

High-Performance Medical Image Registration Using New Optimization Techniques

Mark P. Wachowiak, *Member, IEEE*, and Terry M. Peters, *Senior Member, IEEE*

Abstract—Optimization of a similarity metric is an essential component in intensity-based medical image registration. The increasing availability of parallel computers makes parallelizing some registration tasks an attractive option to increase speed. In this paper, two new deterministic, derivative-free, and intrinsically parallel optimization methods are adapted for image registration. Dividing RECTangles (DIRECT) is a global technique for linearly bounded problems, and multidirectional search (MDS) is a recent local method. The performance of DIRECT, MDS, and hybrid methods using a parallel implementation of Powell's method for local refinement, are compared. Experimental results demonstrate that DIRECT and MDS are robust, accurate, and substantially reduce computation time in parallel implementations.

Index Terms—Dividing RECTangles (DIRECT), medical image registration, multidirectional search (MDS), optimization, parallel computing.

I. INTRODUCTION

MANY RECENT registration methods, both linear and nonlinear, are intensity-based: some measure of similarity between images, computed directly from image intensities, is maximized. The robustness, accuracy, and efficiency of intensity-based registration depend on the similarity metric, search space (linear, nonlinear), interpolation scheme, and the similarity metric optimization approach. For optimization, the Nelder–Mead downhill simplex, Powell's direction set method, and first derivative-based methods such as conjugate gradient and Levenberg–Marquardt, have often been used [1], [2]. These techniques are generally local, with a limited capture range, and are susceptible to convergence to local optima, and thus to misregistration [3]. Additionally, derivatives of many metrics are not available, or cannot be easily estimated (although analytic derivatives exist for some metrics [4]).

Faster processors and memory access have substantially reduced registration time, but parallel computing offers the potential for further reduction, and for facilitating the use of techniques that were formerly considered too computationally expensive. In this paper, two new deterministic derivative-free methods are adapted for medical image registration: Dividing RECTangles (DIRECT) is a global technique for difficult linearly bounded optimization problems [5], while the multidirectional search (MDS) is a local method [6], [7]. Both

methods have well-established convergence properties [5], [7]. The inherent parallelism of DIRECT and MDS can be exploited for improving performance on distributed and shared memory architectures. Other investigators have used DIRECT in registration as a local step after stochastic optimization [8], but in this paper, DIRECT is employed as a global strategy, followed by local refinement with MDS. To our knowledge, MDS has not previously been applied to image registration.

This paper extends our previous work in optimization on distributed memory parallel systems [9] to shared memory systems, and presents parallel optimization for local refinement. Additionally, fine-grained operations, including interpolation and similarity metric computation, are parallelized in Powell's method. Local refinement with MDS, a local variant of DIRECT, and Powell's method are compared. The effect of interpolation is also considered.

II. PARALLEL REGISTRATION APPROACHES

The advantage of parallel computation can be measured by an algorithm's granularity, or computation-to-communication ratio. In fine-grained parallelism, several small tasks are performed on many processors. Although fine granularity can yield substantial efficiency gains (decrease in computation time), scaling, which quantifies efficiency gains with the number of processors, is limited because the communication and synchronization costs increase with the number of threads. In coarse granularity, the computation-to-communication ratio is high, and applications are more scalable. Coarse-grained applications generally run on distributed memory systems or clusters. Shared memory architectures are well-suited to medium-grained applications, although fine- and coarse-grained programs also run well on them.

In previous studies, resampling was multithreaded, and computing similarity metrics, segmentation, and visualization were parallelized on distributed memory clusters [10]. Parallelization of similarity metric computation (by image decomposition) and gradient estimation (by assigning threads to sets of transformations) greatly improved nonlinear registration performance [11].

Global optimization is amenable to coarse-grained parallelism because similarity metrics for several transformations can be computed simultaneously. Stochastic global optimization, which includes genetic algorithms [8], [12], [13], simulated annealing [13], and particle swarm optimization [14], often reduces incorrect registrations over purely local methods, but generally requires more similarity metric evaluations and computation time. In addition, because of stochasticity, multiple runs of the same registration may converge to different

Manuscript received January 26, 2005; revised June 13, 2005 and September 29, 2005. This work was supported by SHARCNet, NSERC R3146-A02, CIHR 14735, the Canadian Foundation for Innovation, the Ontario Innovation Trust, and the Ontario Consortium for Image-guided Therapy and Surgery.

The authors are with the Imaging Laboratories, Robarts Research Institute, London, ON N6A 5K8, Canada, and also with the Department of Medical Biophysics, University of Western Ontario, London, ON N6A 5C1, Canada (e-mail: mwach@imaging.robarts.ca; tpeters@imaging.robarts.ca).

Digital Object Identifier 10.1109/TITB.2005.864476

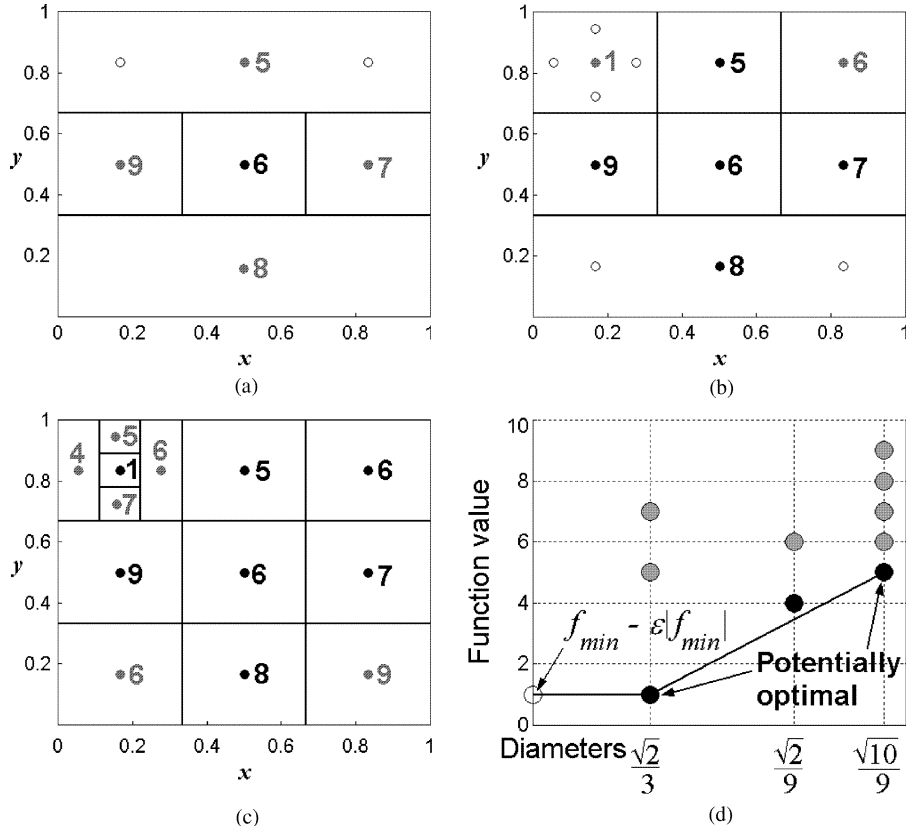


Fig. 1. 2-D division of rectangles in DIRECT. (a) Function values at $\{(x_0 \pm (1/3), y_0), (x_0, y_0 \pm (1/3))\}$, and divisions after the first iteration. (b) Divisions after the second iteration. (c) Division after the third iteration. (d) Convex hull after the third iteration, $\epsilon = 0.01$. New points evaluated in each iteration are shown in gray. Points to be evaluated around centers of potentially optimal rectangles in the next iteration are denoted as open circles.

transformations. Registration with stochastic approaches is also often sensitive to tuning parameters, such as crossover and mutation rates in genetic algorithms, or maximum velocity in particle swarm approaches [14].

