

Combining global and local parallel optimization for medical image registration

Mark P. Wachowiak^a and Terry M. Peters^{a,b}

^aRobarts Research Institute, 100 Perth Drive, London, ON, Canada;

^bUniversity of Western Ontario, London, ON, Canada

ABSTRACT

Optimization is an important component in linear and nonlinear medical image registration. While common non-derivative approaches such as Powell's method are accurate and efficient, they cannot easily be adapted for parallel hardware. In this paper, new optimization strategies are proposed for parallel, shared-memory (SM) architectures. The Dividing Rectangles (DIRECT) global method is combined with the local Generalized Pattern Search (GPS) and Multidirectional Search (MDS) and to improve efficiency on multiprocessor systems. These methods require no derivatives, and can be used with all similarity metrics. In a multiresolution framework, DIRECT is performed with relaxed convergence criteria, followed by local refinement with MDS or GPS. In 3D-3D MRI rigid registration of simulated MS lesion volumes to normal brains with varying noise levels, DIRECT/MDS had the highest success rate, followed by DIRECT/GPS. DIRECT/GPS was the most efficient (5–10 seconds with 8 CPUs, and 10–20 seconds with 4 CPUs). DIRECT followed by MDS or GPS greatly increased efficiency while maintaining accuracy. Powell's method generally required more than 30 seconds (1 CPU) with a low success rate (0.3 or lower). This work indicates that parallel optimization on shared memory systems can markedly improve registration speed and accuracy, particularly for large initial misorientations.

Keywords: Registration, optimization, dividing rectangles, multidirectional search, generalized pattern search, parallel computing, high performance computing

1. INTRODUCTION

Image registration, or the process of geometrically aligning images and volumes from the same or different modalities, is a key component in image-guided surgery and therapy, pre-procedural planning, assessing therapy response, developing patient-specific models, and in basic biomedical research. Because of its importance in imaging applications, registration remains an active research endeavor, especially in such areas as similarity metrics,^{1,2} cardiac registration,^{3,4} and nonlinear registration for modeling and surgical planning.⁵

Many recent registration methods are intensity-based. That is, segmentation, surface extraction, and other pre-processing steps are not required prior to registration. In these methods, some measure of similarity between images, computed only from image intensities, is maximized. The robustness, accuracy, and efficiency of intensity-based registration depend on selecting an appropriate similarity metric suited to the particular application (e.g. statistical, information-theoretic methods), the search space (e.g. linear or nonlinear), the interpolation scheme (linear, partial volume, nonlinear), and the approach used to optimize the similarity metric. For optimization, the Nelder-Mead downhill simplex, Powell's direction set method, and techniques requiring derivatives, such as conjugate gradient, Newton's method, or trust region techniques (e.g. Levenberg-Marquardt) have often been used.^{2,6} Gradient descent methods are preferred if accurate first derivatives can be obtained. However, these techniques are generally local, with a limited capture range, and are susceptible to premature convergence to local optima, and therefore, to misregistration.¹ Furthermore, many of these methods are not conducive to parallel implementation. In derivative-based optimization, partial derivatives (gradients) may be computed simultaneously, but, because of noise or irregularity in many similarity metric functions (especially those based on information theory), gradients often cannot be easily estimated, e.g. from finite differences.

Further author information: (Send correspondence to Mark P. Wachowiak)

Mark P. Wachowiak: E-mail: mwach@imaging.robarts.ca, Telephone: 1 519 663 5777 x34280

Faster processors (central processing units, or CPUs) and memory access have substantially reduced registration time, but parallel computing can potentially further increase efficiency, and facilitate the use of optimization techniques that were formerly considered too computationally expensive. These techniques can decrease computation time because different regions of the search space are explored simultaneously. In this paper, three such methods – DIRECT, the generalized pattern search, and the multidirectional search, are parallelized, and adapted for medical image registration. Dividing RECTangles, or DIRECT, is a global technique that was designed for difficult linear bound-constrained optimization problems.⁷ The generalized pattern search (GPS) evaluates points in the search space by sampling a mesh, or pattern, of points around a base point.⁸ Finally, the multidirectional search, or MDS, is a special case of GPS where points are evaluated on a simplex.⁹ Both GPS and MDS are strictly local methods. Therefore, DIRECT, because of its global search capabilities, can be used to sample a large region of the search space, followed by application of GPS or MDS for local refinement. The intrinsic parallelism of DIRECT, GPS, and MDS can be easily exploited for improving registration performance on both distributed and shared memory architectures. Other investigators have used DIRECT as a final, local step after stochastic optimization for registration.¹⁰ In this paper, however, DIRECT is employed as a bound-constrained global strategy. Because of its convergence properties, GPS and MDS are used for local refinement of parameters obtained with DIRECT, and the results are compared with local refinement with Powell’s method. To our knowledge, GPS and MDS have not previously been applied to image registration.

We have previously shown that parallel implementations of DIRECT can increase capture range and substantially reduce computation time on distributed memory architectures.¹¹ In the current work, we extend these concepts to include local refinement, and present registration performance on shared memory architectures. In addition, we assess the effect of interpolation methods on robustness and efficiency.

2. PARALLEL OPTIMIZATION IN IMAGE REGISTRATION

The performance advantage of parallel computation can be measured by an algorithm’s granularity, or the ratio of computation to communication. In fine-grained parallelism, several small computation tasks are performed on many processors. Although fine-grained algorithms can be implemented on most parallel architectures, they are best suited to vector machines. A drawback of fine granularity is that the ratio of computation to communication is usually low due to the cost of communication and synchronization, which increases with the number of threads. In coarse-grained parallelism, the computation-to-communication ratio is much higher, and applications are more scalable; that is, computation speed increases with the number of threads, assuming that the threads execute on different processors. Coarse-grained parallel programs generally run on distributed memory systems or clusters of computers. There is a continuum of granularity, and medium-grained applications are well-suited to shared memory architectures, although fine- and coarse-grained applications also run well on them.

