Fig. 7(c) shows the mutual information functions computed by the partial volume model using multiple sample points. As we insert more points between pixels, it results in more local maxima and minima. Although the amplitudes tend to become smaller, the many kinks are still there.

We next illustrate the effect of number of stripes on mutual information. Fig. 8 shows the mutual information as a function of horizontal translation (vertical translation is similar). The number of stripes varies from 1 to 4 and we test it on 2 sets of images. To the limit, when an infinite number of stripes are used, it will converge to the analytical PDF for 2D images. As we can see, the mutual information functions are almost identical using 1, 2 or 4 stripes, showing that approximating the analytical PDF using only 1 stripe is already very accurate. In all of our experiments, 1 stripe is used.

We then consider the effect of number of bins. For the T1-CT image pair, Fig. 9 shows the mutual information of translation computed by the interpolation-based (IM), partial volume (PV) and our methods (CM). In the top row, the number of bins varies from 16 to 128. We also test the results using different image resolutions as shown in the bottom row. Note that no bin is used in our method since we compute PDFs rather than discrete probabilities. We instead consider the effect of the number of cells in the computation of entropies (cf. Section III-B). Both the interpolation and partial volume models introduce more and more interpolation artifacts when the number of bins increases or the image resolution decreases.

TABLE I

CONVERGENCE OF POWELL'S AND NELDER-MEAD METHODS FOR IM, PV, AND CM IN 2D IMAGE REGISTRATION. INITIAL GUESS RANGES ARE 0–10 mm FOR TEST 1, 10–20 mm FOR TEST 2, AND 20–30 mm FOR TEST 3

Test 1	Powell's Method		Nelder-Mead Method	
Model	256×256	128×128	256×256	128×128
IM	95%	90%	80%	70%
PV	100%	100%	90%	60%
CM	100%	100%	100%	100%

Test 2	Powell's Method		Nelder-Mead Method	
Model	256×256	128×128	256×256	128×128
IM	95%	90%	85%	35%
PV	100%	95%	75%	30%
CM	100%	100%	100%	100%

Test 3	Powell's Method		Nelder-Mead Method	
Model	256×256	128×128	256×256	128×128
IM	90%	80%	50%	15%
PV	95%	85%	70%	35%
CM	100%	100%	100%	100%

TABLE II

REGISTRATION RESULTS (PIXELS) OF NELDER-MEAD AND TRUST REGION METHODS FOR DIFFERENT MODELS WITH INITIAL GUESSES (10, 12) FOR TEST 1, (8.3, −10.2) FOR TEST 2, AND (−7.3, −5.8) FOR TEST 3

Test 1	Nelder-Mead		Trust Region	
Model	Iter	Computed Solution	Iter	Computed Solution
IM	44	(10.38, 0.26)	10	(9.63, 11.36)
PV	32	(10, 12)	3	(10, 12)
CM	50	(0.19, −0.019)	17	(0.19, −0.022)

Test 2	Nelder-Mead		Trust Region	
Model	Iter	Computed Solution	Iter	Computed Solution
IM	49	(0.26, 0.22)	3	(8.30, −10.20)
PV	48	(0.00, 0.00)	7	(7.99, −9.98)
CM	49	(0.19, −0.019)	18	(0.19, −0.021)

Test 3	Nelder-Mead		Trust Region	
Model	Iter	Computed Solution	Iter	Computed Solution
IM	32	(0.67, −6.57)	5	(−7.30, −5.76)
PV	35	(−7.00, −6.00)	5	(−6.99, −5.99)
CM	56	(0.19, −0.019)	18	(0.19, −0.022)

On the other hand, the continuous method gives smooth mutual information, regardless of the number of cells and image resolutions.

Finally, Fig. 10(a) shows a trace plot of the mutual information function of the translations in both horizontal and vertical directions using our continuous model. It is the function being maximized by the optimization scheme described in Section IV-A. Similar to the 1D plots, there are no interpolation artifacts. The mutual information is smooth and shows only one (global) maximum. Similar plots generated by nonparametric windows [23], GPVE [7], and Parzen window [19] are also given in Fig. 10 for comparison.

B. 2D Image Registration

To illustrate the benefit of the smoothness of the mutual information using the continuous method, two commonly used optimization methods, the Powell's method [31] and the Nelder-Mead method [29], are considered. For the interpolation and partial volume models, we just apply the standard Nelder-Mead or Powell's methods. For the continuous method, we use the optimization scheme in Section IV-A. For the optimization in each principal axis and rotation, we apply Powell's method (i.e. Brent's method in 1D) or the Nelder-Mead method. The images in this section are from R.I.R.E.