In the coarse-grained DIRECT and MDS approaches, the entire evaluation of the similarity metric (applying a geometric transformation, interpolation, density estimation, and calculating the metric) is distributed to multiple processors so that metrics for many transformations can be computed simultaneously. These techniques are suitable for distributed or shared memory implementation, and are also extendible to include parallelization of other operations, such as resampling. For local refinement, a fine-grained parallelization of Powell's method is also presented, and compared with MDS.

A. Dividing Rectangles (DIRECT)

DIRECT [5] is a region-based approach that estimates finite bounds on the rate of change of the objective function (the Lipschitz constant). DIRECT balances global search, which identifies local minima, with local search, which finds the “basin of attraction” of the global minimum. Let n denote the number of dimensions in the search space. The search space is treated as an n -D rectangle, which is recursively divided into smaller rectangles. Rectangles are characterized by an integer level $l \geq 0$, with longest sides of length 3^{-l} , and short sides of at least $3^{-(l+1)}$. Starting with the center \mathbf{x}_0 of the initial rectangle,

(at level $l = 0$), the $2n$ points $\mathbf{x}_0 \pm 3^{-(l+1)} \mathbf{e}_i$, $i = 1, \dots, n$, are evaluated. \mathbf{e}_i denotes the i th column of $\mathbf{I}_{n \times n}$. The longest sides of the rectangle are divided into thirds, first along the dimension with the smallest function value, and these rectangles are then divided along the dimension with the next smallest value, and continuing until all the longest sides have been divided. The rectangles are then grouped by their l_2 -norm diameters

$$d(l, p) = 3^{-l} \sqrt{n - 8p/9} \quad (1)$$

where p is the number of short sides. The set of potentially optimal rectangles, or those that define the bottom of the convex hull of a scatter plot of rectangle diameters versus $f(\mathbf{x}_i)$ for all rectangle centers \mathbf{x}_i , are identified (Fig. 1(d)), and are selected for division in the next iteration. The slopes of the line segments connecting these points are estimates of the Lipschitz constant. The leftmost coordinate of the convex hull is $(0, f_{\min} - \epsilon|f_{\min}|)$, where f_{\min} is the current smallest function value and $\epsilon \geq 0$ (Fig. 1(d)). The smaller the value of ϵ , the more local the search. Because it uses these estimates implicitly, DIRECT is “Lipschitz optimization without the Lipschitz constant” [5].

Division of 2-D rectangles in DIRECT is illustrated in Fig. 1. The first function value, $f(x_0, y_0) = 6$, is obtained at the center $(x_0, y_0) = (1/2, 1/2)$ of a normalized rectangle. In the first iteration (Fig. 1(a)), the points $(x, y) \in \{(x_0 \pm 1/3, y_0), (x_0, y_0 \pm 1/3)\}$ are evaluated. The y dimension of the initial rectangle,

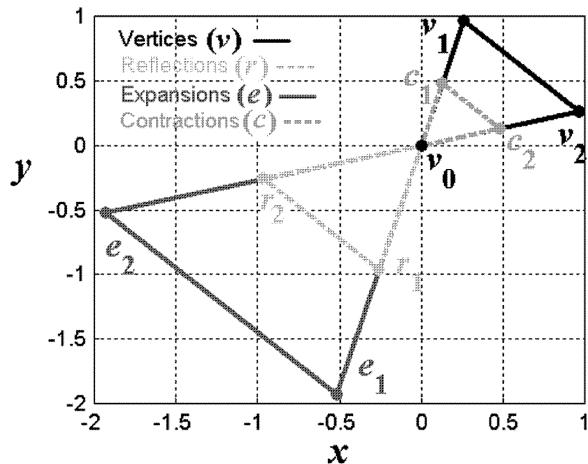


Fig. 2. Simplex transformations in MDS.

containing 5, is divided first. Divisions after the first iteration are shown in Fig. 1(a). The center of the large rectangle with $f(x_0, y_0 + ((1/3))) = 5$ is the only point on the convex hull, and is selected next for division. Divisions after the second iteration are shown in Fig. 1(b). The rectangles whose centers have values 8 and 1 are selected next. Divisions after the third iteration, and the resulting convex hull, are shown in Fig. 1(c) and (d), respectively. At each iteration, the maximum number of function evaluations is $2n$ times the number of points on the convex hull, representing the centers of rectangles that are candidates for division. Their function values can be evaluated in parallel. Full descriptions of DIRECT, as well as its theoretical behavior and convergence properties, are found in [5], [15]–[17].

A locally biased variant of DIRECT has been developed for low-dimensional problems with few local minima [17]. Rectangles are grouped by the l_∞ norm, $d(l, p) = 3^{-l}$, instead of by the l_2 norm.

B. Multidirectional Search (MDS)

Like the Nelder–Mead method, MDS employs a simplex of $n + 1$ n -D vertices. A new simplex is generated at each iteration based on the current best point v_0 . Whereas the edges connecting simplex vertices may become linearly independent in Nelder–Mead (see [18] for examples of functions for which the Nelder–Mead method results in a degenerate simplex), MDS requires that each new simplex is congruent to the previous one [15]. The initial simplex is constructed so that its vertices fully span the n -D space. Starting with $v_0 = (v_0, v_1, \dots, v_{n-1})^T$, one such construction is [19], [20]:

$$v_i = (v_0 + q_n, \dots, v_{i-1} + p_n, \dots, v_{n-1} + q_n)^T \quad (2)$$

where

$$p_n = \frac{\sqrt{n+1} - 1 + n}{n\sqrt{2}}, \quad q_n = \frac{\sqrt{n+1} - 1}{n\sqrt{2}} S \quad (3)$$

and S is a scaling factor. There can be a different S_i for every dimension i , depending on the relative scaling in that dimension. The simplex is modified by reflection, expansion, or contraction, as shown in Fig. 2. In each iteration, the simplex is reflected and

the function values at the new vertices are computed. If a new best vertex has been identified, an expansion step is performed. Otherwise, a contraction step is performed, and the new vertices are accepted. The vertex with the best value is relabeled as v_0 . MDS has two primary loops: an outer loop that determines a new set of search directions, and an inner loop that determines the length of the steps to be taken [7]. These step lengths are calculated with an expansion factor $\mu > 1$ and a contraction factor $\theta \in (0, 1)$. MDS is a descent method, since for each iteration k , $f(v_0^{k+1}) \leq f(v_0^k)$. Under general conditions, MDS has been proven to converge. A full description of MDS is found in [7], [15], [20].

If P processors are available, then each CPU could evaluate about $3n/P$ vertices in parallel. If $P \geq 3n$, then each reflection, contraction, and expansion can be computed simultaneously, in a speculative manner [21].

C. Global Optimization With Local Refinement

Due to its global scope, DIRECT often converges slowly. As the search progresses, the mesh of sample points becomes dense, and DIRECT effectively becomes an exhaustive search [15]. DIRECT performs well in the early iteration phases [5], [15], and in finding the vicinity of the global optimum. Efficient local methods may subsequently be used for refinement.