Parallelism has been previously exploited by other investigators to improve registration speed for clinical applications. For example, fine-grained operations such as resampling and similarity metric computations have been multithreaded, and segmentation and visualization are parallelized on distributed memory clusters.¹²

Coarse granularity is achieved with inherently parallel optimization techniques. Such methods include stochastic global approaches, including genetic algorithms,^{10,13,14} simulated annealing,¹⁴ and particle swarm optimization.¹⁵ Although these methods are less prone to entrapment in local optima, they require many evaluations of the objective function. In addition, because of their stochasticity, multiple runs of the same task may yield different results. Another disadvantage is that performance is dependent on parameters that, in the case of genetic algorithms, include spopulation size, crossover and mutuation rates, and number of generations.

In DIRECT, GPS, and MDS, all of which are deterministic, the entire evaluation of the similarity metric can be distributed to multiple processors so that metrics for many transformations in the search space can be computed simultaneously. Similarity metric computation in linear registration consists of: (1) Determining the transformation matrix T from n transformation parameters. (2) For all n -D coordinates S in the source image, applying the transformation: $S_T = TS$. (3) Discarding the coordinates in S_T that fall outside the target volume. (4) Interpolating the intensities of the coordinates in the target and estimating the joint probability density function from the source-target joint histogram. (5) Computing the target and source marginal densities from the joint density. (6) Calculating the metric.

Such coarse-grained approaches are suitable for distributed, shared memory, and hybrid implementation, and are also extendible to include parallelization of fine-grained operations, such as resampling.

Following the optimization literature, optimization is considered as minimization, and, in the context of image registration, the goal is minimization of the negative of a specified similarity metric.

2.1. Dividing Rectangles (DIRECT)

DIRECT is a technique for finding the global minimum of a multivariate function subject to linear bounds.⁷ It is a region-based Lipschitzian approach,⁷ in that finite bounds on the rate of change of the objective function can be determined. This bound, known as the Lipschitz constant, is used along with function value information at the upper and lower limits in each dimension of a region to compute a bound on the objective function in that region.¹⁶ DIRECT balances global search, which finds the “basin of attraction” of the minimum, with local search, which exploits this region. The search space is treated as a normalized n -D hyper-rectangle, which is recursively divided into smaller hyper-rectangles. Hyper-rectangles at level $l \geq 0$ have long sides of length 3^{-l} and short sides of length at least $3^{-(l+1)}$. Starting with the center \mathbf{x} and level $l = 0$ of the initial rectangle, the $2n$ points $\mathbf{x} \pm 3^{-(l+1)}\mathbf{e}_i$, $i = 1, \dots, n$, are evaluated. The rectangle is divided into thirds, first along the dimension with the smallest function value, and these divisions are then divided along the dimension with the next smallest value, and continuing until division has been performed along all dimensions. The rectangles are then grouped by their l_2 -norm diameters, $d(l, p) = 3^l(n - 8p/9)^{1/2}$, 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 (see Figure 1(a)), are identified, and are used as the centers for the next iteration. The slopes of the line segments connecting these points are estimates of the Lipschitz constant. DIRECT uses these estimates implicitly, by providing a method to select hyper-rectangles for division,¹⁶ and therefore DIRECT is considered as “Lipschitz optimization without the Lipschitz constant”.⁷ An example of rectangle division in the 2D case for the first two DIRECT iterations is illustrated in Figures 1(b-c). The basic DIRECT algorithm is outlined below.

1. Normalize the search space to be the unit n -dimensional hypercube. Let x_0 denote the center of the hypercube. Evaluate $f(x_0)$.
2. Identify the set S of potentially optimal rectangles. These are the n -D rectangles that define the bottom of the convex hull of a scatter plot of rectangle diameters versus $f(x_i)$ for all rectangle centers x_i .
3. For all rectangles $j \in S$:
 - (a) Identify the set T of dimensions with maximum side length 3^{-l} . Let $\delta = 3^{-(l+1)}$, or one-third the maximum side length.
 - (b) Sample the function at points $\mathbf{x} \pm \delta\mathbf{e}_i$ for all $i \in T$, where \mathbf{x} is a rectangle center and \mathbf{e}_i is the i th unit vector.
 - (c) Divide the rectangle containing x into thirds along the dimensions in T , starting with the dimension with the lowest value of $f(\mathbf{x} \pm \mathbf{e}_i)$ and continuing to the dimension with the highest $f(\mathbf{x} \pm \mathbf{e}_i)$.
4. Repeat steps 2-3 until convergence or termination criteria are met.

At each iteration, the maximum number of function evaluations is $2n$ times the number of points on the convex hull, representing the centers of hyper-rectangles that are candidates for division. The function values at the the centers of the resulting divided hyperrectangles can be computed in parallel. The theoretical behavior of DIRECT has been described in the literature.^{7,16,17} and the algorithm has been adapted to many engineering problems.¹⁷ Full descriptions of DIRECT are found in.^{7,16-18}