In our test data, the original target and template images are aligned; i.e. the optimal translations are $(0, 0)$, and the optimal rotation is 0°. We test the convergence of different models using different initial guesses. We randomly generate 20 initial guesses where the x and y coordinates are in the ranges of 0 to 10 mm, 10 to 20 mm, and 20 to 30 mm, respectively. The means of these initial guesses are $(4.68, 5.31)$ mm, $(15.08, 14.91)$ mm and $(24.83, 25.12)$ mm, respectively. Meanwhile we also impose a random initial rotation between $-6°$ and 6°. The number of bins for the interpolation and partial volume models is 16. In practice, rigid registration often needs to recover around 20 mm misalignment. Thus the performance of the optimization methods in the second and third ranges are of particular importance. A test is deemed a failure if the distance between the optimal translation and the calculated translation exceeds 2 mm. The convergence results of the Powell's and Nelder-Mead methods are shown in Table I. The second to fifth columns show the rate of successfully recovered alignments out of 20 experiments with different image resolutions, obtained by the interpolation model (IM), partial volume model (PV), and the continuous method (CM). We note that the success rates for IM and PV are consistent with the results reported in the literature, e.g. [23]. For the continuous method, since the computed mutual information function is smooth and has no artifacts, the optimization methods always converge to the optimal solutions.

Table II shows some sample convergence results of different registration models. In this experiment, the ground truth is at pixel location $(0.2, -0.02)$. The iteration stops when the difference between successive iterates is less than 0.01 pixel. Columns 2 and 4 show the iteration numbers using the Nelder-Mead method and a trust region method [32], respectively. For the CM model, it is the total number of iterations from all optimizations in x, y, and rotation. The trust region method is gradient based method which is used to show the effect of the use of gradient on the convergence of the registration methods. We remark that more computation is needed per iteration for the trust region than the Nelder-Mead method. Columns 3 and 5 shows the computed solutions. Due to interpolation artifacts, both the IM and PV models tend to converge to wrong solutions. The mutual information function is piecewise convex computed by the PV model with local maxima at integer pixel locations. Other than a few cases such as Test 2 (using the Nelder-Mead method), the PV model tends to converge close to an integer pixel location near the initial guess, and in Test 1, it actually stays as the initial guess. The CM model converges to the ground truth successfully in all cases. In general, Nelder-Mead takes more iterations than the trust region method. However, the trust region method

TABLE III
MEAN REGISTRATION RESULTS (PIXELS) OF DIFFERENT MODELS FOR IMAGE PAIRS CT-T1, T2-PD, AND T1-T2. NUMBERS IN SQUARE BRACKETS ARE STANDARD DEVIATIONS

Model	CT-T1	T2-PD	T1-T2
NP	(0.38, 0.39)	(0.39, 0.25)	(−0.55, −0.31)
	[(0.040, 0.043)]	[(0.056, 0.11)]	[(0.19, 0.11)]
GPVE	(0.21, 0.38)	(0.19, 0.067)	(−0.73, −0.21)
	[(0.056, 0.045)]	[(0.051, 0.032)]	[(0.064, 0.038)]
PZW	(0.35, 0.35)	(0.32, 0.20)	(−0.48, −0.30)
	[(0.024, 0.020)]	[(0.12, 0.16)]	[(0.11, 0.031)]
CM	(0.42, 0.047)	(0.16, −0.033)	(−0.39, −0.15)
	[(0.018, 0.015)]	[(0.0027, 0.0032)]	[(0.0051, 0.0030)]

usually converges quickly to the local maxima near the initial guess, probably due to the use of local derivative information. Nelder-Mead, on the other hand, does not use any derivative information. It has more capability to skip the artifacts and computes a solution closer to the global maximum.

