

Parallel Optimization Approaches for Medical Image Registration

Mark P. Wachowiak¹ and Terry M. Peters^{1,2}

¹ Imaging Research Laboratories, Robarts Research Institute

² Department of Medical Biophysics, University of Western Ontario

London, ON N6A 5K8 Canada

{mwach, tpeters}@imaging.robarts.ca

Abstract. Optimization of a similarity metric is an essential component in most medical image registration approaches based on image intensities. The increasing availability of parallel computers makes parallelizing some registration tasks an attractive option. In this paper, two relatively new, deterministic, direct optimization algorithms are parallelized for distributed memory systems, and adapted for image registration. DIRECT is a global technique, and the multidirectional search is a recent local method. The performance of several variants are compared. Experimental results show that both methods are robust, accurate, and, in parallel implementations, can significantly reduce computation time.

1 Introduction

The robustness and performance of intensity-based image registration techniques depend on selecting an appropriate similarity metric suited to the particular application, choice of search space (linear, elastic, nonlinear), interpolation scheme, and the approach used to optimize the similarity metric. For the latter, the Nelder-Mead downhill simplex, Powell's method (line search), and methods requiring derivatives (conjugate gradient, Newton's method, Levenberg-Marquardt) have often been used [1,2]. If accurate first derivatives are available, then gradient descent methods are preferred [3]. However, these techniques are generally local, and are susceptible to premature convergence to local optima, and incorrect registration. Global optimization, usually having a stochastic component (e.g. [4,5,6]), has been shown to improve robustness, but at the cost of increased similarity metric evaluations and slower convergence. Faster processors and memory access have substantially reduced registration time, but parallel computing has the potential to further increase efficiency, and also to facilitate use of optimization techniques that were formerly considered too computationally expensive to be used for registration. Powell's method, while generally robust, is still prone to local minima entrapment for many similarity metrics [7], and is not easily parallelized. In gradient descent methods, gradients can be computed in parallel, but analytical expressions for the derivatives of many metrics (w.r.t. a transformation) are not available, or cannot be easily estimated.

In the current work, two relatively new deterministic direct methods (not requiring evaluation of derivatives) are applied to medical image registration: DIRECT (for DIviding RECTangles) is a global technique that was designed for difficult optimization problems [8], and the multidirectional search (MDS) is a local method [3,9]. The inherent parallelism of both DIRECT and MDS can be easily exploited for improving registration performance. Other investigators have used DIRECT as a final, local step after stochastic optimization for registration [4], but in this paper, DIRECT is employed as a global strategy. In addition, a version of DIRECT with local bias is adapted for registration. To our knowledge, MDS has not previously been used in this application.

Other investigators have exploited parallelism to improve registration speed for clinical applications. In [10], resampling is multithreaded, and computation of similarity values, as well as segmentation and visualization, are parallelized on distributed memory clusters. In [11], vector field computations and resampling are performed in a hybrid distributed/shared memory environment. In the current paper, coarse-grained (high computation-to-communication ratio) parallelism is achieved by distributing the entire evaluation of the similarity metric, so that metrics for many transformations can be computed simultaneously. This technique is suitable for distributed, shared memory, and hybrid implementation, and is also extendible to include parallelization of other operations.

2 Dividing Rectangles (DIRECT)

DIRECT is a relatively recent algorithm for finding the global minimum of a multivariate function subject to linear bounds. It is essentially a Lipschitzian approach, but no Lipschitz constant needs to be specified [8]. DIRECT balances global search, which finds the “basin of attraction” of the minimum, with local search, which exploits this basin. In DIRECT, the search space is treated as an n -D rectangle with sides normalized to length 1, which is then recursively divided into smaller rectangles. Every rectangle i is centered at a point \mathbf{x}_i , and has a level $l = 0, 1, \dots$ and stage p , $0 \leq p < n$. The rectangle has p short sides of length $3^{-(l+1)}$ and $n-p$ long sides of length 3^{-l} . Rectangles are grouped by their l_2 norm diameters, $d(l, p) = 3^l(n - 8p/9)^{1/2}$ [12]. The set of potentially optimal rectangles, or those that define the convex hull of a scatter plot of rectangle diameters versus $f(\mathbf{x}_i)$ for all rectangle centers \mathbf{x}_i , are identified, and are used as the centers for the next iteration. For each identified center, the $2n$ points $\mathbf{x}_i \pm 3^{-(l+1)}\mathbf{e}_j$, $j = 1, \dots, n$, are evaluated. The rectangle is divided into thirds, first along the dimension with the smallest function value, and the resulting rectangles are then divided along the dimension with the next smallest value, and continuing until each new point is the center of a new rectangle. Division of an initial 2D rectangle with center $(\frac{1}{2}, \frac{1}{2})$ is shown in Fig. 1(a). Here, the initial normalized 2D rectangle is divided first along the y -axis, which is the side containing the smallest function value, $f(\frac{1}{2}, \frac{1}{2} - \frac{1}{3}) = 0.4$, and next along the x -axis, which is the side having the second smallest value, $f(\frac{1}{2} - \frac{1}{3}, \frac{1}{2}) = 0.8$.