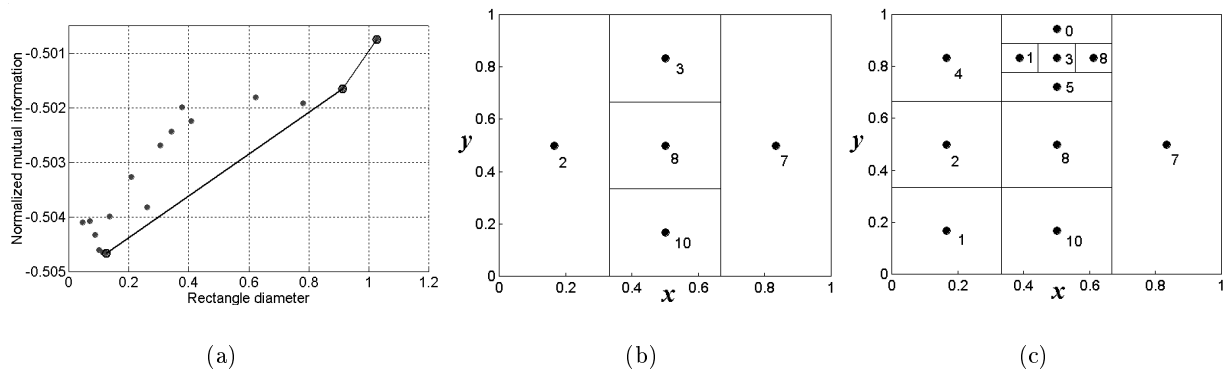


Figure 1. Convex hull. The rectangles whose centers have function values that are located on the solid black line are identified as potentially optimal, and are selected for subdivision in the next DIRECT iteration.

2.2. Generalized Pattern Search

The generalized pattern search is an iterative method for unconstrained minimization in which the search space around a base point $\mathbf{x} \in \mathbf{R}^n$ is explored until a new, better iterate is found.⁸ Points are sampled on a deterministic mesh, or pattern constructed around the current iterate at iteration k , \mathbf{x}_k (Figure 2(a)). GPS provides flexibility in how the mesh points are sampled. At one extreme, the algorithm may “guess” which direction would produce a simple decrease in function value. At the other extreme, all points on the mesh can be sampled, and the point with the largest decrease is chosen as \mathbf{x}_{k+1} for the next iteration. In the following discussion, let \mathbf{Z} and \mathbf{R} respectively denote the set of integers and real numbers. Let $x_k \in \mathbf{R}^n$ be a point at iteration k around which a pattern is constructed. Let $B \in \mathbf{R}^{n \times n}$ denote a nonsingular basis matrix, and let $C_k \in \mathbf{Z}^{n \times p}$, $p > 2n$, denote a generating matrix $C_k = [M_k - M_k L_k]$, where M_k is a finite set of nonsingular matrices from $\mathbf{Z}^{n \times n}$, and $L_k \in \mathbf{Z}^{n \times (p-2n)}$, which contains at least one column of zeros. Let $\Delta_k \in \mathbf{R}$, $\Delta_k > 0$ denote the step size of the exploratory moves. The direction of the exploratory moves is defined by a pattern $P_k = BC_k$. A trial step s_k^i is the i -th column of the matrix $\Delta_k P_k$, and $\mathbf{x}_k + s_k^i$ is a trial point. The GPS algorithm is outlined as follows⁸:

1. Initialization
 - (a) Set $k \leftarrow 0$.
 - (b) Choose $\Delta_0 > 0$.
 - (c) Select B and C_0 .
 - (d) Set $\tau > 1$.
2. Evaluate $f(\mathbf{x}_0)$.
3. Determine the step sizes s_k^i from $\Delta_k BC_k = \Delta_k P_k$.
4. Evaluate $f(\mathbf{x}_k) + s_k^i$ for $i = 1, \dots, p$.
5. If there is an s_k^i such that $f(\mathbf{x}_k + s_k^i) < f(\mathbf{x}_k)$ (i.e. there is an improving point), then:
 - (a) $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + s_k^i$.
 - (b) $\Delta_k = \lambda_k \Delta_k$, where $\lambda_k \in \{\tau^{w_1}, \dots, \tau^{w_L}\}$, and $w_i \geq 0$ for $i = 1, \dots, L$.
6. Otherwise (there is no improving point):
 - (a) $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k$.

(b) $\Delta_k = \theta \Delta_k$, where $\theta \in \{\tau^{w_1}, \dots, \tau^{w_L}\}$, and $w_0 < 0$.

7. Set $k \leftarrow k + 1$.

8. Repeat until convergence criteria are reached.

In the simplest implementations, B and M_k are set to $I^{n \times n}$, and L_k is a column of zeros. Specific cases of GPS are derived by the method in which pattern points are sampled for evaluation. If all p points (or a large subset) of the pattern P_k are to be evaluated, then parallel evaluation of $f(\mathbf{x}_k + s_k^i)$ can substantially increase efficiency. The convergence properties of GPS have also been studied.⁸

2.3. Multidirectional Search (MDS)

The multidirectional search is a special case of the GPS. Like the Nelder-Mead method, MDS utilizes a simplex consisting of $n + 1$ n -D vertices. A new simplex is generated at each iteration based on the current best point \mathbf{v}_0 , i.e., the point attaining the lowest function value in the simplex. Given an n -dimensional vertex \mathbf{v}_0 , the initial simplex consisting of vertices v_i , $i = 1, \dots, n$, is constructed so that the vertices fully span the n -dimensional space. One such construction is given as¹⁹:

$$\mathbf{v}_i = (v_0 + p_0, v_1 + p_1, \dots, v_{n-1} + p_{n-1})^T, \quad (1)$$

where

$$p_j = \begin{cases} \frac{\sqrt{n+1-1+n}}{n\sqrt{2}} S & \text{if } j = i - 1 \\ \frac{\sqrt{n+1-1}}{n\sqrt{2}} S & \text{otherwise} \end{cases} \quad (2)$$

and S is a scaling factor. There can be different S_i for each dimension i , based on the relative scaling in that dimension. In linear registration, translation and rotation have different characteristic scales.