We compare our continuous model (CM) with three other models: nonparametric windows (NP) [23], third order generalized partial volume (GPVE) [7], and Parzen window (PZW) [19], which are different approaches for reducing interpolation artifacts. The tested images are from the R.I.R.E. training data (see Fig. 1), slice 11, aligned by the gold standard provided. Since the images are registered in 3D, it is assumed that the 2D slices are also registered with ground truth close to (0, 0). Similar to Table I, we randomly generate 60 initial guesses with 20 from each of the ranges: 0 to 8 pixels, 8 to 16 pixels, and 16 to 24 pixels. For simplicity, we do not include rotation in this case. Nelder-Mead is used for the NP, GPVE, and PZW models. We use the optimization scheme in Section IV-A with each 1D problem solved by Nelder-Mead for our model. The registration results are shown in Table III for the image pairs CT-T1, T2-PD, and T1-T2. Columns 2-4 show the mean computed solutions in x and y, and the standard deviations. All models converge to a solution close to (0, 0) with the solutions given by the CM model are slightly closer to the origin. The standard deviations are typically small, showing the models consistently converge to a similar solution, regardless of locations of the initial guesses.

C. 3D Image Registration

We apply the continuous mutual information registration method to 3D image pairs (from Brainweb). Similar to 2D, we also randomly generate 20 initial guesses in the ranges of 0 to 10 mm, 10 to 20 mm, and 20 to 30 mm for image pairs with different image resolutions. The means of these initial guesses are $(4.2, 4.7, 4.8)$ mm, $(15.0, 14.6, 15.6)$ mm, and $(25.6, 26.4, 24.2)$ mm, respectively. We also impose random initial rotations about x-, y- and z-axis between -6° and 6°. The number of bins for interpolation and partial volume models is 32. A test is deemed a failure if the distance between the optimal translation and the calculated translation exceeds 2 mm. The convergence results of the Powell's method and Nelder-Mead method are shown in Table IV. Similar to 2D, the success rates of the interpolation and partial volume models deteriorate as the initial guesses get farther away from the ground truth due to the interpolation artifacts of the mutual information function. The continuous method is much more robust and it is able to converge for all cases.

TABLE IV
CONVERGENCE OF POWELL'S AND NELDER-MEAD METHODS FOR DIFFERENT MODELS IN 3D IMAGE REGISTRATION. INITIAL GUESS RANGES ARE 0–10 mm FOR TEST 1, 10–20 mm FOR TEST 2, AND 20–30 mm FOR TEST 3

Test 1	Powell's Method		Nelder-Mead Method	
Model	256×256×64	128×128×32	256×256×64	128×128×32
IM	100%	100%	85%	80%
PV	100%	100%	80%	65%
CM	100%	100%	100%	100%

Test 2	Powell's Method		Nelder-Mead Method	
Model	256×256×64	128×128×32	256×256×64	128×128×32
IM	100%	95%	75%	65%
PV	100%	100%	60%	55%
CM	100%	100%	100%	100%

Test 3	Powell's Method		Nelder-Mead Method	
Model	256×256×64	128×128×32	256×256×64	128×128×32
IM	50%	65%	45%	40%
PV	55%	60%	55%	55%
CM	100%	100%	100%	100%

TABLE V
MEAN REGISTRATION ERRORS (MM) OF FIRST, SECOND, THIRD ORDERS GENERALIZED PV MODELS AND OUR CONTINUOUS MODEL FOR CT-PD, CT-T1, AND CT-T2 IMAGE DATASETS

CT-PD	1st GPVE	2nd GPVE	3rd GPVE	CM
Patient001	1.9977	1.9228	1.9153	1.7066
Patient002	1.2299	1.2432	1.2937	0.8784
Patient003	2.0026	1.8938	1.7069	1.6509
Patient004	3.6220	2.9257	2.8353	3.2663
Patient005	1.8771	1.9415	2.0097	1.5921
CT-T1	1st GPVE	2nd GPVE	3rd GPVE	CM
Patient001	1.7707	1.2560	1.2345	1.0885
Patient002	1.2745	0.8215	0.8475	0.6259
Patient003	1.2328	1.3193	1.3035	0.7297
Patient004	2.9410	1.7740	1.8078	2.3266
Patient005	1.3967	1.2782	1.3057	0.9580
CT-T2	1st GPVE	2nd GPVE	3rd GPVE	CM
Patient001	2.1906	2.7672	2.8046	2.4859
Patient002	1.5538	1.7799	1.8891	1.1927
Patient003	1.3920	1.2336	1.2265	2.4316
Patient004	2.7021	2.6294	2.7959	2.8869
Patient005	1.7265	1.9959	1.9801	2.0896

We now show an example to compare the continuous method (CM) with the generalized partial volume method (GPVE) [7] on a set of testing datasets from R.I.R.E. The ground truth transforms were obtained from a prospective marker-based technique and they do not have perfect grid alignment in general. Registration results can be compared with the ground truth through an online interface which returns the errors of the registrations [28].