A locally-biased form of the DIRECT algorithm has been developed for low-dimensional problems with few local minima [12]. The rectangles are grouped by the l_∞ norm, or maximum side length, given by: $d(l, p) = 3^{-l}$.

At each iteration, the number of function evaluations is $2n \times$ the number of points on the convex hull, and these can be performed in parallel. Several strategies exist, from simple master-slave models to load-balancing strategies [13]. The theoretical behavior of DIRECT has been explored, [8,12], and the algorithm has been adapted to engineering problems [12], and for local optimization in image registration [4]. Full descriptions of DIRECT are found in [8,12,13].

3 Multidirectional Search (MDS)

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{x}_0 , i.e., the point attaining the lowest function value in the simplex. Initial vertices are usually chosen as $[\mathbf{I}_n - \mathbf{1}_{n \times 1}]$ [9], where \mathbf{I}_n is the $n \times n$ identity matrix, and $-\mathbf{1}_{n \times 1}$ is a column vector of n -1's. The simplex changes by reflection, expansion, or contraction, as shown in Fig. 1 (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 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 , $f(\mathbf{x}_0^k) \leq f(\mathbf{x}_0^{k+1})$. Furthermore, the simplex angles do not change, and thus, unlike the Nelder-Mead simplex, the search directions are guaranteed to be linearly independent. A full description of the MDS algorithm and its convergence properties is found in [9].

If P processors are available, then each CPU could evaluate about $3n/P$ vertices in parallel. However, if $P \geq 3n$, then each rotation, contraction, and expansion can be computed simultaneously, in a speculative manner[14].

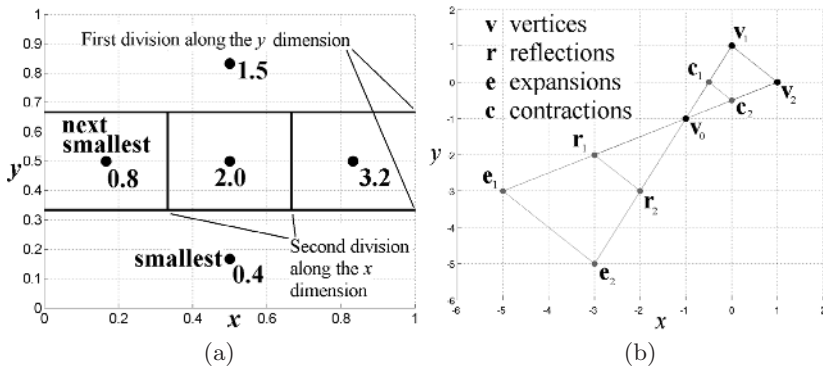


Fig. 1. (a) Division of rectangles in DIRECT. (b) MDS simplex operations.

4 Methods

Multiresolution rigid-body ($n = 6$) registration of 3D to 3D medical images was performed with global DIRECT (GD); locally-biased DIRECT (LD); GD at low resolution and LD at full resolution (GLD); MDS with $\mu = 2.0$ and $\theta = 0.5$; and Powell's (serial) method (P). Similarity metrics were mutual information (MI), normalized mutual information (NMI), entropy correlation coefficient (ECC), the correlation ratio (CR), and CR/NMI (CR at low and NMI at high resolutions). For MI, NMI, and ECC, histograms computed with trilinear partial volume interpolation [2] and 64 bins were found to give the best results.