Powell’s direction set technique is most extensively used in derivative-free optimization for medical image registration [1]–[3]. A set of conjugate directions corresponding to each of the n dimensions is iteratively built, and line minimizations along these directions are performed (e.g., with Brent’s method) until stopping criteria are satisfied [22]. Powell’s method converges rapidly, and is highly accurate for initial points in the vicinity of the optimum. Although it is inherently serial, fine-grained (loop-level) operations are possible, as explained below. Alternatively, MDS may be used for local refinement. As a strictly local technique, MDS is not suitable for the early, global stages of the search [9]. DIRECT, MDS, and Powell’s method may be incorporated into a multiresolution framework, wherein DIRECT is employed for initial low-resolution searches, and local methods are used in the high-resolution stage.

III. METHODS

A. Similarity Metrics

The most popular and widely studied similarity metrics for intensity-based registration are those based on information theory [1], [3]. They have been used for multimodal (including functional and structural) registration (see [1] for a recent review). The Shannon mutual information is

$$MI \equiv I(X, Y) = H(X) + H(Y) - H(X, Y) \quad (4)$$

where $H(X)$ and $H(X, Y)$ respectively denote the entropy of the image intensities X , and the joint entropy of X and Y . Normalized mutual information (NMI) is less sensitive to the size of the source/target overlap, and is defined as [23]:

$$NMI \equiv \tilde{I}(X, Y) = \frac{H(X) + H(Y)}{H(X, Y)}. \quad (5)$$

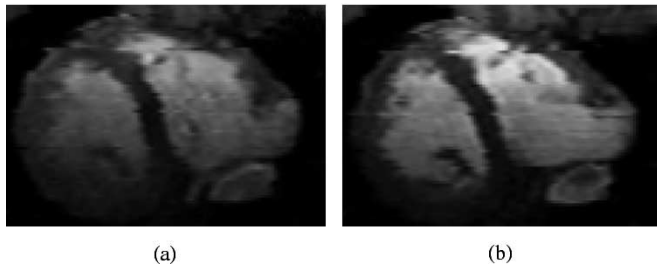


Fig. 3. Images used in cardiac registration experiments. Axial view of (a) source, phase three (systole). (b) Target volume, phase 15 (diastole).

B. Interpolation

Interpolation during registration greatly affects performance, particularly for information-theoretic metrics [24], [25]. As a goal of the current study is to reduce registration time, three relatively fast methods were selected for comparison: trilinear (or simply linear) (LI), partial volume (PV) [26], and jittered-blurred (JB) interpolation [24]. In LI, the nearest target neighbors to the source voxel are identified, and the resulting intensity of the target voxel is computed as a linear combination of the target neighbors. The histogram bin for this intensity is updated. PV interpolation was proposed to reduce oscillations in the similarity metric function, facilitating optimization. The bins for the neighbors are updated according to their distance from the source voxel. In JB interpolation, zero-mean Gaussian random variates with standard deviation 0.5 are added to the x , y , and z coordinates of the source voxel, and the bin corresponding to the intensity of the nearest neighbor is updated, resulting in a “jittering” effect. The final joint histogram is then filtered (“blurred”) [24]. While JB interpolation reduces oscillations, smaller “hills” and “valleys” are often introduced. However, smoothness in the registration function does not necessarily provide the most accurate or efficient registration [24].

The number of histogram bins also affects the registration function. Using the full range of intensities often increases oscillations and artifacts, while using too few bins may over-smooth and displace the location of the optimum. In previous work by the authors [3], [9], [14], 64 bins provided a compromise between smoothness and retaining important histogram features. Consequently, in this paper, all results were obtained with 64-bin histograms.

C. Data

The data for the registration experiments were simulated T1 and PD (proton density) MRI brain volumes generated with BrainWeb (www.bic.mni.mcgill.ca/brainweb) [27]; real MRI cardiac volumes of a human heart collected over 20 phases of the complete cardiac cycle (Fig. 3); and clinical T1 and T2 MRI brain images (Fig. 4). The BrainWeb reconstructions are highly realistic images, generated from known “ground truth” data, that are used extensively in the neuroscience community for validating segmentation and registration algorithms. They were selected to test robustness with respect to morphology dif-

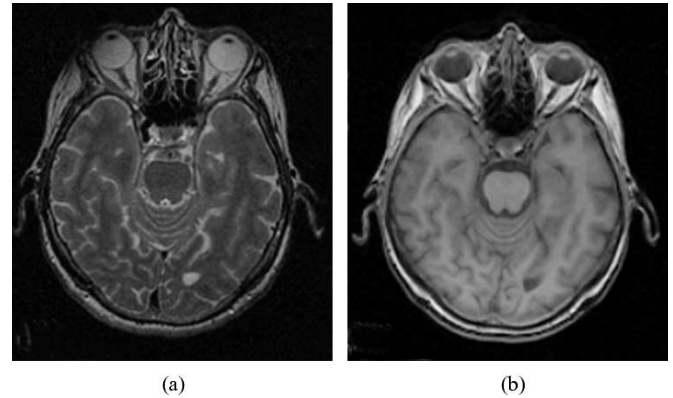


Fig. 4. Images used in clinical brain registration experiments. Axial view of (a) source volume (T2) and (b) target (T1).

ferences (normal tissue versus. multiple sclerosis regions) and noise (20% intensity nonuniformity and 3% noise relative to intensity for the PD volume, and 0% intensity nonuniformity and 0% and 9% noise for the T1 volumes). The noise statistics are Rayleigh in background regions, and Rician where signal is present. The noise percentage is the ratio of the standard deviations of the noise and signal for the reference tissue from which the images were generated.

The cardiac cine MRI volumes were selected at the diastolic and systolic phases of the cardiac cycle, as well as in mid-cycle, determined through prospective ECG gating. Each slice in each phase was acquired on the same breath-hold, with approximately 42 ms between phases. A set of volumes with minimal breathing motion artifacts was chosen for registering midcycle and peak systolic phases to the diastolic phase. As the position of the patient and the MRI scanner was constant throughout the acquisition, ground truth orientations for the cardiac volumes were known. The purpose is to globally align the volumes to facilitate successful nonrigid matching between the phases. Rigid registration of this type is useful as a preliminary step in validating dynamic heart models generated with nonrigid registration [28], and to study the effect of phase differences on cardiac image alignment [29]. A third set of experiments consisted of T1 and T2 clinical brain volumes generated from axial slices, obtained from a patient undergoing treatment for Parkinson’s disease. Ground truth orientation was obtained from 12 landmarks placed on a rigid frame attached to the patient’s skull.

The images are described below:

- 1) **MRI brain SOURCE:** PD MRI with multiple sclerosis lesions, 3% noise, size $101 \times 101 \times 35$ voxels, with voxel size 1.0 mm^3 .
 - a) **TARGET 1:** T1 MRI normal brain, 0% noise, $181 \times 217 \times 181$, voxel size 1.0 mm^3 .
 - b) **TARGET 2:** Same as TARGET 1, with 9% noise.
- 2) **MRI cardiac TARGET:** Peak diastole (phase 15), $256 \times 256 \times 75$, voxel size 1.5 mm^3 .
 - a) **SOURCE 1:** MRI cardiac, peak systole (phase 3), $169 \times 111 \times 5$, voxel size 1.5 mm^3 .