The database includes 3D CT, T1, T2, and PD weighted MR image volumes for a number of patients. Here, we perform image registrations for 3 pairs of images: CT-PD, CT-T1, and CT-T2. Table V shows the mean registration errors of our continuous model (last column) obtained from the online database interface for the first five patients. The results for the other patients are similar. The continuous model successfully

finds the correct transformation parameters for all the testing cases. As a comparison, we also include the registration results (columns 2 to 4) reported in [7] where they applied the GPVE model to alleviate the interpolation artifacts on the same datasets. Their main idea is to generalize the partial volume model (first order) by using higher order splines (second and third orders). Table V shows that the continuous model produces better results than the GPVE model in 10 cases and worse results in the other 5 cases. We remark that the purpose of this example is to illustrate how the GPVE and our methods perform. There are likely other methods which are able to obtain more accurate registration results on these datasets.

VII. Conclusion

We have proposed a method for computing mutual information to address the issue of interpolation artifacts. Images are modeled as functions defined on a continuous image domain. We have described a construction for the probability and joint probability density functions of the image functions. We have also presented a numerical scheme to compute the entropies and joint entropies of the images. We have generalized the idea from 1D to 2D and 3D by optimizing each translation parameter separately. The mutual information function computed seems to be smooth. Numerical results show that the continuous model always yields convergence in solving the registration problem. We have demonstrated the robustness and accuracy of the continuous model by the 2D and 3D images obtained from the public domain databases. In addition, an analysis is presented to explain the special piecewise convex shape of the mutual information obtained from the partial volume model.

Acknowledgement

The authors would like to thank the referees for their valuable comments and suggestions, especially the suggested clean formula for the joint PDF in one dimension.

References

[1] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[2] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," in *Proc. 5th Int. Conf. Comput. Vis.*, Jun. 1995, pp. 16–23.

[3] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multimodality medical image registration using information theory," in *Information Processing in Medical Imaging*, Y. Bizais, C. Barillot, and R. D. Paola, Eds. Boston, MA, USA: Kluwer, 1995, pp. 263–274.

[4] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, Sep. 1997.

[5] 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, no. 2, pp. 187–198, Apr. 1997.

[6] H. M. Chen and P. K. Varshney, "Registration of multimodal brain images: Some experimental results," *Proc. SPIE*, vol. 4731, pp. 122–133, Apr. 2002.

[7] H. M. Chen and P. K. Varshney, "Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1111–1119, Sep. 2003.

[8] H.-M. Chen, P. K. Varshney, and M. K. Arora, "Performance of mutual information similarity measure for registration of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2445–2454, Nov. 2003.

[9] G. Egnal and K. Daniilidis, "Image registration using mutual information," Dept. Comput. & Inf. Sci., Univ. Pennsylvania, Philadelphia, PA, USA, Tech. Rep. MS-CIS-00-05, 2000.

[10] J. X. Ji, H. Pan, and Z.-P. Liang, "Further analysis of interpolation effects in mutual information-based image registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1131–1140, Sep. 2003.

[11] J. Liu and J. Liu, "Artifacts reduction in mutual information-based image registration using prior information," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1. Sep. 2003, pp. 1113–1116.

[12] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual information matching and interpolation artefacts," *Proc. SPIE*, vol. 3661, pp. 56–65, May 1999.

[13] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Interpolation artefacts in mutual information-based image registration," *Comput. Vis. Image Understand.*, vol. 77, no. 2, pp. 211–232, Feb. 2000.

[14] G. K. Rohde, A. Aldroubi, and J. D. M. Healy, "Interpolation artifacts in sub-pixel image registration," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 333–345, Feb. 2009.

[15] M. Seppä, "Continuous sampling in mutual-information registration," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 823–826, May 2008.

[16] J. Tsao, "Interpolation artifacts in multimodality image registration based on maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 22, no. 7, pp. 854–864, Jul. 2003.

[17] H. Zhang, X. Zhou, J. Sun, and J. Zhang, "A novel medical image registration method based on mutual information and genetic algorithm," in *Proc. Int. Conf. Comput. Graphics, Imag. Vis., New Trends*, Jul. 2005, pp. 221–226.