The data were (Fig. 2): (1) Source volume: Simulated PD MRI brain volume (BrainWeb, Montreal Neurological Institute [15]) with multiple sclerosis lesions and 3% noise, with $181 \times 217 \times 35$ voxels, each voxel of size 1.0 mm^3 . Target: Simulated normal T1 MRI volume with 9% noise $181 \times 217 \times 181 \text{ mm}^3$ voxels. (2) Same as (1), but with 9% noise in the source volume. (3) Source and target: porcine heart CT volumes, $143 \times 195 \times 168$ voxels of size 0.63 mm^3 . The source and target were imaged with different CT parameters so that the target had very low contrast relative to the source. For all data, ground truth orientations were known. The source volumes were initially misregistered at random directions and with distances (d_0) of 5mm, 10mm, and 20mm from the target, and misrotated at $\pm 10^\circ$ about all three axes ($\pm 30^\circ$ for the heart volumes). For each d_0 , 24 experiments were run. Convergence criteria were (1) a maximum budget of 2000 function evaluations, or (2) a tolerance of 0.005 between the best function values, or (3) for DIRECT, there are more than five iterations where the improved \mathbf{x} is less than 0.1 mm and 0.01° from the previous best \mathbf{x} , with a tolerance of 0.005.

Performance was judged on: (1) The success rate (less than 2mm translation error and a maximum rotation error of $\pm 2^\circ$), (2) Translation and rotation accuracy for satisfactory registrations, and (3) Computation time. All experiments were run on a gigabit ethernet cluster of 900 MHz CPUs, each with 1 GB of memory, with MPI using MPICH libraries (www.mcs.anl.gov/mpi/mpich/).

5 Results

The success rates for the three volumes are shown in Table 1. The mean computation time and the ratio of computation time to number of similarity metric evaluations (T/E ratio) for the MRI registration trials are shown in Fig. 3, and in Fig. 4 for the CT heart trials. The mean translation and rotation errors, with minimum and maximum error values, are shown in Fig. 5. For the brain volumes, the DIRECT algorithms, especially LD, had the highest success rate. GD, GLD, and LD also outperformed MDS and P for $d_0 = 20\text{mm}$. MDS was successful for $d_0 = 5\text{mm}$ and 10mm , but failed for $d_0 = 20\text{mm}$. GD, GLD, LD, and P performed best for MI, NMI, ECC, and CR/NMI (especially NMI), but MDS was most successful with MI. For the CT heart volumes, MDS had a low success rate (highest with MI). GD, GLD, LD, and P were successful for $d_0 = 5\text{mm}$ and 10mm (with P the overall best for 10mm), but, as with the brain volumes, for $d_0 = 20\text{mm}$, the DIRECT methods had a high success rate while both P

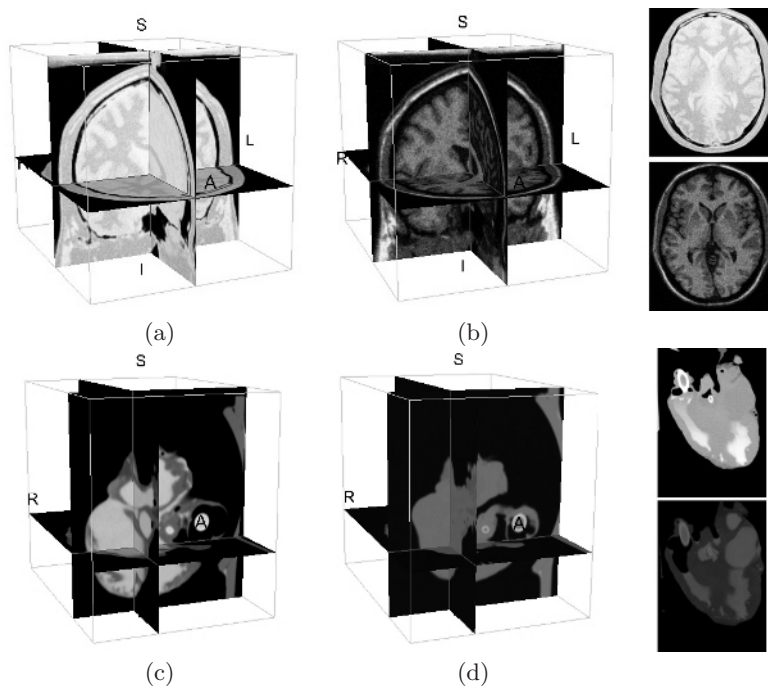


Fig. 2. (a) PD MRI brain, 3% noise. (b) T1 MRI brain, 9% noise. (c) Source CT heart volume. (d) Target CT heart volume.