The simplex changes by reflection, expansion, or contraction, as shown in Figure 2(b). In each iteration, the simplex is reflected and the new vertices are evaluated. If a new best vertex has been identified, an expansion step is computed. Otherwise, a contraction step is performed, and the new vertices are accepted. MDS has two primary loops: an outer loop that determines a new set of search directions by considering the best vertex, and an inner loop that determines the length of the steps to be taken.⁹ These step lengths are determined by an expansion factor $\mu > 1$ and a contraction factor $\theta \in (0, 1)$. MDS is a descent method, since for each iteration k , $f(\mathbf{v}_0^k) \leq f(\mathbf{v}_0^{k+1})$. Furthermore, the simplex angles do not change, and thus, unlike the Nelder-Mead simplex, the search directions are guaranteed to be linearly dependent. Under certain general conditions, MDS has been proven to converge. A full description of the MDS algorithm and its convergence properties is found in.⁹

MDS parallelizes in a natural way. The initial simplex requires $n + 1$ function evaluations, and reflections, contractions, and expansion each require n evaluations, all of which can be performed in simultaneously by multiple CPUs. The basic schema of MDS is given below.

1. Start with an initial simplex S_0 with $n + 1$ vertices, $\langle \mathbf{x}_0^0, \mathbf{x}_1^0, \dots, \mathbf{x}_n^0 \rangle$, and select two rational numbers: an expansion factor $\mu \in (1, +\infty)$, and a contraction factor $\theta \in (0, 1)$.
2. Compute $f(\mathbf{x}_n^0)$ for $i = 0, \dots, n$.
3. Initialize the iteration k : $k \leftarrow 0$.
4. Find the best vertex, j , where $j = \operatorname{argmin}_i \{f(\mathbf{x}_i^k) : i = 0, \dots, n\}$.
5. Swap \mathbf{x}_j^k and \mathbf{x}_0^k .
6. Repeat until a vertex has been replaced.

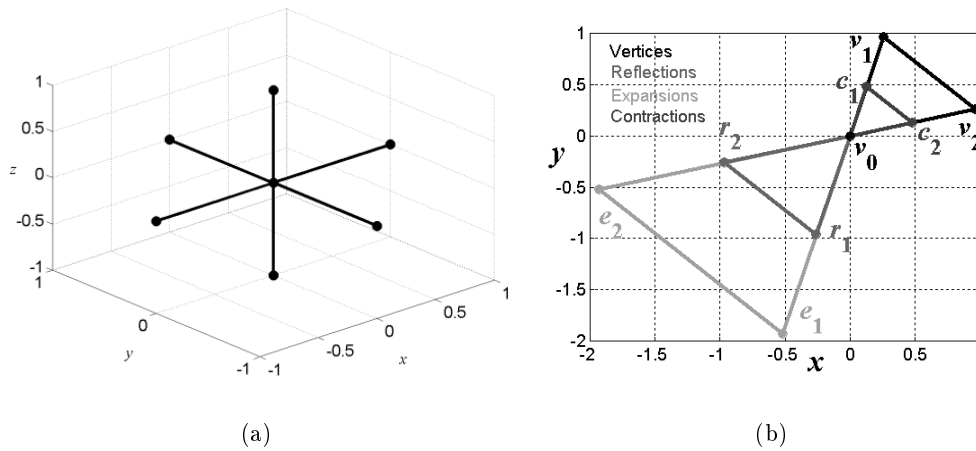


Figure 2. (a) Example 3D pattern in GPS. (b) Simplex transformations in MDS.

- (a) For all i , compute reflections, $\mathbf{r}_i^k = \mathbf{x}_0^k - (\mathbf{x}_i^k - \mathbf{x}_0^k)$, and compute $f(\mathbf{r}_i^k)$.
 - (b) If $\min \{f(\mathbf{r}_i^k) : i = 1, \dots, n\} < f(\mathbf{x}_0^k)$, then $\text{replaced} \leftarrow 1$.
 - (c) If replaced , then compute an expansion:
 - i. For all i , compute expansions: $\mathbf{e}_i^k = \mathbf{x}_0^k - \mu(\mathbf{x}_i^k - \mathbf{x}_0^k)$, and compute $f(\mathbf{e}_i^k)$.
 - ii. If $\min \{f(\mathbf{e}_i^k)\} < f(\mathbf{r}_i^k)$, then accept the expansion: $\mathbf{x}_i^k \leftarrow \mathbf{e}_i^k$, else, accept the reflection: $\mathbf{x}_i^k \leftarrow \mathbf{r}_i^k$.
 - (d) Else (if *not replaced*) then compute contractions:
 - i. $\mathbf{c}_i^k = \mathbf{x}_0^k + \theta(\mathbf{x}_i^k - \mathbf{x}_0^k)$, and compute $f(\mathbf{c}_i^k)$.
 - ii. If $\min \{f(\mathbf{c}_i^k) : i = 1, \dots, n\} < f(\mathbf{x}_0^k)$, then $\text{replaced} \leftarrow 1$.
 - iii. Accept the contraction: $\mathbf{x}_i^k \leftarrow \mathbf{c}_i^k$.
7. Repeat until replaced .
 8. $k \leftarrow k + 1$.
 9. Repeat steps 4–6 until convergence criteria are achieved.

3. METHODS

To demonstrate the use of DIRECT and MDS in image registration, and to compare accuracy and timing results with Powell's method, rigid body ($n = 6$, with translations and rotations about the x -, y -, and z - axes) registration experiments were performed on 3D images where ground truth values were known. The data were simulated T1 and PD BrainWeb brain volumes (www.bic.mni.mcgill.ca/brainweb),²⁰ shown in Figure 3. The source volume is a simulated proton-density (PD) 3D MRI volume of a brain with multiple sclerosis (MS) lesions (20% intensity non-uniformity and 3% noise relative to intensity). The volume contains $101 \times 101 \times 35$ voxels (volume elements, or 3D pixels), each of size 1.0 mm^3 . The target volumes (to which the source is registered) are simulated T1 MRI volumes of a normal brain, containing $181 \times 217 \times 181$ 1.0 mm^3 voxels with 0% intensity non-uniformity. The first volume contained 0% noise, and the second contained 9% noise. The BrainWeb data were selected to test robustness with respect to differences in morphology (normal brain tissue vs. multiple sclerosis regions) and noise.

