An Adaptive Monte Carlo Approach to Phase-Based Multimodal Image Registration

Alexander Wong, *Member, IEEE*

***Abstract*—In this paper, a novel multiresolution algorithm for registering multimodal images, using an adaptive Monte Carlo scheme is presented. At each iteration, random solution candidates are generated from a multidimensional solution space of possible geometric transformations, using an adaptive sampling approach. The generated solution candidates are evaluated based on the Pearson type-VII error between the phase moments of the images to determine the solution candidate with the lowest error residual. The multidimensional sampling distribution is refined with each iteration to produce increasingly more plausible solution candidates for the optimal alignment between the images. The proposed algorithm is efficient, robust to local optima, and does not require manual initialization or prior information about the images. Experimental results based on various real-world medical images show that the proposed method is capable of achieving higher registration accuracy than existing multimodal registration algorithms for situations, where little to no overlapping regions exist.**

***Index Terms*—Adaptive Monte Carlo, image registration, multimodal, Pearson error, phase.**

I. Introduction

IMAGE registration is the process of finding the optimal geometric transformation that aligns images of the same scene acquired under different conditions (e.g., time, view angle, sensor modality, etc.). Image registration plays an important role in various biomedical applications, such as medical image super-resolution [1], [2], medical image fusion [3], disease diagnosis [4], and computer-assisted surgery [5]. Of particular interest is multimodal image registration, where the underlying goal is to find the optimal geometric transformation that aligns images acquired, using different imaging modalities [e.g., magnetic resonance imaging (MRI)/computed tomography (CT)/positron emission tomography (PET)]. Traditionally, multimodal image registration required the manual selection of control point pairs between the images being registered, which are used to estimate the transformation that aligns the images. This manual selection of control point pairs is very time-consuming and laborious. Therefore, methods that register images acquired, using different modalities in an automated manner, is desired.

Multimodal image registration is a very challenging problem for many reasons. The same scene acquired by different imaging modalities are represented by different intensity values. This makes it very difficult to align images based on their intensity values. This disparity in intensity mappings is further complicated by the presence of local image nonuniformities (e.g., static field and RF nonhomogeneities for MRI [6], [7]) and noise. Furthermore, such disparities can result in local minima along the convergence plane if evaluated in a direct manner, thereby affecting the ability of iterative optimization techniques, such as conjugate gradient [8] and Nelder–Mead simplex [9] to converge to the global optima. Finally, solving the multimodal registration problem can be very computationally expensive, particularly for large images. Therefore, multimodal image registration methods that can address all of these issues are highly desired.

The main contribution of this paper is a novel approach to the problem of registering images from different modalities, using an adaptive Monte Carlo scheme. The proposed method utilizes an adaptive sampling scheme to draw increasingly more plausible solution candidates from a multidimensional solution space. Solution candidate evaluation is performed based on the Pearson type-VII error between the phase moments of the images to determine the alignment between the images. To the best of the author's knowledge, there are currently no methods that utilize the concept of adaptive Monte Carlo method for the purpose of multimodal image registration. The key motivation of using an adaptive Monte Carlo method for the purpose of multimodal image registration is that it allows for efficient optimization while avoiding convergence issues under situations characterized by many local optima along the convergence plane and small cost gradients toward the global optima.

The paper is organized as follows. Previous work in multimodal registration is discussed in Section II. The theory underlying the proposed method is described in detail in Section III. The proposed method is presented in Section IV. The testing methods and experimental results are discussed in Section V. Finally, conclusions are drawn and future work is discussed in Section VI.

II. Previous Work

A large number of algorithms have been proposed for the purpose of registration images acquired, using different imaging modalities. Amongst the most popular multimodal image registration techniques are mutual information and entropy-based methods [10]–[15]. The underlying goal of entropy-based methods is to minimize the joint intensity entropy between the images being registered. These methods take advantage of the fact that correctly registered images correspond to tightly packed joint distributions and the transformation with minimum joint intensity entropy should theoretically be the optimal alignment. The main advantage of entropy-based methods is that it allows images acquired, using different imaging modalities to be

Manuscript received May 19, 2008; revised December 19, 2008 and April 5, 2009. First published November 3, 2009; current version published January 15, 2010.

The author is with the Systems Design Engineering, University of Waterloo, Waterloo N2L 3G1, Canada (e-mail: a28wong@engmail.uwaterloo.ca).

Digital Object Identifier 10.1109/TITB.2009.2035693

compared in a direct manner. Currently, entropy-based methods have been shown to be very effective in multimodal registration, and are considered to be state-of-the-art.

There are several important drawbacks to entropy-based methods. First, entropy-based methods are typically underconstrained with respect to intensity relationships. As such, the convergence planes associated with entropy-based methods possess high nonmonotonicity with many local optima [16]. This is problematic since most entropy-based methods utilize iterative optimization methods to solve for the optimal alignment between images, which rely on the monotonicity of the convergence plane. Furthermore, entropy-based methods require the calculation of marginal and joint entropies, which is computationally expensive to perform.