and MDS performed poorly. All methods were accurate, with mean translation and rotation errors less than 1.5 mm and 2° , respectively. The parallel methods scaled well, with an almost linear decrease in time from 4 to 8 processors (DIRECT), and from 6 to 12 processors (MDS). The effect of similarity metric on timings was minimal. For “speculative” MDS, there was marginal improvement with $3n = 18$ CPUs. For the MR brain data, the mean computation time for Powell’s method was 70.81, 138.22, 116.80, 72.79, and 86.90 sec. for NMI, MI, ECC, CR, and CR/NMI, respectively, and for the CT heart data, 217.3, 338.62, 339.44, 318.00 sec. for NMI, MI, ECC, CR, and CR/NMI, respectively, which were much higher than for DIRECT and MDS (except for LD with 4 CPUs). The T/E ratio was also generally higher with P. DIRECT has high overhead (rectangle division, finding the convex hull, etc.). Thus, T/E for these methods was high for 4 CPUs. MDS has low overhead, and had the best T/E for all trials.

6 Discussion and Conclusions

In this paper, optimization with parallel similarity metric computations was presented. These methods were applied to linear registration, an important global step prior to nonlinear matching. Many nonlinear methods compute deformation vectors from small blocks of voxels at different resolutions which are matched linearly [16], and therefore can also benefit from improved optimization.

Table 1. Ratio of successful registrations. (1) Brain MRI, PD-T1; MS lesions, 3% noise. (2) Brain MRI, PD-T1, MS lesions, 9% noise. (3) Heart CT.

Volume	Metric	5 mm					10 mm					20 mm				
		P	GD	GLD	LD	MDS	P	GD	GLD	LD	MDS	P	GD	GLD	LD	MDS
(1)	NMI	1.00	1.00	1.00	1.00	0.79	0.92	0.92	0.92	1.00	0.38	0.29	0.71	0.75	0.88	0.00
	MI	1.00	1.00	1.00	1.00	1.00	1.00	0.46	0.46	0.54	1.00	0.54	0.33	0.38	0.38	0.08
	ECC	1.00	1.00	1.00	1.00	1.00	1.00	0.54	0.54	0.54	1.00	0.29	0.33	0.42	0.38	0.12
	CR	1.00	0.88	0.92	0.96	1.00	1.00	0.29	0.38	0.38	0.83	0.25	0.21	0.21	0.29	0.00
	CR/NMI	1.00	0.96	0.92	0.96	1.00	1.00	0.33	0.38	0.71	0.75	0.17	0.29	0.21	0.67	0.00
(2)	NMI	1.00	1.00	1.00	1.00	0.12	1.00	0.71	0.62	0.83	0.12	0.08	0.75	0.71	0.88	0.00
	MI	1.00	1.00	1.00	1.00	1.00	1.00	0.58	0.46	0.58	0.88	0.17	0.29	0.42	0.38	0.00
	ECC	1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.46	0.54	0.83	0.25	0.29	0.42	0.33	0.04
	CR	1.00	0.92	1.00	1.00	1.00	1.00	0.33	0.38	0.54	0.88	0.21	0.21	0.33	0.29	0.00
	CR/NMI	1.00	0.96	1.00	1.00	1.00	0.96	0.46	0.46	0.71	0.75	0.17	0.21	0.38	0.54	0.00
(3)	NMI	0.97	0.97	1.00	1.00	0.53	0.94	0.81	0.81	0.91	0.31	0.94	0.97	0.94	0.94	0.19
	MI	0.62	0.72	0.69	0.72	0.66	0.50	0.59	0.62	0.72	0.66	0.62	0.75	0.75	0.81	0.47
	CR/NMI	0.50	0.94	0.78	0.88	0.66	0.56	0.84	0.66	0.88	0.59	0.47	0.94	0.75	0.91	0.34