- b) **SOURCE 2:** MRI cardiac, midcycle (phase 11), $169 \times 111 \times 5$, voxel size 1.5 mm^3 .
- 3) **T2 MRI clinical brain SOURCE:** $229 \times 189 \times 30$, voxel size 1.0 mm^3 . **T1 MRI clinical brain TARGET:** $229 \times 189 \times 193$, voxel size 1.0 mm^3 .

D. Experiments

Two-level multiresolution rigid ($n = 6$) registration of 3-D to 3-D medical images were performed with:

- 1) DIRECT (D) throughout the registration process;
- 2) DIRECT in the low resolution stage, then high-resolution refinement with MDS (D/MDS);
- 3) DIRECT in the first stage, followed by high resolution refinement with Powell's method (D/P);
- 4) DIRECT in the first stage, then high-resolution refinement with locally-biased DIRECT (D/LD); and
- 5) Powell's method (P) throughout the registration.

We used MI and NMI as similarity metrics, which were optimized by minimizing their negative values. LI, PV, and JB interpolation were used during the registration process. For JB, the filter coefficients $[1/16 \ 4/16 \ 6/16 \ 4/16 \ 1/16]$ were applied along both axes of the joint histogram [24]. During the low resolution phase, the number of voxels used to compute the similarity metric was $1/2^3$ the full number of voxels in the high-resolution phase. From previous studies [3], [9], [14], [29], we observed that Gaussian blurring in the low resolution phase does not increase registration accuracy or efficiency, and hence simple downsampling was employed. To test scaling of parallelization, all experiments were performed with 1, 2, 4, 8, and 12 CPUs.

The source volumes were initially mistranslated at distances (d_0) of 5, 10, 20, and 30 mm from ground truth, and at initial misrotations of $\pm 2^\circ$, $\pm 10^\circ$, $\pm 20^\circ$, and $\pm 30^\circ$ about all three axes. For each d_0 and misrotation, the source was initially centered on four different points (32 experiments for each d_0).

Performance was judged on 1) the ratio of satisfactory registrations to the total number of registration experiments and 2) computation time. An experiment is considered successful if the translation error is less than 1 mm, and the maximum of the rotation errors along the x -, y -, and z -axes is less than $\pm 1^\circ$.

E. Parallelization

Once a new transformation is determined, the following steps are performed to compute the similarity metric.

- 1) Given a set of n transformation parameters, determine the transformation matrix T .
- 2) For all voxel coordinates S in the source image, apply the transformation: $S_T = TS$.
- 3) Determine the set of coordinates in S_T that fall within the target volume (S'_T).
- 4) Interpolate the intensities of the coordinates in the target corresponding to the intensity of each coordinate in S'_T , and subsequently estimate the joint probability density function by calculating the source-target joint histogram.
- 5) Compute the marginal densities of the source and target from the joint density.
- 6) Calculate the similarity metric.

For DIRECT and MDS, different threads perform the transformation, interpolation, and similarity metric computation for their own set of transformation parameters. Since each CPU performs a large amount of computation relative to the communication cost (passing the metric back to a "master" thread), coarse granularity is achieved. In "speculative" MDS, all of the simplex reflections, expansions, and contractions are computed in parallel, requiring $3n = 18$ CPUs for rigid transformations. An analysis of this fully-speculative MDS behavior revealed that synchronization overhead with 18 threads on a shared memory system reduced efficiency. Because MDS is used for local refinement, wherein the initial point \mathbf{v}_0 is already close to the optimum, fewer expansion (exploratory) steps are generally performed. In our analysis, we found that many more contractions than expansions were performed as the simplex contracted around the optimum. Thus, in our speculative MDS implementation, reflections and contractions were computed in parallel, requiring $2n = 12$ CPUs, making full utilization of all 12 CPUs at any given iteration more likely.

For Powell's method, loop-level (fine-grained) parallelism was applied to Steps 2–6 above. Through runtime profiling, we found that determining valid coordinates and calculating the joint histogram are particularly expensive, and can benefit from parallelization.

F. Implementation

All experiments were performed on an SGI Altix 3000 shared memory system with 20 1.3 GHz Intel Itanium-2 CPUs. The system has 20 GB main memory, and a 6 MB L3 floating point cache. The programs were written in C/C++, with parallelism implemented with OpenMP. Transformations (Step 2, described previously) were applied with BLAS numerical routines. DIRECT and MDS were implemented as described in [5] and [7], respectively. Powell's method with Brent's line minimization [22] was performed as described in [2]. For DIRECT, $\epsilon = 0.01$. The expansion and contraction factors for MDS were set respectively to $\mu = 2$ and $\theta = 1/2$, and the translation and rotation scales S_i were chosen as 2 mm and 8° , but MDS was robust with respect to other scale values.

With the current state of the art in imaging systems, it can be assumed that the initial orientation would normally lie within ± 5 cm of optimal alignment. Therefore, in DIRECT, the initial rectangle was chosen to be ± 5 cm from the initial point along all 3 axes, and rotations of $\pm \pi$ radians about the three axes.

G. Convergence Criteria

We chose the stopping/convergence criteria for DIRECT, MDS, and Powell's method to be as similar as possible. Powell's method was stopped if $2|f_p - f| \leq f_{\text{tol}}(|f_p| + |f|)$ [22], or if more than 500 iterations in Brent's line search had been expended. Here, f_p and f denote the previous and current function values, respectively, and $f_{\text{tol}} = 0.005$ is the tolerance at which two values are considered to be equal. For DIRECT, the best results were obtained with an empirical method based on the volume of the rectangles.

TABLE I
BRAINWEB SUCCESS RATE (ALL MISROTATIONS)

d_0 (mm)		NMI				MI					
		5	10	20	30	mean	5	10	20	30	mean
D	PV	1.00	0.89	1.00	0.87	0.94	0.98	0.74	0.82	0.66	0.80
	LI	0.99	0.92	0.98	0.95	0.96	0.96	0.73	0.78	0.68	0.79
	JB	0.96	0.59	0.65	0.32	0.63	0.96	0.46	0.36	0.16	0.48
P	PV	0.62	0.50	0.07	0.00	0.30	0.72	0.61	0.16	0.01	0.38
	LI	0.60	0.41	0.03	0.00	0.26	0.67	0.55	0.06	0.01	0.32
	JB	0.64	0.46	0.07	0.00	0.29	0.79	0.71	0.24	0.02	0.44
D/	PV	1.00	0.86	0.66	0.79	0.83	0.97	0.70	0.44	0.59	0.68
	M	0.98	0.84	0.80	0.79	0.85	0.94	0.57	0.46	0.46	0.61
	JB	0.95	0.57	0.64	0.33	0.62	0.92	0.44	0.35	0.17	0.47
D/	PV	0.98	0.73	0.65	0.68	0.76	0.98	0.72	0.45	0.59	0.68
	P	0.95	0.73	0.77	0.75	0.80	0.94	0.59	0.46	0.47	0.62
	JB	0.93	0.36	0.36	0.35	0.50	0.93	0.36	0.33	0.28	0.47
D/	PV	1.00	0.86	1.00	0.82	0.92	0.98	0.68	0.79	0.61	0.76
	LD	0.98	0.89	0.98	0.92	0.94	0.96	0.71	0.74	0.64	0.76
	JB	0.92	0.54	0.48	0.32	0.56	0.96	0.52	0.40	0.25	0.53