[18] M. Wei, J. Liu, and J. Liu, "Artifact reduction in mutual-information-based CT-MR image registration," *Proc. SPIE*, vol. 5370, pp. 1176–1186, Feb. 2004.

[19] P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Process.*, vol. 9, no. 12, pp. 2083–2099, Dec. 2000.

[20] P. Thévenaz, M. Bierlaire, and M. Unser, "Halton sampling for image registration based on mutual information," *Sampling Theory Signal Image Process.*, vol. 7, no. 2, pp. 141–171, May 2008.

[21] O. Salvado and D. Wilson, "Removal of local and biased global maxima in intensity-based registration," *Med. Image Anal.*, vol. 11, no. 2, pp. 183–196, Apr. 2007.

[22] P. Thévenaz, T. Blu, and M. Unser, "Short basis functions for constant-variance interpolation," *Proc. SPIE*, vol. 6914, pp. 69142L-1–69142L-8, Feb. 2008.

[23] N. Dowson, T. Kadir, and R. Bowden, "Estimating the joint statistics of images using nonparametric windows with application to registration using mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1841–1857, Oct. 2008.

[24] N. Joshi, T. Kadir, and M. Brady, "Simplified computation for non-parametric windows method of probability density function estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1673–1680, Aug. 2011.

[25] A. Rajwade, A. Banerjee, and A. Rangarajan, "Probability density estimation using isocontours and isosurfaces: Application to information-theoretic image registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 475–491, Mar. 2009.

[26] P. Aljabar, J. Hajnal, R. Boyes, and D. Rueckert, "Interpolation artefacts in non-rigid registration," in *Proc. 8th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 3750. 2005, pp. 247–254.

[27] B. Karacali, "Information theoretic deformable registration using local image information," *Int. J. Comput. Vis.*, vol. 72, no. 3, pp. 219–237, May 2007.

[28] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, Jr., R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, E. Van den, A. Petra, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J. Hawkes, C. Studholme, J. B. Maintz, V. Antoine, M. A. Viergever, G. Malandain, X. Pennec, M. E. Noz, G. Q. Maguire, Jr., M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *J. Comput. Assist. Tomography*, vol. 21, no. 4, pp. 554–566, Jul.–Aug. 1997.

[29] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.

[30] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, Jun. 2001.

[31] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, MA, USA: Cambridge Univ. Press, 1988.

[32] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *Soc. Ind. Appl. Math. J. Optim.*, vol. 6, no. 2, pp. 418–445, May 1996.

Lin Xu was born in Tianjin, China, in 1984. He received the Bachelor of Science degree in information and numerical science from the School of Mathematical Sciences, Nankai University, Tianjin, China, in 2006, and the Master of Mathematics degree in computer science from the David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada, in 2008, under the supervision of Prof. J. W. L. Wan.

He joined the University Health Network, Toronto, ON, Canada, in 2008, as a Research Assistant. Since 2009, he has been with Oracle America, Inc., Redwood City, CA, USA, where he is currently a Software Engineer. His current research interests include numerical solutions and parallel computation on high-performance platforms, with applications in intensity-based image registration.

Justin W. L. Wan received the M.S. and Ph.D. degrees in applied mathematics from the University of California, Los Angeles, CA, USA, in 1998.

He was with the Scientific Computing and Computational Mathematics (SCCM) Program, Stanford University, Stanford, CA, USA, as a Forsythe Fellow from 1998 to 2001, before he joined the University of Waterloo, Waterloo, ON, Canada, as an Assistant Professor. He is currently an Associate Professor with the David Cheriton School of Computer Science, University of Waterloo. His current research interests include scientific computing with special interests in developing efficient numerical techniques for solving partial differential equations arising from modeling, image segmentation, registration, tracking of cell images, scientific visualization, and physically-based simulation of natural phenomena in computer graphics.

Dr. Wan is currently is a Canada Research Chair in scientific computing as well as the Director of the Centre for Computational Mathematics in Industry and Commerce.

Tiantian Bian received the B.Eng. degree in mechanical engineering from Zhejiang University, Hangzhou, China, the M.Eng. degree in software engineering from Tsinghua University, Beijing, China, and the M.Math. degree in computer science from the University of Waterloo, Waterloo, ON, Canada, in 2003, 2007, and 2010, respectively.

He has been with Microsoft, Redmond, WA, USA, as a Software Development Engineer, since 2011. His current research interests include computer vision and high performance computing.