Another popular group of multimodal image registration techniques are feature-based methods [17]–[22]. In feature-based methods, the images are transformed into a common feature space prior to cost evaluation. Therefore, such methods attempt to find image correspondences in an indirect manner by finding correspondences between extracted features that exist in a common feature space. Features used in such methods include intensity gradient information [17], [18], local frequency information [19]–[21], and shape properties [23]. There are several important advantages to the use of feature-based methods. First, since images are converted to a common feature space prior to comparison, objective functions that are more constrained than those used in entropy-based methods with respect to interimage feature relationships can be used. As a result, the convergence planes associated with feature-based methods typically possess higher monotonicity with fewer local optima. Second, feature-based methods allow more computationally efficient objective functions, such as sum of squared distances and cross correlation to be utilized. In particular, efficient techniques for evaluating such objective functions exhaustively for all possible translations and rotations on a pixel level have been proposed [18].

There are several important drawbacks to feature-based methods that need to be addressed. First, while methods exist for performing objective function evaluation exhaustively on a pixel level for simple transformations, this type of exhaustive evaluation becomes intractable to perform on a subpixel level and/or more complex transformations due to high computational costs. Second, while the convergence plane for feature-based methods are generally more monotonic than entropy-based methods, whether the global optima corresponds to the optimal alignment depends heavily on the selection of appropriate features as well as objective functions.

Other recent techniques include correlation ratio [24] and regression-based methods [25], [26]. Such methods make the assumption that the relationship between intensity values from the images being registered can be represented as a function (e.g., polynomials [24], [26] and piecewise linear [25]). One issue with such methods is that the functional assumption is often not true and are not easily customizable to handle scenarios with different intensity relationships [27].

In addition to the specific problems associated with each group of multimodal image registration methods, all of the aforementioned methods encounter difficulties when faced with situations characterized by: 1) large misregistrations; and 2) little to no initial region overlap between the images. Methods that utilize iterative optimization methods, such as gradient descent, conjugate gradient [8], Nelder–Mead simplex [9], Levenberg–Marquardt method, Powell's method, and quadratic programming are often unable to converge under such situations due to local optima along the convergence plane and small cost gradients toward the global optima (i.e., moving toward the global optima yields little to no decrease in cost), even when multiresolution methods are used. Multistart methods attempt to reduce the effects of local minima on the convergence to the global optima by initializing local optimizations at multiple starting points. However, this can become computationally expensive for large search spaces where many local optimizations must be performed at different starting points, and choosing such starting points can also be a challenging task. Methods that utilize exhaustive search over all possible transformations are able to avoid the issue of local optima along the convergence plane, but at the cost of high-computational complexity that is only tractable for simple transformations and pixel-level accuracy. The goal of the proposed study is to introduce a feature-based image registration method that addresses the convergence and computational complexity issues associated with large misregistrations as well as situations where there is little to no initial region overlap between the images through the use of an adaptive Monte Carlo scheme.

It is important to distinguish the proposed adaptive Monte Carlo scheme from other image registration schemes that utilize stochastic sampling concepts, particularly the multistart registration method proposed by Song *et al.* [15] and the particle filter-based approach proposed by Florin *et al.* [22]. The underlying goal of the registration method proposed by Song *et al.* is to identify several good starting points from which to initialize a set of local optimizations. The selection of desirable starting points that would lead to a correct solution during the local optimization processes was achieved by Song *et al.* through a prior learning strategy, which constructs nonparametric prior models for each transform parameter, using stochastic sampling. As such, the use of stochastic sampling in the method proposed by Song *et al.*, is in the prior learning process and not in the iterative optimization process. This use of stochastic sampling is fundamentally very different from the proposed method, which utilizes stochastic sampling in the iterative optimization process and not in the learning process. As such, the proposed method can be initialized at an arbitrary starting point without affecting convergence to a good solution. This divergence of ideas lead to very different optimization approaches as well as very different learning models. Since the fundamental improvement of the method proposed by Song *et al.* over existing multistart methods, is in selecting good starting points based on learnt priors, a nonparametric model is crucial for the success of such an algorithm. However, since the proposed method is a single-start method that attempts to converge to a single good solution through a sequence of stochastically sampled solution candidates, an adaptive parametric model is more suitable for guiding the stochastic sampling process toward the single good solution. Similarly, the method proposed by Florin *et al.* [22]

utilizes stochastic sampling in the learning stage, where the Bayesian posterior probability density function is estimated at each iteration through stochastic sampling, and then, used to compute a cost function gradient. This is fundamentally different from the proposed method, which does not rely on cost function gradients in finding the solution. Furthermore, the repeated distribution reestimation process required for the method proposed by Florin *et al.* is very computationally expensive.

III. Theory

Prior to explaining the proposed method, it is important to first explain the theory behind the proposed method. First, the proposed adaptive Monte Carlo scheme is described and explained in detail to justify the use of such a method for multimodal image registration. Finally, the proposed candidate solution evaluation scheme is presented.