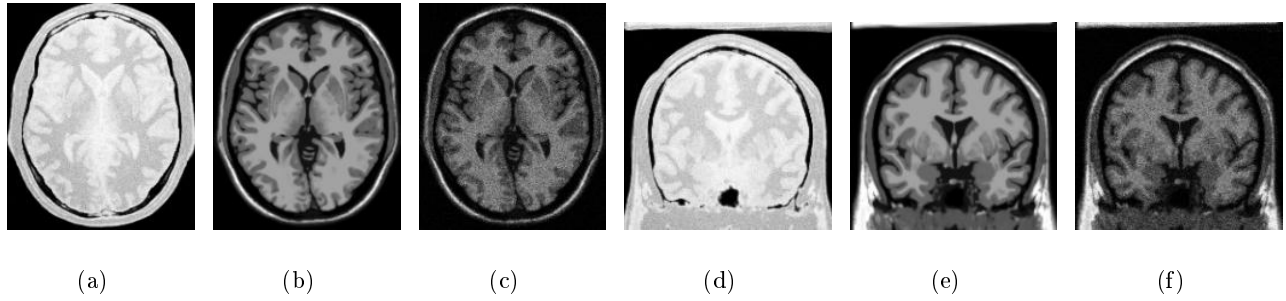


Figure 3. Images used in brain registration experiments. (a-c) Axial view of (a) the source volume - PD MRI brain with MS lesions, 3% noise (b) Target volume - T1 MRI brain, (c) 9% noise. (d-f) Coronal view of the same volumes.

Two-level multiresolution registration of 3D to 3D medical images were performed with: (1) DIRECT in the low resolution stage, followed by high-resolution refinement with GPS (DIRECT/GPS); (2) DIRECT in the low resolution stage, followed by high-resolution refinement with MDS (DIRECT/MDS); and (3) Powell's method.

The most popular and widely-studied similarity metrics for intensity-based registration are those based on information theory.^{1,2} These measures have been used for multimodality registration, and for aligning functional modalities (e.g. PET) with structural modalities (e.g. MRI, CT). Let X and Y respectively denote the intensities of the source image and of the corresponding transformed coordinates in the target (reference) volume. Let $p(x_i)$, $i = 1, \dots, n$, and $p(y_j)$, $j = 1, \dots, m$, denote the probability distribution of the intensities of X and Y , respectively, and $p(x_i, y_j)$ represent their joint distribution. The Shannon entropy of X is $H(X) = -\sum_i^n p(x_i) \ln p(x_i)$ and the joint entropy of X and Y , $H(X, Y)$, is $H(X, Y) = -\sum_j^m \sum_i^n p(x_i, y_j) \ln p(x_i, y_j)$. The Shannon mutual information (MI) of X and Y is defined as:

$$MI \equiv I(X, Y) = H(X) + H(Y) - H(X, Y). \quad (3)$$

Mutual information normalized by the joint entropy, or simply normalized mutual information (NMI) is less sensitive to the size of the source/target overlap, and is defined as:

$$NMI \equiv \tilde{I}(X, Y) = \frac{H(X) + H(Y)}{2H(X, Y)}. \quad (4)$$

During the low resolution phase of the registration, the number of voxels used to compute the similarity metric was half the full the number of voxels in the high resolution, local refinement phase. It has been found from previous studies^{1,11,15} that simple downsampling, as opposed to filtering or blurring, is efficient, and just as accurate. During the registration, both linear (LI) and partial volume interpolation (PV), were used to generate joint histograms, from which the probability density was estimated. In 3D linear (trilinear) interpolation, a new intensity is computed as a linear combination of a voxel's eight nearest neighbors, and the corresponding histogram bin is incremented. In partial volume interpolation, the bins corresponding to the eight neighbors is updated according to the weights of the linear combination.² We have observed from previous work^{1,11} that the best results were obtained with 64-bin histograms, and consequently 64-bins were also used in the current work. To test the scaling of the optimization methods, all experiments were performed with sets of 1, 2, 4, and 8 CPUs.

The source volumes were initially mistranslated at distances (d_0) of 5mm, 10mm, and 20mm with respect to the target, and at initial misrotations of $\pm 30^\circ$ about all three axes, resulting in 8 misrotations. For each d_0 , 4 initial translation vectors were computed.

The experiments were performed on two architectures: a large SMP computer and a dual-processor system with multithreading capabilities. The SMP system was an SGI Altix 3000 shared memory system running Linux, with 20 Intel Itanium 2 CPUs, each running at a clock speed of 1.3 GHz. The system 20 GB main memory and a 6 MB L3 floating point precision cache. The programs were written in C and C++ and compiled with the Intel compiler, with parallelism implemented through OpenMP.

To test parallel performance on a smaller system, registrations were also performed on a dual-processor Intel Xeon system, with each processor running at 3.06 GHz. This system employs “hyperthreading”, which was previously known as “simultaneous multithreading”. Each processor appears to the operating system as two logical processors, so that the system is treated as a 4-CPU SMP computer. Instead of each execution thread running on a single CPU, as is the case with larger SMP systems, parallel processing is performed as two execution threads running on a single CPU. Small SMP computers with 2 or 4 processors are becoming increasingly popular, and therefore a performance analysis of the parallel optimization methods are presented.

DIRECT, GPS, and MDS were implemented as described in the literature.⁷⁻⁹ Powell’s method with Brent’s line minimization,²¹ and optimization along the 6 dimensions were performed in the same order as described in.⁶ With the current state of the art in image acquisition and visualization systems, it is reasonable to assume that in real linear registration situations, the initial orientation would 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 coordinate axes, and $\pm \pi$ radians for the initial rotations about the three axes. For GPS and MDS, the expansion and contraction factors were respectively set to $\mu = 2$ and $\theta = \frac{1}{2}$. The translation scale S was set to 2 mm, and the rotation scale was set to 8° , but the algorithm was found to be robust with respect to different scale values.