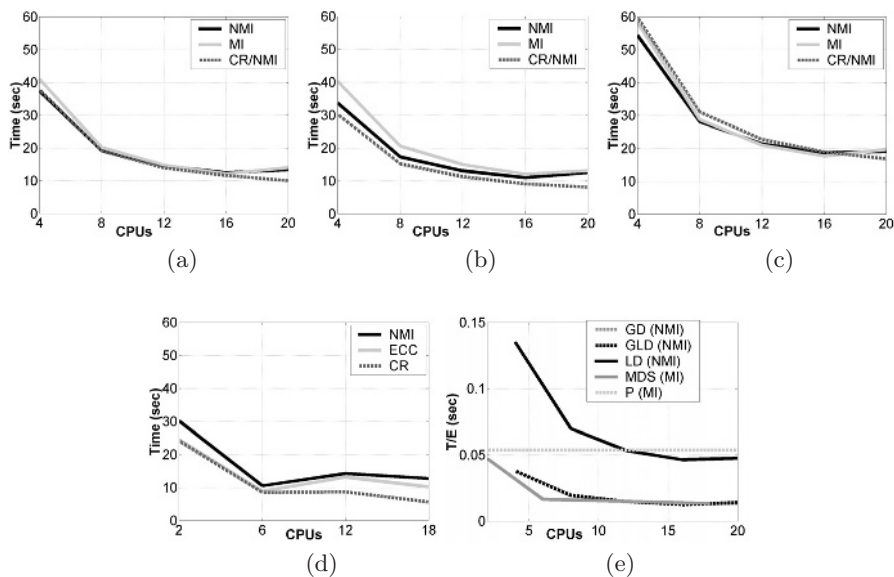


Fig. 3. Registration of brain volumes. Computation time for (a) GD (b) GLD (c) LD (d) MDS (e) Time per similarity metric evaluation.

Although Powell’s method is robust and accurate, DIRECT and MDS easily exploit parallelism. DIRECT, especially LD, is also more robust for greater initial misregistrations (20mm). As a local method, MDS works well only for small capture ranges. Thus, MDS may be used as a local, high resolution search step after global optimization with DIRECT. DIRECT and MDS were also robust with respect to noise (e.g. 9% noise in the T1 MRI volume).

Because of the coarse granularity, the parallel methods scaled well. Function computation times are mostly consistent, and load balancing [13] is not expected to greatly increase speed. At least 8 CPUs for the DIRECT methods, and at

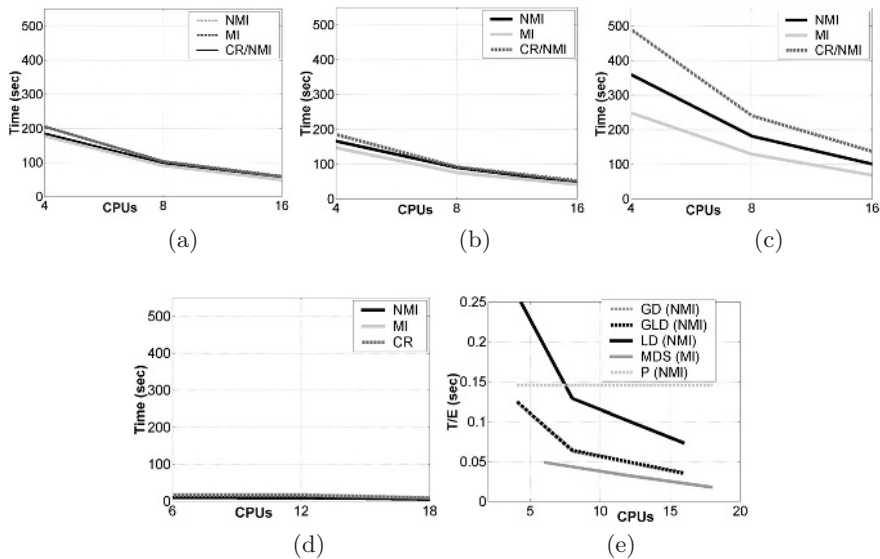


Fig. 4. Heart volume registration. Computation time for (a) GD (b) GLD (c) LD (d) MDS (efficient, but very low success rate) (e) Time per similarity metric evaluation.

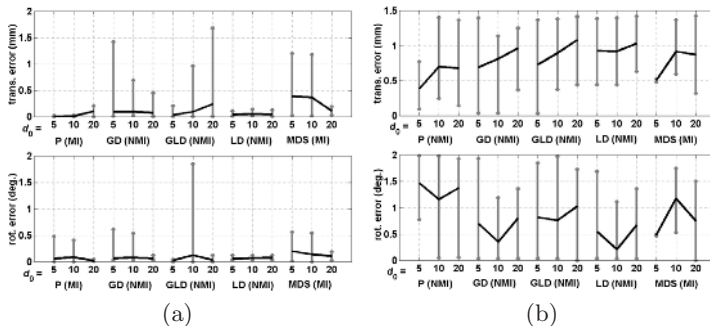


Fig. 5. Mean translation and rotation errors. (a) Brain volumes. (b) Heart volume.

least 12 for MDS, are required for speedup. Faster processors (> 2 GHz), shared memory, and optimized numerical libraries can also improve speed.