A. Adaptive Monte Carlo Method

Suppose we wish to register two different images f and g acquired, using different imaging modalities. The optimal transformation $\hat{T}$ that bring f and g into alignment can be formulated as an optimization problem

$$\hat{T} = \arg\min_{T} \left[C\left(f\left(T(\underline{x}) \right), g\left(\underline{x} \right) \right) \right] \tag{1}$$

where $\underline{x}$ represents a point in image space and C is the objective function that evaluates the dissimilarity between the images. Based on this formulation, the goal is to find a feasible solution from the solution space of possible geometric transformations that minimizes the objective function.

Many iterative optimization schemes have been proposed for finding the optimal solution [8], [9], [28], [29]. Such methods work based on the assumption that the convergence plane is monotonic in nature. However, this assumption of monotonicity is often not the case, particularly for situations characterized by high-dimensional solution spaces. Therefore, such methods often become trapped in local optima along the convergence plane. This issue is especially problematic in situations, where characterized by large misregistration and little region overlap between the images. In such cases, moving toward the global optima yields little to no decrease in cost, and hence, iterative methods would fail to find the global optima in such cases. A method to alleviate this problem is to evaluate all feasible solutions in the solution space. While methods exist to perform such exhaustive solution evaluation efficiently for low-dimensional solution spaces on a pixel level [18], it is intractable to evaluate $\hat{T}$ in such a manner for high-dimensional solution spaces or on a subpixel level from a computational perspective. For example, to exhaustively evaluate solutions from the solution space of all possible integer 2-D translations and rotations for two 256×256 images would require the evaluation of over 23 million solution candidates.

To address this important issue, we propose that we instead generate random solution candidates for $\hat{T}$ from the solution space of possible geometric transformations in an efficient manner, using a novel adaptive Monte Carlo scheme. Let us consider a m-D random field S representing the solution space of all possible geometric transformations as defined by m model parameters, T be a random variable in S, and p be an arbitrary probability density function on S. If we were to take n random solution candidates $T_1, \ldots, T_n$ based on p, the Monte Carlo estimate of $\hat{T}$ can be defined as follows:

$$\hat{T} = \arg\min_{T \in \{T_1, \ldots, T_n\}} \left[C\left(f\left(T(\underline{x}) \right), g\left(\underline{x} \right) \right) \right]. \tag{2}$$

There are several advantages to the use of Monte Carlo method for the purpose of multimodal image registration. First, it avoid the issues associated with local optima along the convergence plane faced by iterative optimization methods, since it does not rely on local cost gradients to guide it toward the global optima. Therefore, such a scheme would not require manual initialization since the initial alignment of the images does not affect its ability to determine the global optima. Second, the use of Monte Carlo methods allows optimization problems that are infeasible to evaluate exhaustively (e.g., problems involving high-dimensional solution spaces) to be solved in an efficient manner.

One major problem with this "naive" Monte Carlo (NMC) approach to image registration is that it generates too many solution candidates that are either infeasible or far from the desired global optima. This can lead to an unnecessary increase in computational overhead from too many irrelevant solution candidates being evaluated. The reason that the conventional Monte Carlo method produces so many irrelevant solution candidates is that it generates solution candidates from the solution space based on an arbitrary probability density function p. Therefore, the probability density function used to generate the solution candidates may differ significantly from that of solution candidates that are more plausible to be the global optima. A particularly effective technique for addressing the issue of solution candidate irrelevancy is adaptive sampling, where the underlying concept is that random variables with greater impact on the estimate (in this case, solution candidates $T_1, \ldots, T_n$ that are more plausible to be the global optima $\hat{T}$) should be sampled more frequently. A sampling density function p_* is used to emphasize important regions in the solution space in an attempt to significantly reduce irrelevant solution candidates and improve computational performance.

The fundamental issue associated with adaptive sampling is the selection of sampling density function p_*, which is critical to the computational performance of the Monte Carlo method. In the case of multimodal image registration, the probability density that describes whether a given solution candidate is close to the desired solution is generally unknown. As such, it is very difficult to select a good sampling density for the image registration problem, particularly in situations where the solution exists in a high-dimensional solution space. To address this issue, we propose an adaptive sampling scheme, where an initial sampling density is corrected and refined with each iteration to produce increasingly more plausible solution candidates for the optimal alignment.

The proposed adaptive sampling scheme can be described as follows. Let $T = (t_1, t_2, \ldots, t_n)$ be a solution candidate in the

n-dimensional solution space of possible transformations for 2-D image alignment, where t_i is the ith parameter of the transformation model, and t_1 and t_2 correspond to the translation along the x- and y-axis, respectively. First, an initial sampling density p_*^1 is used to generate an initial set of random solution candidates $T_1^1, \ldots, T_n^1$. The initial sampling density p_*^1 is defined as follows:

$$p_*^1(t_1) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(t_1 - m_1)^2}{2\sigma_1^2}\right),\ t_1^{\max} > t_1 > t_1^{\min} \tag{3}$$

$$p_*^1(t_2) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{(t_2 - m_2)^2}{2\sigma_2^2}\right),\ t_2^{\max} > t_2 > t_2^{\min} \tag{4}$$