For Powell’s method, the termination condition was $2|f_p - f| \leq 0.005(|f_p| + |f|)$,²¹ or if the number of iterations in Brent’s line minimization exceeded 500. f and f_p respectively denote the current function value and the best value from the previous line search. For DIRECT, convergence criteria were the same, with f_p denoting the best value from the division of all rectangles in the previous iteration. Additionally, to ensure adequate search space exploration, iterations (maximum of 200) continued until the volume of the smallest hyperrectangle reached 0.1mm^3 for translations and 0.004rad^3 for rotations. For GPS and MDS, the convergence criteria were the same as for Powell’s method, with an additional condition based on the maximum distance between vertices of the MDS simplex¹⁹:

$$\frac{1}{\max\{1, \|\mathbf{v}_0\|\}} \max_{1 \leq i \leq n} \|\mathbf{v}_i^k - \mathbf{v}_0\| < 0.5, \quad (5)$$

Performance was judged on: (1) The ratio of satisfactory registrations to the total number of registration experiments performed, where a registration 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 $\pm 1^\circ$; (2) Computation time. Since in the case of image registration the parameter values determined by the optimization (and not the actual numerical objective function value) have primary importance, computation times are only considered for successful registrations.

4. RESULTS

The success rates for the brain registration experiments using DIRECT/GPS and DIRECT/MDS are shown in Table 1. Timing results for the 20-CPU SMP experiments with the 0% noise and 9% noise source volumes are found in Figures 4 and 5, respectively. Because Powell’s method is serial, only 1-CPU timing results are presented. They are shown alongside multiple-CPU results for comparison.

Computation times for the dual-processor system are shown in Figure 6. Only the DIRECT/MDS results using NMI and MI with PV interpolation are presented, as the timing results with linear interpolation are similar.

The DIRECT methods had substantially higher success rates than Powell’s method, as seen in Table 1. This result is expected, due to the global scope of DIRECT. For small initial misorientations (e.g. $< 20^\circ$ rotation error and < 20 mm translation error), Powell’s method is usually the best choice for accurate and efficient derivative-free local optimization. However, it fails for the large misrotations used in the experiments presented in this paper

Table 1. Brain registration success rates.

Opt. method		DIRECT/GPS				DIRECT/MDS				Powell			
Metric		NMI		MI		NMI		MI		NMI		MI	
Source	d_0	L	PV	L	PV	L	PV	L	PV	L	PV	L	PV
0% noise	5	1.00	0.97	0.97	1.00	1.00	1.00	0.97	1.00	0.16	0.25	0.25	0.31
	10	0.75	0.63	0.81	0.69	0.75	0.84	0.84	0.84	0.00	0.06	0.16	0.22
	20	0.69	0.56	0.63	0.53	0.88	0.88	0.66	0.72	0.00	0.00	0.06	0.06
9% noise	5	0.97	0.97	0.97	1.00	1.00	1.00	0.97	1.00	0.16	0.09	0.16	0.22
	10	0.66	0.63	0.66	0.59	0.69	0.78	0.63	0.63	0.00	0.03	0.09	0.03
	20	0.44	0.47	0.53	0.50	0.53	0.84	0.56	0.59	0.00	0.00	0.00	0.06

(30°), as does MDS when used without DIRECT.¹¹ Local refinement with MDS resulted in marginally higher success rates than with GPS local optimization. For all methods, the number of correct registrations decreased as d_0 increased. The success rates for the registrations with the 0% noise source volume were marginally higher than for the 9% noise volume, but the degree of noise was not a major factor in registration performance. There was no significant difference in the success rate for LI or PV, or between NMI and MI, for the DIRECT methods or for Powell’s technique.

From Figures 4 and 5, efficiency for the parallel methods improve with the number of processors used. Local refinement with GPS was more efficient and consistent (with a lower standard deviation) than with MDS. Registration with linear interpolation was generally faster than PV. One possible explanation is that because the lower degree of “arch” or “oscillation” artifacts in the mutual information objective function with PV, local minima far from the global minimum are more pronounced, consequently causing the global part of the search (DIRECT) to expend more function evaluations exploring local optima. The effect of interpolation on registration performance appears to be heavily dependent on the characteristics of the source and target volumes, and research into these effects continues.²²

5. DISCUSSION

In clinical registration applications, the initial orientation is often close to optimal alignment, and the purpose of registration – especially rigid or linear – is to “fine-tune” this initial guess. In these cases, the objective function in the vicinity the minimum is generally smooth with few local optima, and local techniques such as Powell’s method, gradient-descent, and even MDS are very fast and accurate. The main advantage of DIRECT is that its scope is global, and it is best used when there is uncertainty regarding the initial guess.

Because DIRECT requires a large number of objective function evaluations, it is best utilized in coarse grained parallel implementations. When executed serially, DIRECT often requires more computation time than Powell’s method. For a sufficient number of CPUs (4 or more in for the experiments presented in this paper), DIRECT often reaches the vicinity of the minimum faster than does Powell’s method. Once this region has been identified, a local technique can quickly converge to the minimum. Powell’s method or gradient-descent techniques may serve this purpose, but we elected to employ GPS and MDS, which, like DIRECT, are conducive to coarse grained parallelization. Although GPS is generally more efficient than MDS, the latter consistently has a higher success rate, and, at present, is recommended over GPS as a local refinement mechanism.