TABLE II
CARDIAC SUCCESS (ALL MISROTATIONS). TARGET AT PEAK DIASTOLE

d_0 (mm)		NMI				MI					
		5	10	20	30	mean	5	10	20	30	mean
D	PV	0.91	0.90	0.95	0.95	0.93	0.95	0.92	0.96	0.97	0.95
	LI	0.89	0.88	0.95	0.94	0.92	0.91	0.92	0.95	0.94	0.93
	JB	0.95	0.94	0.97	0.94	0.95	0.98	0.96	0.95	0.89	0.95
P	PV	0.73	0.66	0.50	0.46	0.59	0.80	0.64	0.55	0.55	0.64
	LI	0.52	0.48	0.33	0.28	0.40	0.52	0.47	0.31	0.31	0.40
	JB	0.91	0.79	0.71	0.66	0.77	0.92	0.93	0.81	0.77	0.86
D/	PV	0.99	0.93	0.96	0.86	0.93	0.99	0.95	0.94	0.88	0.94
	M	0.78	0.45	0.78	0.73	0.69	0.87	0.64	0.87	0.81	0.80
	JB	0.96	0.87	0.87	0.84	0.88	0.95	0.91	0.92	0.79	0.89
D/	PV	0.93	0.94	0.95	0.94	0.94	0.97	0.95	0.96	0.96	0.96
	P	0.91	0.81	0.96	0.90	0.89	0.96	0.91	0.96	0.93	0.94
	JB	0.91	0.80	0.87	0.77	0.84	0.99	1.00	1.00	0.82	0.95
D/	PV	0.86	0.87	0.95	0.95	0.91	0.91	0.92	0.96	0.97	0.94
	LD	0.89	0.86	0.94	0.94	0.91	0.91	0.89	0.95	0.94	0.92
	JB	0.94	0.97	0.95	0.87	0.93	0.94	0.90	0.94	0.93	0.93

- 1) Determine the volumes generated by the smallest rectangle using translations (V_t) and rotations (V_r).
- 2) If $V_t < 0.001 \text{ mm}^3$ and $V_r < 0.0001^\circ$, then a) If $2|f_p - f| \leq f_{\text{tol}}(|f_p| + |f|)$, stop; otherwise, increment the number of nonimproving iterations. b) Stop if there is no improvement in four iterations.

Considering rectangle volumes ensures that the search space has been adequately sampled. If there is no improvement with small V_r and V_t , then DIRECT is likely “trapped,” and local search is then employed. The maximum number of DIRECT iterations was set to 1000.

For MDS, stopping criteria were the same as for Powell’s method, with an additional criterion based on the maximum distance between the three translation components of simplex vertices: $(1/\Delta) \max_{1 \leq i \leq 3} \|\mathbf{v}_i^k - \mathbf{v}_0^k\| < \epsilon$ [20], where $\Delta = \max\{1, \|\mathbf{v}_0\|\}$, and $\epsilon = 0.5$ was empirically found to work well.

IV. RESULTS

A. Success Rates

The success rates for the simulated brain MRI, cardiac MRI, and clinical brain registrations are shown in Tables I, II, and III, respectively. The success rates for PD to T1 (9% noise) and peak systole-peak diastole experiments for large initial misrotations (30°) are shown in Table IV.

TABLE III
CLINICAL BRAIN SUCCESS RATE (ALL MISROTATIONS)

d_0 (mm)		NMI				MI					
		5	10	20	30	mean	5	10	20	30	mean
D	PV	1.00	1.00	0.97	1.00	0.99	0.98	1.00	0.93	0.99	0.97
	LI	0.97	0.99	0.96	0.98	0.97	0.98	0.99	0.96	0.97	0.97
	JB	0.96	0.99	0.95	0.99	0.97	0.99	0.99	0.91	0.80	0.92
P	PV	0.86	0.79	0.55	0.31	0.63	0.93	0.91	0.71	0.71	0.81
	LI	0.93	0.84	0.54	0.33	0.66	0.97	0.91	0.79	0.69	0.84
	JB	0.74	0.68	0.46	0.26	0.53	0.90	0.91	0.83	0.58	0.80
D/	PV	0.73	1.00	0.79	0.93	0.86	0.77	0.96	0.85	0.90	0.87
	M	0.66	0.79	0.69	0.77	0.73	0.73	0.86	0.78	0.83	0.80
	JB	0.70	0.66	0.59	0.74	0.67	0.83	0.72	0.63	0.86	0.76
D/	PV	0.71	0.89	0.70	0.81	0.78	0.78	0.96	0.85	0.89	0.87
	P	0.69	0.86	0.72	0.73	0.75	0.74	0.97	0.88	0.90	0.87
	JB	0.61	0.66	0.68	0.70	0.66	0.83	0.94	0.84	0.86	0.87
D/	PV	0.88	1.00	0.88	0.98	0.93	0.85	0.99	0.80	0.92	0.89
	L	0.89	0.99	0.91	0.97	0.94	0.82	0.98	0.82	0.91	0.88
	JB	0.91	0.92	0.76	0.97	0.89	0.91	0.81	0.92	0.91	0.89

TABLE IV
SUCCESS RATES (JB AND NMI) FOR PD TO T1 (9% NOISE) BRAIN MRI REGISTRATION AND FOR PEAK SYSTOLE TO PEAK DIASTOLE CARDIAC MRI REGISTRATION, 30° INITIAL MISROTATION

Volume	Opt.	d_0 (mm)				
		5	10	20	30	Overall
Brain MRI	D	1.00	0.72	1.00	0.66	0.84
	P	0.16	0.03	0.00	0.00	0.05
	D/M	1.00	0.53	0.69	0.66	0.72
	D/P	1.00	0.47	0.50	0.53	0.63
	D/L	1.00	0.69	1.00	0.69	0.84
Cardiac MRI	D	1.00	1.00	1.00	0.97	0.99
	P	0.59	0.53	0.34	0.25	0.43
	D/M	1.00	0.88	0.94	0.88	0.92
	D/P	1.00	1.00	1.00	0.94	0.98
	D/L	1.00	1.00	1.00	1.00	1.00

Because the assumptions of normality and equal variances were not strictly met, Wilcoxon rank sum tests were performed to assess statistical significance at the $\alpha = 0.05$ level. For interpolation over both MI and NMI ($N = 768$ for all experiments), for the BrainWeb data, there was no significant difference between the success rates for PV and LI for either NMI or MI, using the DIRECT methods. For Powell’s method, PV was significantly better than LI. For the cardiac experiments, PV was significantly better than LI for all methods, except D/LD. There was no significant difference between PV and JB for DIRECT and D/LD, but PV was significantly better than JB for D/MDS and for D/P. For Powell’s method, JB was significantly better than PV. For the clinical brain images, there was no significant difference among the interpolation methods, except for D/MDS, where PV was better than LI and JB, and D/P, where PV was better than JB.

Comparing the effect of the similarity metric with PV interpolation ($N = 384$), for the BrainWeb data, NMI had a significantly higher success rate than MI for all DIRECT methods, but MI was better than NMI with Powell’s method. However, for the cardiac images, MI was significantly better than NMI for all methods except D/MDS. For the clinical brain images, significant differences were observed for P and D/MDS (MI better) and for D/LD (NMI better).