$$p_*^1(t_i) = \frac{1}{t_i^{\max} - t_i^{\min}},\ t_i^{\max} > t_i > t_i^{\min} \text{ and } i > 2 \tag{5}$$

where σ_1 and σ_2 are the base standard deviations of p_* for t_1 and t_2, respectively, m_1 and m_2 represent the translation along the x- and y-axis, respectively that aligns the center of masses of f and g, and $t_i^{\min}$ and $t_i^{\max}$ represent the minimum and maximum allowable values for parameter t_i. Each solution candidate is then tested using the objective function C to determine the solution $\hat{T}^{k-1}$ that minimizes the objective function from the set of solution candidates. At each iteration k, the sampling density p_*^k is adaptively refined based on the cost gradient ΔC_k between iteration $k-1$ and $k-2$, and solution candidate $\hat{T}^{k-1}$ as follows:

$$p_*^k(t_i) = \frac{1}{\sigma_i\left(\Delta C_k/\Delta C_3\right)\sqrt{2\pi}} \exp\left(-\frac{(t_i - \hat{t}_i^{k-1})}{2(\sigma_i(\Delta C_k/\Delta C_3))^2}\right), \tag{6}$$

$$t_i^{\max} > t_i > t_i^{\min} \tag{7}$$

where σ_i is the base standard deviations of p_* for t_i and can be defined as follows:

$$\sigma_i = \frac{t_i^{\max} - t_i^{\min}}{4}, \tag{8}$$

$\hat{t}_i^{k-1}$ is the ith parameter of $\hat{T}^{k-1}$, and ΔC_k and ΔC_2 are the cost gradients at iterations k and 3 as defined as follows:

$$\Delta C_k = C(\hat{T}^{k-1}) - C(\hat{T}^{k-2}) \tag{9}$$

$$\Delta C_3 = C(\hat{T}^2) - C(\hat{T}^1). \tag{10}$$

It can be observed from (6) that at each iteration, the mean and variance of sampling density p_* are refined relative to the parameters of the optimal solution from the previous iteration $\hat{T}^{k-1}$ and the reduction in cost between the previous two iterations ΔC_k. This adaptive sampling density estimation is based on the theory that as the algorithm converges to the global optima, the most plausible solutions for the optimization problem should be found in regions in the solution space that is increasingly closer to the previous best solution. Based on this assumption, a good solution candidate distribution model would intuitively be the Gaussian distribution model, where the sampling density of solution candidates is concentrated near the previous best solution, and steadily declines as we move away from the previous best solution. The main advantage to the proposed adaptive sampling scheme is that no prior information is needed about the images beforehand to select an appropriate sampling density for generating plausible solution candidates from the solution space. Therefore, the proposed scheme encourages solution locality while still allowing for the possibility of solution candidates that are far from the previous solution.

B. Solution Candidate Evaluation

For each solution candidate T generated by the adaptive sampling scheme, it is necessary to test the solution candidate, using an objective function C to determine the associated cost. Therefore, registration accuracy depends heavily on the objective function being used. To allow for well-constrained similarity evaluation between images acquired under different modalities, we propose that each generated solution candidate is evaluated, using an objective function based on the Pearson type-VII [30] error between the phase moments of the images being registered. The proposed solution candidate evaluation process can be described in the following manner. Upon initialization, the phase moments ρ associated with each point in f and g are calculated based on the iterative estimation scheme we previously proposed in [31], which was shown to be highly robust to image nonuniformities and noise. Given an image f_0, the initial local phase coherence estimate P_0 at orientation θ is obtained at iteration $t = 0$, using the following expression [32]:

$$P(\underline{x}, \theta) = \frac{\sum_n W(\underline{x}, \theta)\, \lfloor A_n(\underline{x}, \theta)\, \Delta\Phi(\underline{x}, \theta) - T \rfloor}{\sum_n A_n(\underline{x}, \theta) + \varepsilon} \tag{11}$$

$$\Delta\Phi(\underline{x}, \theta) = \cos(\phi_n(\underline{x}, \theta) - \bar{\phi}(\underline{x}, \theta)) - |\sin(\phi_n(\underline{x}, \theta) - \bar{\phi}(\underline{x}, \theta))| \tag{12}$$

where W is the frequency-spread weighting factor (coherence across a wide frequency spread is weighted more than coherence across a narrow frequency spread), $\bar{\phi}$ is the weighted mean phase, T is a noise threshold, and ε is a small constant used to avoid division by zero. The parameters used during implementation were those described in [32].