On the 20-CPU system, computation time decreased almost linearly using 2 and 4 CPUs for the DIRECT methods. The rate of decrease (scaling) was lower with additional processors, but computation time continued to improve. On the dual-processor system, efficiency of MI registration improved substantially with 2 threads, and improved marginally with 3 and 4 threads. This result is expected, as as true multiprocessing is possible with two threads, which utilize the full resources of two separate CPUs. In contrast, with hyperthreading, CPU resources are shared among multiple threads.

For NMI registration, a very sharp decrease in computation time (super-linear scaling) occurred with 2 CPUs, probably due to caching effects and other memory optimizations specific to the Intel processor. With additional

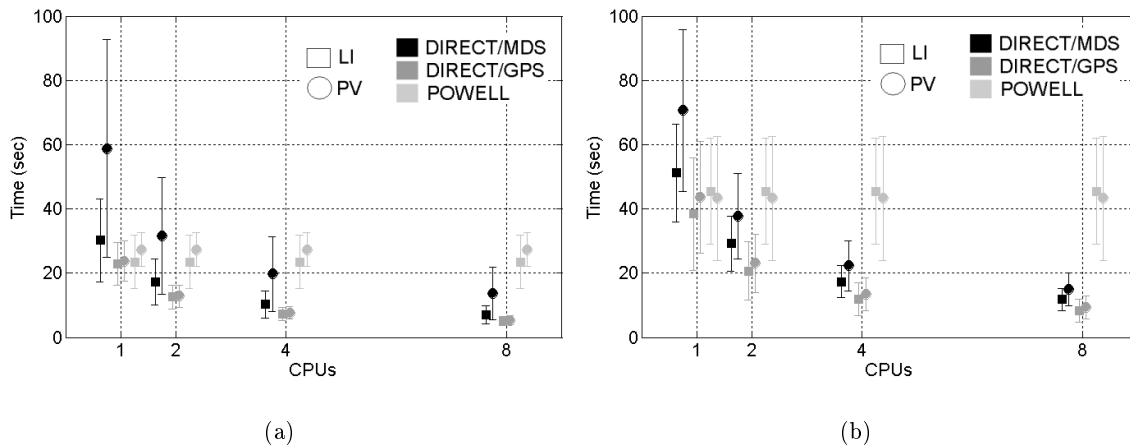


Figure 4. Computation time for 0% noise source volume. (a) NMI. (b) MI.

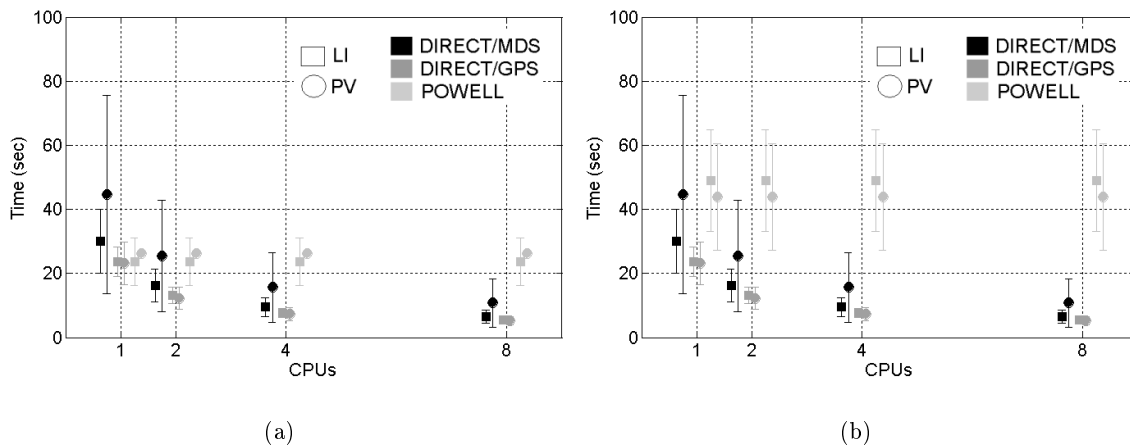


Figure 5. Computation time for 9% noise source volume. (a) NMI. (b) MI.

CPUs, performance actually was worse than with 2 CPUs, but still better than with a single processor. This observation could also be explained by the smoothness of NMI near the minimum. Furthermore, in previous work^{1,11} we found that NMI is often easier to optimize than MI. Performance and accuracy issues regarding NMI and MI are still being studied.²

6. CONCLUSION

In this paper, coarse grained parallel optimization approaches for both global searches and for local refinement were presented for medical image registration. Experimental results were presented for a large shared memory system as well as a dual-processor architecture. This work shows that parallel optimization on shared memory systems can markedly decrease computation time during registration. Compared with Powell's method, combining global (DIRECT) and local (GPS and especially MDS) parallel derivative-free optimization yields substantial improvements in efficiency with as few as 4 CPUs, or with 2 CPUs on dual-processor computers. The coarse granularity of DIRECT, the generalized pattern search, and the multidirectional search can be implemented on both distributed and shared memory architectures. However, on the latter, fine grained registration operations, such