Finally, comparing optimization approaches by similarity metric, and using PV interpolation ($N = 384$) for all experiments, for the brain experiments with NMI, DIRECT had a significantly higher success rate than P, D/MDS, and D/P (no

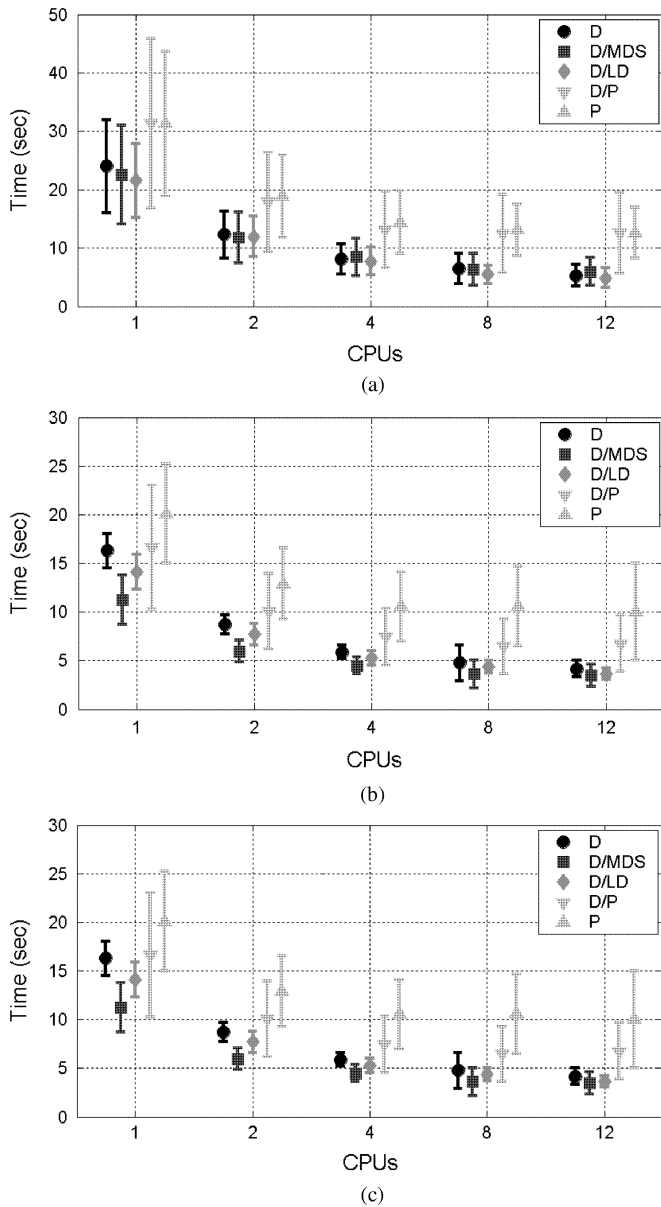


Fig. 5. Timing results using MI. (a) BrainWeb MRI PD-T1 (9% noise), PV. (b) Systole-diastole cardiac, JB. (c) Clinical brain, PV.

difference with D/LD). With MI, however, DIRECT was significantly better than all the other methods, with D/LD having the success rate closest to that of DIRECT. For the cardiac experiments, there was no significant difference between DIRECT and any of the other DIRECT methods for either metric. DIRECT was significantly better than Powell's method. For the clinical brain experiments, DIRECT was significantly better than all other methods, for both MI and NMI. From this analysis, it is observed that 1) PV generally results in the highest success rate, but for the cardiac volume, PV and JB performed equally; 2) there is no clear advantage for either NMI or MI, as performance of the metrics was dependent on the images registered, except for Powell's method, where MI has a higher success rate than NMI; and 3) DIRECT either outperforms other methods, or there is no significant difference between DIRECT and its

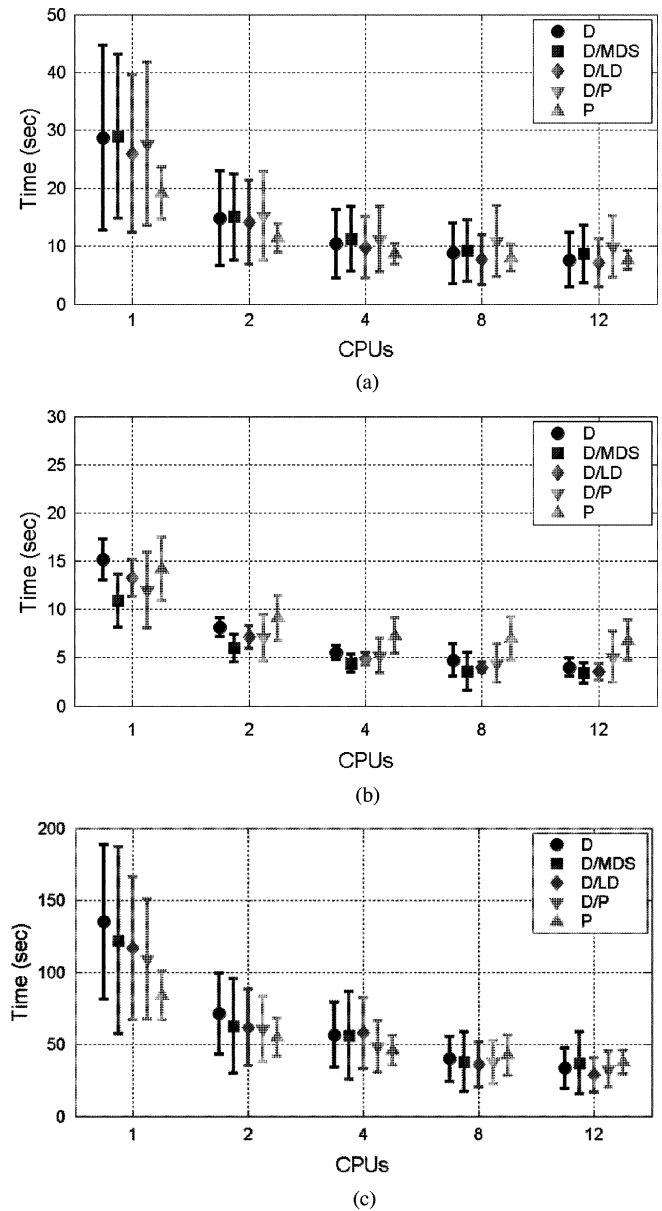


Fig. 6. Timing results using NMI. (a) BrainWeb MRI PD-T1 (9% noise), PV. (b) Systole-diastole cardiac, JB. (c) Clinical brain, PV.

hybrids. However, DIRECT has a significantly higher success rate than Powell's method.