At each iteration t, phase moments ρ_t are computed based on P_{t-1} as follows:

$$\begin{aligned}\rho_t(\underline{x}) = {} & \frac{1}{2}\sum_\theta P_{t-1}(\underline{x}, \theta)^2 \\ & + \frac{1}{2}\Bigg[4\Bigg(\sum_\theta (P_{t-1}(\underline{x}, \theta)\sin(\theta))(P_{t-1}(\underline{x}, \theta)\cos(\theta))\Bigg)^2 \\ & + \Bigg(\sum_\theta [(P_{t-1}(\underline{x}, \theta)\cos(\theta))^2 \\ & \quad - (P_{t-1}(\underline{x}, \theta)\sin(\theta))^2]\Bigg)^2\Bigg]^{1/2}.\end{aligned} \tag{13}$$

A new estimate of f_t is then computed based on a moment-adaptive bilateral estimation scheme

$$f_t(\underline{x}) = \frac{\sum_{\psi} w\big(\underline{x}, \psi, \rho_t\,(\underline{x})\big) f_{t-1}\,(\psi)}{\sum_{\psi} w\big(\underline{x}, \psi, \rho_t(\underline{x})\big)} \tag{14}$$

where ψ is a local neighborhood around $\underline{x}$, and the estimation weighting function w consists of a spatial weighting function w_s, and an amplitudinal weighting function w_a

$$w(\underline{x}, \psi, \varpi_t(\underline{x})) = w_a(\underline{x}, \psi, \varpi_t(\underline{x}))w_s(\underline{x}, \psi, \varpi_t(\underline{x})) \tag{15}$$

$$w_s(\underline{x}, \psi, \varpi_t(\underline{x})) = e^{-1/2(\|\underline{x}-\psi\|/(\sigma_s(\varpi_t(\underline{x}))))^2} \tag{16}$$

$$w_a(\underline{x}, \psi, \varpi_t(\underline{x})) = e^{-1/2(\|I(\underline{x})-I(\psi)\|/(\sigma_a(\varpi_t(\underline{x}))))^2}. \tag{17}$$

The estimated image f_t becomes the basis for estimating phase coherence P_{t+1} during the next iteration. The iterative estimation scheme is performed for f and g to determine the phase moments ρ_f and ρ_g. The parameters used during implementation were those described in [31].

To evaluate the cost associated with solution candidate T, the proposed objective function is defined as the cumulative Pearson type-VII [30] error between the phase moments ρ_f and ρ_g

$$C(f(T(\underline{x}), g(\underline{x})) = \sum_{\underline{x}} \ln(1 + (\rho_f(T(\underline{x})) - \rho_g(\underline{x}))^2)^{1/2}. \tag{18}$$

One of the main advantages of using the Pearson type-VII error metric is that it is highly robust to outliers. Common methods for evaluating error, such as Manhattan and quadratic error metrics are highly sensitive to outliers. To illustrate this, the influence of outliers on an error metric can be studied based on its derivative [33]. For example, the derivative of the quadratic error metric e^2 is $2e$. This indicates that the influence of outliers on quadratic error metric increases linearly and without bound. The derivative of the Pearson type-VII error metric $\ln((1 + e^2)^{1/2})$, on the other hand, is $e/(1 + e^2)$. Therefore, the influence of outliers on Pearson type-VII error is bounded.

IV. Proposed Method

Based on the above theory, the proposed method can be described as follows. Given images f and g, the corresponding phase moments ρ_f and ρ_g are computed as described in Section III-B. At each iteration k, random solution candidates $T_1^k, \ldots, T_n^k$ are generated from the solution space S based on the adaptive sampling density p_*^k proposed in Section III-A. The generated solution candidates $T_1^k, \ldots, T_n^k$ are evaluated based on the cumulative Pearson type-VII error between the phase moments of the images being registered as described in Section III-B to determine the best solution $\hat{T}^k$ from the set of generated solution candidates. This process is repeated until the termination criteria are satisfied to obtain the solution $\hat{T}$.

A multiresolution scheme involving three different scales $(s = (1/4), (1/2), 1)$ was used to improve convergence speed. For testing purposes, 100 solution candidates are generated and evaluated at each iteration of the proposed method, and for the case of affine transformations $(t_1^{\min}, t_1^{\max}, t_2^{\min}, t_2^{\max}, t_3^{\min}, t_3^{\max}) = (-(2w/3), (2w/3), -(2h/3), (2h/3), 0, 2\pi)$ represents translation along x- and y-axis, and rotation, respectively, and w and h are the width and height of the reference image, respectively, and $(t_4^{\min}, t_4^{\max}, t_5^{\min}, t_5^{\max}, t_6^{\min}, t_6^{\max}, t_7^{\min}, t_7^{\max}) = (0.8, 1.2, 0.8, 1.2, 0, 0.2, 0, 0.2)$ represents scale and shear along x- and y-axis, respectively.

V. Experimental Results

The proposed method was implemented in MATLAB and evaluated in two different experiments, using real medical image sets obtained from the National Library of Medicine visible human project (VHP) and whole brain atlas (WBA) [34]. A summary of each image set is given below.

1) BRAIN: Brain, axial, 1 mm resolution, T1-T2.
2) PELVIS: Pelvis, coronal, 1.875 mm resolution, T1-T2.
3) TORSO: Torso, coronal, 1.875 mm resolution, T1-T2.
4) BPET: Brain, sagittal, 1 mm resolution, T1-PET.
5) BCT: Brain, axial, 1 mm resolution, proton density (PD)-CT.