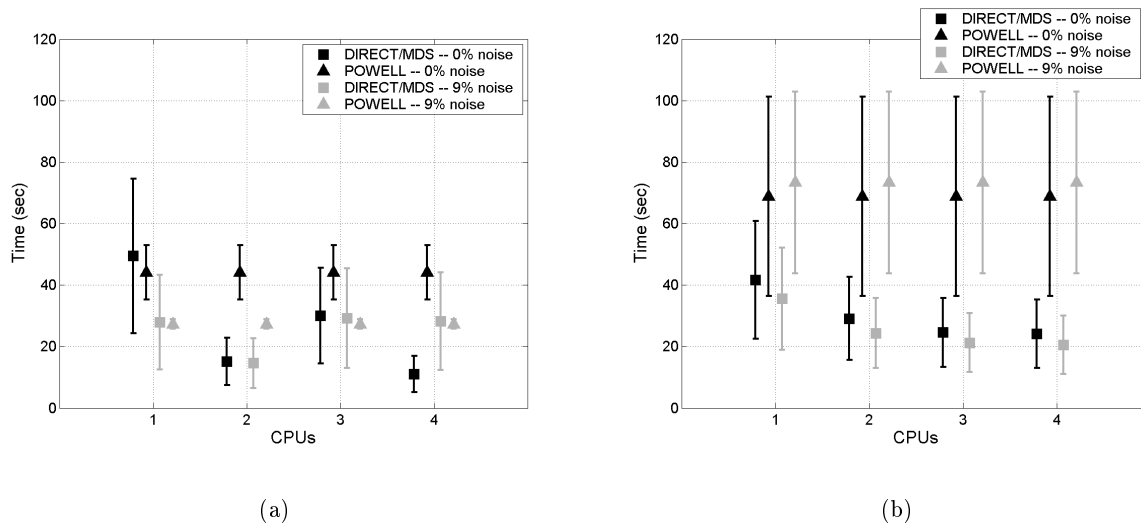


Figure 6. Computation time with hyperthreading. (a) NMI. (b) MI.

as interpolation and resampling, can also be parallelized. Because of the robustness and efficiency of inherently parallel methods presented here, the application of global, and especially parallel optimization, to medical image registration merits further investigation. In future work, mixed granularity optimization for registration will be implemented, as well as combining DIRECT with Powell’s method. Additionally, the optimization approaches presented in this paper will be adapted for non-rigid registration, and validated on clinical images.

7. ACKNOWLEDGMENTS

The authors thank Kevin Wang, Chris Wedlake, and Dr. Baolai Ge for technical assistance, the Montréal Neurological Institute at McGill University for making their BrainWeb images available, and Dr. Renata Smolíková-Wachowiak, Dr. Gerard Guiraudon, Dr. Hualiang Zhong, and John Moore for valuable discussions. This work was supported by SHARCNet, NSERC R3146-A02, CIHR 14735, and ORDCF.

REFERENCES

1. M. P. Wachowiak, R. Smolíková, and T. M. Peters, “Multiresolution biomedical image registration using generalized information measures,” in *Proceedings of MICCAI 2003*, R. E. Ellis and T. M. Peters, eds., *Lecture Notes in Computer Science* **2879**, pp. 846–853, Springer Verlag, 2003.
2. J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, “Mutual-information-based registration of medical images: a survey,” *IEEE Transactions on Medical Imaging* **22**, pp. 986–1004, 2003.
3. T. Makela, P. Clarysse, O. Sipila, N. Pauna, Q. C. Pham, T. Katila, and I. E. Magnin, “A review of cardiac image registration methods,” *IEEE Transactions on Medical Imaging* **21**, pp. 1011–1021, 2002.
4. R. Smolíková, M. P. Wachowiak, and M. Drangova, “Registration of fast cine cardiac MR slices to 3D preprocedural images: toward real time registration for mri-guided procedures,” in *Proceedings of SPIE: Medical Imaging*, J. M. Fitzpatrick and M. Sonka, eds., **5370**, pp. 1195–1205, 2004.
5. 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,” *Medical Image Analysis* **8**, pp. 387–401, 2004.
6. F. Maes, D. Vandermeulen, and P. Suetens, “Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information,” *Medical Image Analysis* **3**, pp. 373–386, 1999.

7. D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the Lipschitz constant," *Journal of Optimization Theory and Applications* **79**, pp. 157–181, 1993.
8. V. Torczon, "On the convergence of pattern search algorithms," *SIAM Journal on Optimization* **7**, pp. 1–25, 1997.
9. V. Torczon, "On the convergence of the multidirectional search algorithm," *SIAM Journal on Optimization* **1**, pp. 123–145, 1991.
10. R. He and P. A. Narayana, "Global optimization of mutual information: application to three-dimensional retrospective registration of magnetic resonance images," *Computerized Medical Imaging and Graphics* **26**, pp. 277–292, 2002.
11. M. P. Wachowiak, , and T. M. Peters, "Parallel optimization approaches for medical image registration," in *Proceedings of MICCAI 2004*, C. Barillot, D. R. Haynor, and P. Hellier, eds., *Lecture Notes in Computer Science* **3216**, pp. 781–788, Springer Verlag, 2004.
12. S. K. Warfield, F. A. Jolesz, and R. Kikinis, "A high performance computing approach to the registration of medical imaging data," *Parallel Computing* **24**, pp. 1345–1368, 1998.
13. M. Jenkinson and S. Smith, "A global optimization method for robust affine registration of brain images," *Medical Image Analysis* **5**, pp. 143–156, 2001.
14. G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Transactions on Information Technology in Biomedicine* **3**, pp. 47–60, 1999.
15. 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 Transactions on Evolutionary Computation* **8**, pp. 289–301, 2004.
16. C. T. Kelley, *Iterative methods for optimization*, SIAM Press, 1999.
17. J. M. Gablonsky and C. T. Kelley, "A locally-biased form of the direct algorithm," *Journal of Global Optimization* **21**, pp. 27–37, 2001.
18. L. T. Watson and C. A. Baker, "A fully-distributed parallel global search algorithm," *Engineering Computation* **18**, pp. 155–169, 2001.
19. V. Torczon, *Multi-directional search: A direct search algorithm for parallel machines*. PhD thesis, Rice University, May 1989.
20. R. K. S. Kwan, A. C. Evans, and G. B. Pike, "MRI simulation-based evaluation of image processing and classification methods," *IEEE Transactions on Medical Imaging* **18**, pp. 1085–1097, 1999.
21. W. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C. The Art of Scientific Computing*, Cambridge University Press, 2 ed., 1992.
22. J. Tsao, "Interpolation artifacts in multimodality image registration based on maximization of mutual information," *IEEE Transactions on Medical Imaging* **22**(7), pp. 854–864, 2003.