The robustness of DIRECT (and MDS for small capture ranges) and their scalability suggest that these methods should be considered as an alternative to Powell’s method when many CPUs are available. However, on single or dual processor machines, the overhead in the DIRECT methods is prohibitive, and standard approaches should be used. The success of DIRECT and MDS for image registration merits further research into other non-stochastic parallel optimization approaches [17]. Future work includes implementation of these methods on shared-memory architectures, and on parallel methods in which other time consuming steps (resampling, joint histogram estimation) can also be parallelized.

Acknowledgments. The authors thank Hua Qian for heart CT data, Baolai Ge for technical assistance, Renata Smolíková-Wachowiak for helpful discussions, and the anonymous reviewers for constructive suggestions. This work was supported by SHARCNet, NSERC R3146-A02, CIHR 14735, and ORDCF.

References

1. Maes, F., Vandermeulen, D., Suetens, P: Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Med. Image Anal.* **3** (1999) 373–386.
2. Pluim, J. P. W., Maintz, J. B. A., Viergever, M. A.: Mutual-information-based registration of medical images: a survey. *IEEE T. Med. Imag.* **22** (2003) 986–1004.
3. Kolda, T. G., Lewis, R. M., Torczon, V.: Optimization by direct search: new perspectives on some classical and modern methods. *SIAM Rev.* **45** (2003) 385–482.
4. He, R., Narayana, P. A.: Global optimization of mutual information: application to three-dimensional retrospective registration of magnetic resonance images. *Comput. Med. Imag. Graph.* **26** (2002) 277–292.
5. Jenkinson, M., Smith, S.: A global optimization method for robust affine registration of brain images. *Med. Image. Anal.* **5** (2001) 143–156.
6. Matsopoulos, G. K., Mouravliansky, N. A., Delibasis, K. K., Nikita, K. S.: Automatic retinal image registration scheme using global optimization techniques. *IEEE T. Inf. Technol. B.* **3** (1999) 47–60.
7. Wachowiak, M. P., Smolíková, R., Peters, T. M.: Multiresolution biomedical image registration using generalized information measures. *Proc. MICCAI 2003, LNCS 2879*, Ellis, R. E., Peters, T. M., Eds., (2003) 846–853.
8. Jones, D. R., Perttunen, C. D., Stuckman, B. E.: Lipschitzian optimization without the Lipschitz constant. *J. Optimiz. Theory App.* **79** (1993) 157–181.
9. Torczon, V.: On the convergence of the multidirectional search algorithm. *SIAM J. Optimiz.* **1** (1991) 123–145.
10. Warfield, S. K., Jolesz, F. A., Kikinis, R.: A high performance computing approach to the registration of medical imaging data. *Parallel Comput.* **24** (1998) 1345–1368.
11. Ourselin, S., Stefanescu, R., Pennec, X.: Robust registration of multi-modal images: Towards real-time clinical applications. *Proc. MICCAI 2002, LNCS 2489*, Dohi, T. and Kikinis, R. (Eds.) (2002) 140–147.
12. Gablonsky, J. M., Kelley, C. T.: A locally-biased form of the DIRECT algorithm. *J. Global Optim.* **21** (2001) 27–37.
13. Watson, L. T., Baker, C. A.: A fully-distributed parallel global search algorithm. *Engineering Computation* **18** (2001) 155–169.
14. Dennis, J. E., Torczon, V.: Direct search methods on parallel machines. *SIAM J. Optimiz.* **1** (1991) 448–474.
15. Kwan, R. K., Evans, A. C., Pike, G. B.: MRI simulation-based evaluation of image processing and classification methods. *IEEE T. Med. Imag.* **18** (1999) 1085–1097.
16. Wierzbicki, M., Peters, T. M.: Determining epicardial surface motion using elastic registration. *Proc. MICCAI 2003, LNCS 2878*, Ellis, R. E., Peters, T. M., Eds., (2003) 722–729.
17. García-Palomares, U., Rodríguez, J.: New sequential and parallel derivative-free algorithms for unconstrained minimization. *SIAM J. Optimiz.* **13** (2002) 79–96.