A. Experiment 1

The first set of experiments test, the registration accuracy of the proposed method, using all five of the image sets. Each image set was distorted, using 30 randomly generated affine transformations, resulting in a total of 150 test cases. Since the original image sets were aligned, the gold-standard transformations are known for all 150 tests. For comparison purposes, state-of-the-art methods, such as normalized mutual information (NMI), phase mutual information (PMI) [14], least-squares intensity remapping (LSIR) [26], as well as a NMC approach were also tested for each of the 150 test cases. To evaluate the registration accuracy of the methods under test, the rmse for 30 ground-truth control point pairs was computed. Both the NMI and PMI methods were implemented, using smoothed histograms with 64 intensity bins within a multiresolution framework involving three different scales $(s = (1/4), (1/2), 1)$. Furthermore, to ensure a fair comparison, a common sequential quadratic programming [29] local optimization scheme is used for all tested methods. Finally, a multiresolution scheme involving three different scales $(s = (1/4), (1/2), 1)$ was used to improve convergence speed for the tested methods.

The registration accuracy results for the first set of experiments are summarized in Table I. It can be observed that both the NMC approach and the proposed method achieved lower rmse than all of the other tested methods for all of the test cases. The high rmse results associated with the NMI and PMI methods, is largely due to high monotonicity along the convergence plane as a result of the underconstrained nature of such methods. This high monotonicity along the convergence plane causes iterative optimization methods to be trapped in local minima, and thus, unable to converge to the global minima. As such, the NMI and PMI methods only work well when the initial alignment is reasonably close to the global optima in the first place. However, the underlying goal of automatic image registration is to not rely on good initial alignments and to be able to converge to the global optimal given any initial conditions. The NMC approach and the proposed method avoid the issues associated

TABLE I
REGISTRATION ACCURACY FOR EXPERIMENT 1

Test Set	RMSE[1]				
	LSIR [26]	NMI	PMI [14]	NMC	Proposed Method
BRAIN	1.126/2.271/5.165	0.2958/27.47/94.29	0.2879/24.43/93.13	0.2753/0.5062/0.8453	0.2631/0.4814/0.8254
PELVIS	2.547/5.822/7.899	0.5315/32.98/106.3	0.1736/27.67/95.18	0.1576/0.2940/0.5895	0.1473/0.2981/0.5957
TORSO	2.784/4.255/5.439	0.2620/28.84/114.1	0.1965/23.17/106.9	0.1301/0.1929/0.3965	0.1285/0.1839/0.3718
BPET	11.596/13.40/15.694	0.9257/41.48/112.4	0.5682/36.19/112.4	0.4733/0.7507/1.5045	0.4684/0.7556/1.5615
BCT	2.246/3.056/4.899	0.6235/22.1830/99.78	0.3276/24.29/90.94	0.1915/0.4353/0.8799	0.1851/0.4261/0.8607

1: The RMSE is computed over 30 random distortions in pixels relative to the reference image. The table reports the min/mean/max of the 30 trials.

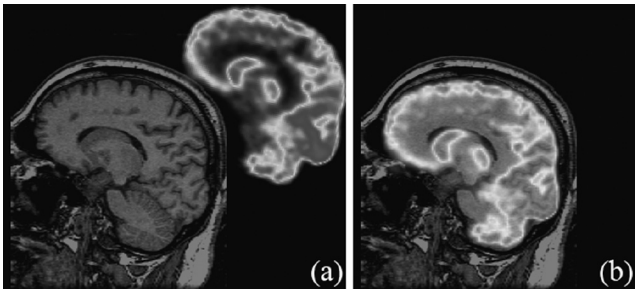


Fig. 1. BPET: (a) misaligned MRI T1 and PET images, (b) aligned using proposed method. It can be observed that successful registration was achieved despite the lack of region overlap.

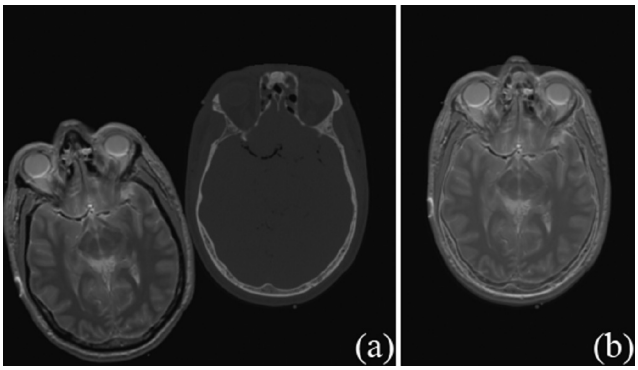


Fig. 2. BCT: (a) misaligned MRI PD and CT images, (b) aligned using proposed method. It can be observed that successful registration was achieved despite the lack of region overlap.