B. Efficiency

Timing results for the various methods are shown in Figs. 5 and 6. For the BrainWeb and clinical brain experiments (testing the effect of noise), PV had the highest success rates and, for the cardiac experiments (testing morphology difference effects), JB generally had high success rates and, consequently, those results are shown. Both DIRECT and D/MDS were faster than P and D/P when used with four or more processors. DIRECT and D/MDS also scale better than Powell's method or D/P (Fig. 7). DIRECT and local DIRECT scale particularly well in the local optimization stage, followed by speculative

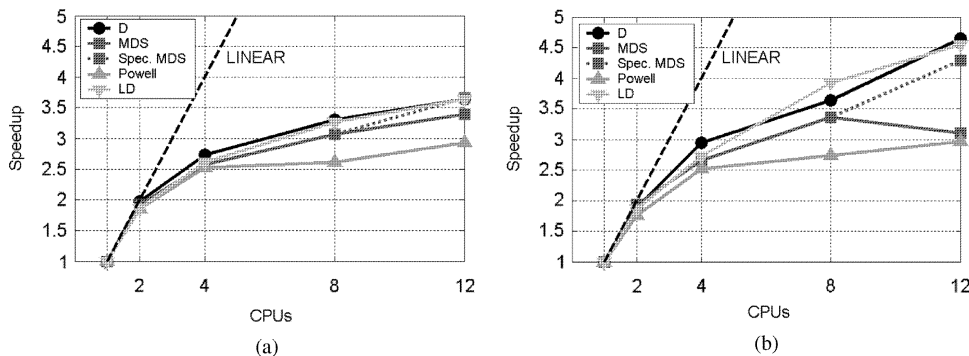


Fig. 7. Speedup for optimization methods, NMI, and PV interpolation. (a) PD-T1 (9% noise) registration of MRI brain images. (b) Systole-diastole registration of MRI cardiac images.

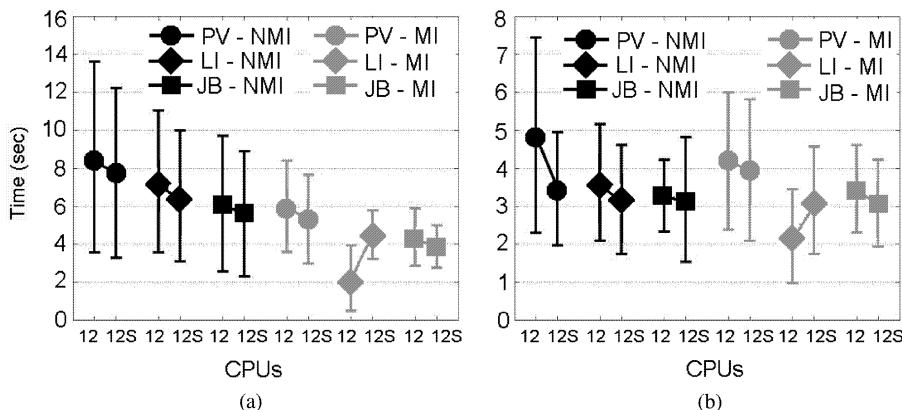


Fig. 8. Timing results for DIRECT/MDS and DIRECT/Speculative MDS (12S) using 12 CPUs. (a) PD-T1 (9% noise) registration of MRI brain images. (b) Systole-diastole registration of MRI cardiac images.

MDS and MDS. Although MDS has coarser granularity than DIRECT, and therefore should scale better, we found that there is a large number of points on the convex hull to be evaluated in the local stage with DIRECT, slightly outweighing the low overhead with MDS. For MDS, the synchronization costs for idle threads increased when more than 8 CPUs were used, but this limitation was overcome with speculative MDS, where 12 CPUs were highly utilized in each iteration (Fig. 8).

V. DISCUSSION

Although Powell's method and gradient techniques are robust and accurate, especially for good initial orientations, DIRECT is more robust for greater initial misregistrations (>10 mm, or $>10^\circ$ initial misorientation). As seen in Table IV, Powell's method has a very low success rate for large initial misrotations of $\pm 30^\circ$. As a strictly local algorithm, MDS is best suited to improvement in the basin of attraction of the optimum. [9]. The performance of DIRECT and DIRECT hybrids is not dependent on the similarity metric or interpolation method. The new techniques had higher success rates than Powell's method alone for both metrics and for all three interpolation methods. The nonsmoothness and local optima generally present with information-theoretic measures contribute to the smaller capture range of Powell's method, a limitation overcome with the more global scope of DIRECT.

Additionally, with current clinical imaging technology, the initial transformation (prior to any registration) may be very close to the optimal alignment. Registration is employed only to "fine-tune" the initial guess, especially where real-time performance is required. In these cases, although Powell's or gradient methods are good choices, MDS should also be considered, due to its coarse granularity and good scaling properties. To improve efficiency on parallel hardware, an algorithm must scale well. The two new derivative-free methods presented in this paper parallelize in a natural way, employing coarse granularity, whereas (parallel) Powell's method employs fine-grained parallelism at the expense of more thread communication overhead and synchronization. Thus, DIRECT and MDS scale better than Powell's method (Fig. 7). Powell's method achieves fine-grained parallelism through computation of the similarity metric and, therefore, scalability is limited. For the DIRECT methods, close to linear scaling is observed when two or four CPUs are used. However, scaling is generally sub-linear for all the optimization techniques, including DIRECT. This behavior is expected, as DIRECT is a complex algorithm with substantial serial overhead. Timing results (Figs. 5 and 6) show that efficiency gains can be achieved for DIRECT, D/P, and D/MDS if as few as four CPUs are available. Although modest gains are seen with 8 or 12 CPUs, using a large number of CPUs is not expected to markedly reduce computation time.

Powell's methods (P and D/P) are slower than DIRECT and D/MDS for MI (Fig. 5). However, with NMI and 1–4 CPUs, Powell's method is substantially faster than the DIRECT methods, which become competitive when more than four CPUs are used (Fig. 6). Experimentally, NMI has been shown to be more conducive to optimization than MI, although the latter is frequently more robust and accurate [3]. Additionally, efficiency gains with DIRECT and DIRECT hybrids are expected only when used in parallel, and efficiency for all methods is heavily influenced by the cost function.

Because the similarity metric computation times do not vary widely among transformations, load balancing strategies [16] are not expected to significantly improve performance. It is, however, possible to parallelize (in a fine-grained manner) some overhead operations in DIRECT [16]. Efficiency will further improve with faster processors and with a larger floating point cache.

VI. CONCLUSION

In this paper, we presented coarse-grained optimization approaches for medical image registration, wherein multiple similarity metrics were computed in parallel. The primary advantages of DIRECT and its hybrids are global exploration of the search space, and increased capture range. Experimental results suggest that multiresolution registration with DIRECT is the most successful. When a small number of CPUs is available (eight or less), combining DIRECT with MDS reduces computation time while maintaining a high success rate. If at least $2n$ CPUs are available, DIRECT/speculative MDS should be considered.

The DIRECT methods were more successful than Powell's method for all interpolation methods, and for both similarity metrics. PV interpolation was marginally more successful than LI or JB interpolation in most cases. Performance was also influenced by the metric, and which metric to use for specific image characteristics or modalities merits further study [1].

In parallel implementation, the new methods generally require less computation time than (parallelized) Powell's method, especially when mutual information is used. While some of the observed decreases (e.g., from 30 to 10 s) may not be significant for registrations taking less than 1 min, efficiencies of this order can be important for more complex procedures lasting minutes, especially in operating room environments. Additionally, efficiency gains are important when the objective is to achieve near real-time performance.

Although the new optimization methods were performed for rigid registration (an important global step prior to nonrigid matching), the coarse-grained approaches may also be extended to some nonrigid registration techniques. For example, some approaches compute deformation vectors from small blocks of voxels at different resolutions that are matched rigidly. Thus, similarity metrics are computed for each block, rather than globally [28]. Other nonrigid methods using stochastic global optimization (e.g., [30]) may also benefit from DIRECT and MDS. Hence, DIRECT and MDS should be considered as compliments to existing methods.