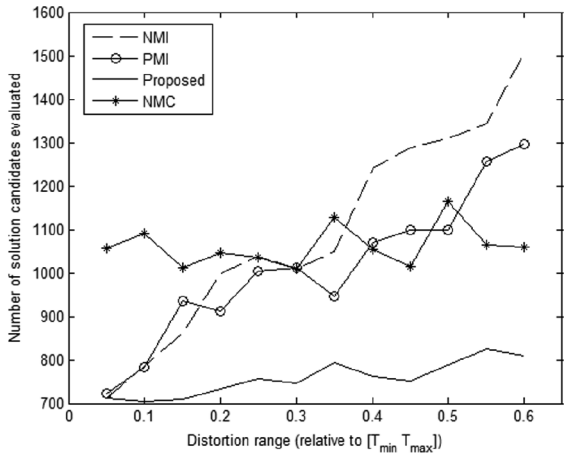


Fig. 3. Efficiency of the tested methods over various ranges of distortions. It can be observed that the number of solution candidates tested remains relatively constant over all ranges of distortions for the NMC method and the proposed method, with the NMC method evaluating a noticeably higher quantity of solution candidates. Furthermore, it can be observed that the average number of solution candidates evaluated for the tested NMI and PMI methods are noticeably higher than that evaluated by the proposed method for all ranges of distortion.

with local optima along the convergence plane faced by iterative optimization methods, since they do not rely on local cost gradients to guide it toward the global optima like the NMI and PMI methods. Therefore, these methods do not require good initialization since the initial alignment of the images does not affect its ability to determine the global optima.

Examples of T1-PET and CT-PD registration achieved using the proposed method for situations characterized by little to no region overlap between the images are shown in Figs. 1 and 2, respectively. It can be observed that in both cases, the proposed method is able to achieve good registration results. These results illustrate the effectiveness of the proposed method for registering images acquired, using different imaging modalities.

B. Experiment 2

The second set of experiments investigates the efficiency of the proposed method, using image sets TORSO, BPET, and BCT under situations characterized by large misregistrations and little to no region overlap between the images. Each of the image sets were distorted, using 30 randomly generated transformations over 12 ranges of x and y translation and rotation, ranging from 5% to 60% of the x and y image dimensions and rotation. This results in a total of 360 test cases. The proposed method was then performed on each test case and the average number of solution candidates tested by the proposed method to achieve convergence for each range of distortion, was measured. For comparison purposes, the NMI, PMI, and NMC methods were also evaluated. This set of experiments is designed to justify the claim that the proposed method maintains efficiency under situations characterized by large misregistrations and little to no region overlap.

A plot of the number of solution candidates evaluated with respect to increasing levels of misregistration is shown in Fig. 3. It can be observed that the quantity of solution candidates evaluated for both the NMC approach and the proposed method remains largely constant for all ranges. However, the quantity of solution candidates evaluated by the NMC approach is noticeably higher than that evaluated by the proposed method, thus validating the efficiency gained by using an adaptive Monte Carlo approach. Furthermore, it can be observed that the

average number of solution candidates evaluated for the tested NMI and PMI methods are noticeably higher than that evaluated by the proposed method for all ranges of distortion. Therefore, these results reinforce the claim that the proposed method is able to maintain efficiency under situations characterized by large misregistrations and little to no region overlap.

VI. Conclusion

In this paper, we introduced a novel method for registering images from different imaging modalities, using an adaptive Monte Carlo scheme. An adaptive sampling scheme was introduced for generating plausible solution candidates from the solution space of possible transformations. An objective function for solution candidate evaluation was introduced, based on the Pearson type-VII error between the phase moments of the images being registered. The proposed method is highly efficient and addresses issues associated with large misregistrations and little to no region overlap between images. Experimental results using real multimodal image sets indicate that high-registration accuracy can be achieved. Future study involves investigating alternate adaptive sampling density models to reduce computational complexity while improving solution candidate plausibility. Furthermore, we intend on extending the proposed method for true 3-D image registration.

References

[1] J. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom, and H. Azhari, "Super-resolution in PET imaging," *IEEE Trans. Med. Imag.*, vol. 25, no. 2, pp. 137–147, Feb. 2006.

[2] D. Robinson, S. Farsiu, J. Lo, P. Milanfar, and C. Toth, "Efficient multiframe registration of aliased X-ray images," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2007, pp. 1–5.

[3] J. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom, and H. Azhari, "Improved image fusion in PET/CT using hybrid image reconstruction and super-resolution," *Int. J. Biomed. Imag.*, vol. 2007, pp. 46846-1–46846-10, 2007.

[4] V. Walimbe, O. Dandekar, F. Mahmoud, and R. Shekhar, "Automated 3-D elastic registration for improving tumor localization in whole-body PET-CT from combined scanner," in *Proc. IEEE EMBC*, 2006, vol. 1, pp. 2799–2802.

[5] O. Sadowski, Z. Yaniv, and L. Joskowicz, "Comparative in-vitro study of contact and image-based rigid registration for computer-aided surgery," *Comput.-Aided Surg.*, vol. 7, no. 4, pp. 223–236, 2002.

[6] A. Simmons, P. Tofts, G. Barker, and S. Arridge, "Sources of intensity nonuniformity in spin echo images at 1.5 T," *Magn. Reson. Med.*, vol. 32, no. 1, pp. 121–128, 1994.

[7] M. Oghabian, S. Mehdipour, and N. Alam, "The impact of RF inhomogeneity on MR image nonuniformity," presented at the Proc. Image Vis. Comput., New Zealand, 2003.

[8] M. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Nat. Bureau Stand.*, vol. 49, pp. 409–436, 1952.

[9] J. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.

[10] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multimodality image registration based on information theory," in *Proc. IPMI*, 1995, pp. 263–274.

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

[12] C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3-D medical image alignment," *Pattern Recogn.*, vol. 32, no. 1, pp. 71–86, 1999.

[13] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[14] M. Mellor and M. Brady, "Phase mutual information as a similarity measure for registration," *Med. Image Anal.*, vol. 9, pp. 330–343, 2005.

[15] G. Song, B. Avants, and J. Gee, "Multistart method with prior learning for image registration," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, 2007, pp. 1–8.

[16] A. Roche, G. Malandain, and N. Ayache, "Unifying maximum likelihood approaches in medical image registration," *Int. J. Imag. Syst. Technol.*, vol. 11, no. 1, pp. 71–80, 2000.

[17] E. Haber and J. Modersitzki, "Intensity gradient-based registration and fusion of multi-modal images," in *Proc. MICCAI*, 2006, pp. 726–733.

[18] J. Orchard, "Globally optimal multimodal rigid registration: An analytic solution using edge information," in *Proc. IEEE ICIP*, 2007, vol. 1, pp. I-485–I-488.

[19] J. Liu, B. Vemuri, and J. Marroquin, "Local frequency representations for robust multimodal image registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 5, pp. 462–469, May 2002.

[20] M. Hemmendorff, M. Andersson, T. Kronander, and H. Knutsson, "Phase-based multidimensional volume registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1536–1543, Dec. 2002.

[21] A. Wong and D. Clausi, "ARRSI: Automatic registration of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1483–1493, May 2007.

[22] C. Florin, J. Williams, A. Khamene, and N. Paragios, "Registration of 3D angiographic and X-ray images using sequential Monte Carlo sampling," in *Proc. CVBIA*, 2005, pp. 427–436.

[23] M. Ali and D. Clausi, "Automatic registration of SAR and visible band remote sensing images," in *Proc. IEEE IGARSS*, 2002, vol. 3, pp. 1331–1333.

[24] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *Proc. MICCAI*, 1998, vol. 1496, pp. 1115–1124.

[25] F. Candocia, "Jointly registering images in domain and range by piecewise linear comparametric analysis," *IEEE Trans. Image Process.*, vol. 12, no. 4, pp. 409–419, Apr. 2003.

[26] J. Orchard, "Efficient least-squares multimodal registration with a globally exhaustive alignment search," *IEEE Trans. Image Process.*, vol. 10, no. 16, pp. 2526–2534, Oct. 2007.

[27] J. Orchard, "Multimodal image registration using floating regressors in the joint intensity scatter plot," *Med. Image Anal.*, vol. 12, no. 4, pp. 385–396, 2008.

[28] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[29] P. Boggs and J. Tolle, "Sequential quadratic programming," in *Acta Numerica*, Cambridge, U.K.: Cambridge Univ. Press, 1995, pp. 1–51.

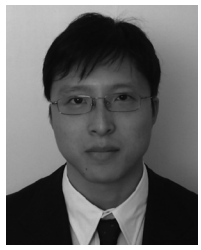
[30] K. Pearson, "Mathematical contributions to the theory of evolution, XIX: Second supplement to a memoir on skew variation," *Trans. Philos. R. Soc. Lond. Ser. A, Containing Papers Math. Phys. Character*, vol. 216, pp. 429–457, 1916.

[31] A. Wong, "An iterative approach to improved local phase coherence estimation," in *Proc. CRV*, 2008, pp. 301–307.

[32] P. Kovesi, "Phase congruency detects corners and edges," in *Proc. Aust. Pattern Recogn. Soc. Conf.*, 2003, pp. 309–318.

[33] M. Black and A. Rangarajan, "The outlier process: Unifying line processes and robust statistics," in *Proc. IEEE CVPR*, 1994, pp. 15–22.

[34] K. Johnson and J. Becker. (2008). The whole brain atlas [Online]. Available: http://www.med.harvard.edu/AANLIB/home.html



Alexander Wong (M'05) received the B.Sc. degree in computer engineering and the M.Sc. degree in electrical and computer engineering in 2005 and 2007, respectively, from the University of Waterloo, Waterloo, ON, Canada.

He is affiliated with the Vision and Image Processing Research Group, University of Waterloo, Waterloo, Canada. He has authored or coauthored refereed journal and conference papers in various fields, such as computer vision, graphics, image processing, biomedical signal processing, and multimedia systems. His current research interests include biomedical image processing and analysis, computer vision, and pattern recognition. He has been engaged with projects in image registration, image denoising, image super-resolution, image segmentation, biomedical tracking, biomedical image analysis, and image and video coding.