These new optimization methods can be executed on multiprocessor systems or distributed memory clusters, with computed transformations sent over high-speed networks to quickly update visual displays and interoperative data, and are therefore suitable for use in image-guided surgery. Future work includes applying DIRECT and MDS to nonrigid registration; additional validation on clinical data, including ultrasound; further improvement of robustness by application of multistart techniques; and testing with other similarity metrics; and incorporating DIRECT/MDS into image-guidance frameworks, including cardiac intervention planning platforms.

ACKNOWLEDGMENT

The authors would like to thank K. Wang, C. Wedlake, and B. Ge for technical assistance; J. Moore, T. Guo and the Montréal Neurological Institute for providing images; and Dr. R. Smolikova-Wachowiak for useful discussions. The authors also thank the anonymous reviewers for valuable suggestions.

REFERENCES

- [1] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
- [2] F. Maes, D. Vandermeulen, and P. Suetens, "Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information," *Med. Image Anal.*, vol. 3, pp. 373–386, 1999.
- [3] M. P. Wachowiak, R. Smolková, and T. M. Peters, "Multiresolution biomedical image registration using generalized information measures," in *Proc. MICCAI 2003, Montréal, Canada, Ser. LNCS*, R. E. Ellis and T. M. Peters, Eds., vol. 2879. New York: Springer-Verlag, 2003, pp. 846–853.
- [4] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multimodal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, pp. 35–52, 1996.
- [5] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian Optimization without the Lipschitz Constant," *J. Optim. Theory Appl.*, vol. 79, pp. 157–181, 1993.
- [6] T. G. Kolda, R. M. Lewis, and V. Torczon, "Optimization by direct search: New perspectives on some classical and modern methods," *SIAM Rev.*, vol. 45, pp. 385–482, 2003.
- [7] V. Torczon, "On the convergence of the multidirectional search algorithm," *SIAM J. Optim.*, vol. 1, pp. 123–145, 1991.
- [8] R. He and P. A. Narayana, "Global optimization of mutual information: Application to three-dimensional retrospective registration of magnetic resonance images," *Comput. Med. Imaging Graph.*, vol. 26, pp. 277–292, 2002.
- [9] M. P. Wachowiak and T. M. Peters, "Parallel optimization approaches for medical image registration," in *Proc. MICCAI 2004, St. Malo, France, Ser. LNCS*, C. Barillot, D. R. Haynor, and P. Hellier, Eds., vol. 3216. New York: Springer-Verlag, 2004, pp. 781–788.
- [10] S. K. Warfield, F. A. Jolesz, and R. Kikinis, "A high performance computing approach to the registration of medical imaging data," *Parallel Comput.*, vol. 24, pp. 1345–1368, 1998.
- [11] T. Rohlfing and C. R. Maurer, "Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 1, pp. 16–25, Mar. 2004.
- [12] M. Jenkinson and S. Smith, "A global optimization method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, pp. 143–156, 2001.
- [13] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Trans. Inf. Technol. Biomed.*, vol. 3, no. 1, pp. 47–60, Mar. 1999.

- [14] M. P. Wachowiak, R. Smolíková, Y. Zheng, J. M. Zurada, and A. S. Elmaghraby, "An approach to multimodal biomedical image registration utilizing particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 289–301, Jun. 2004.
- [15] C. T. Kelley, *Iterative Methods for Optimization*. Philadelphia, PA: SIAM, 1999.
- [16] L. T. Watson and C. A. Baker, "A fully-distributed parallel global search algorithm," *Eng. Comput.*, vol. 18, pp. 155–169, 2001.
- [17] J. M. Gablonsky and C. T. Kelley, "A locally-biased form of the direct algorithm," *J. Global Optim.*, vol. 21, pp. 27–37, 2001.
- [18] K. I. M. McKinnon, "Convergence of the Nelder-Mead simplex method to a nonstationary point," *SIAM J. Optim.*, vol. 9, pp. 148–158, 1998.
- [19] S. L. S. Jacoby, J. S. Kowalik, and J. T. Pizzo, *Iterative Methods for Nonlinear Optimization Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1972.
- [20] V. Torczon, "Multi-directional search: A direct search algorithm for parallel machines," Ph.D. dissertation, Dept. Math. Sci., Rice Univ., May 1989.
- [21] J. E. Dennis and V. Torczon, "Direct search methods on parallel machines," *SIAM J. Optim.*, vol. 1, pp. 448–474, 1991.
- [22] W. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C. The Art of Scientific Computing*, 2nd ed., Cambridge, U.K.: Cambridge Univ. Press, 1992.
- [23] C. Studholme, D. J. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, pp. 71–86, 1999.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] R. K. S. Kwan, A. C. Evans, and G. B. Pike, "MRI simulation-based evaluation of image processing and classification methods," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1085–1097, Nov. 1999.
- [28] M. Wierzbicki, M. Drangova, G. Guiraudon, and T. M. Peters, "Validation of dynamic heart models obtained using non-linear registration for virtual reality training, planning, and guidance of minimally invasive cardiac surgeries," *Med. Image Anal.*, vol. 8, pp. 387–401, 2004.
- [29] R. Smolíková-Wachowiak, M. P. Wachowiak, A. Fenster, and M. Drangova, "Registration of two-dimensional cardiac images to preprocedural three-dimensional images for interventional applications," *J. Magn. Reson.*, vol. 22, pp. 219–228, 2005.
- [30] N. P. Castellanos, P. L. D. Angel, and V. Medina, "Nonrigid medical image registration technique as a composition of local warpings," *Pattern Recognit.*, vol. 37, pp. 2141–2154, 2004.



Mark P. Wachowiak (S'99–M'02) received the M.S. degree in computer science and the Ph.D. degree in computer science and engineering from the University of Louisville, Louisville, KY, in 1997 and 2002, respectively.

He is currently a Researcher in the Imaging Laboratories at the Robarts Research Institute in London, ON, Canada. His research interests include computational biophysics, mathematical modeling of biological and physiological systems, medical imaging, signal processing, systems biology, proteomics, high-performance computing, and computational science. Most recently, he has been developing parallel and high-performance techniques for applications in image-guided minimally invasive therapy and surgery.



Terry M. Peters (M'64–SM'97) received the B.E. (Honors) and the Ph.D. degrees in electrical engineering from the University of Canterbury, Christchurch, New Zealand, in 1969 and 1974, respectively.

Currently, he is a Scientist in the Imaging Research Laboratories at the Robarts Research Institute (RRI), London, ON, Canada, and Professor in the Departments of Radiology and Nuclear Medicine and Medical Biophysics at the University of Western Ontario. For the past 20 years, his research has focused on the application of computational hardware and software to medical imaging modalities in surgery and therapy. Starting at the Montréal Neurological Institute, his laboratory pioneered many of the concepts and applications used for the surgical treatment of epilepsy and Parkinson's Disease. At RRI, he established an image-guided surgery and therapy group. He has authored over 150 peer-reviewed papers and book chapters, and has delivered over 130 invited presentations.

Dr. Peters is a Fellow of the Canadian College of Physicists in Medicine. He is an Associate Editor of *IEEE TRANSACTIONS ON MEDICAL IMAGING* and a Member of the Editorial Boards of *Magnetic Resonance Imaging; Computer Methods in Biomechanics and Biomedical Engineering; Australasian Physical and Engineering Sciences in Medicine; Medical Image Analysis; and Current Medical Imaging Reviews*. He is an Executive Member of the Board of the MICCAI Society, as well as its Treasurer.