Differential Evolution for Medical Image Registration

Michel Salomon*
LSIIT-ICPS, CNRS
Université Louis Pasteur
Strasbourg
67400 Illkirch, France

Guy-René Perrin
LSIIT-ICPS, CNRS
Université Louis Pasteur
Strasbourg
67400 Illkirch, France

Fabrice Heitz
LSIIT-AIDI, CNRS
Université Louis Pasteur
Strasbourg
67400 Illkirch, France

Abstract *A common framework for 3D image registration consists in minimizing a cost (or energy) function that expresses the pixel or voxel similarity of the images to be aligned. Standard cost functions, based on voxel similarity measures, are highly non-linear, non-convex, exhibit many local minima and thus yield hard optimization problems.*

In this paper we consider a general purpose global optimization algorithm based on random sampling and evolutionary principles: differential evolution. Beside yielding accurate registrations, differential evolution appears as a robust algorithm, exhibits fast convergence and is easy to use, since few parameters are required.

Keywords: 3D medical image registration, optimization, differential evolution

1 Introduction

Many tasks in computer vision and image analysis have been expressed as global optimization problems. The general issue is to find the global minimum of an objective (also called cost or energy) function which describes the interaction between the different variables modeling the image features in a given problem [1].

The purpose of image registration (also called image matching) is to geometrically align one image (the floating or source image) to another (the reference or target image) so that voxels (or pixels) representing the same physical structure may be superimposed. Image registration is an important preliminary step in many computer vision tasks involving for instance, the reconstruction of 3D information from 2D views [2] or the fusion of multimodal images [3]. 3D medical imaging with its wide variety of sensors is probably one of the first application field, as are remote sensing, military imaging or multisensor robot vision. Applications of medical image registration range from computer assisted surgery to the analysis of functional images used for example to follow the evolution of diseases or to assess the efficiency of a treatment.

Standard cost functions, are highly nonlinear, non-convex and exhibit many local minima. Local, deterministic optimization algorithms are known to be sensitive to local minima. Therefore a global optimization algorithm is required to compute close to optimal solutions [4]. Early in the sixties, several researchers suggested to simulate the process of natural evolution. This early work led to the development of a new class of robust and efficient global search algorithms, known as evolutionary algorithms: genetic algorithms [5], evolution strategies [6, 7], and so on.

In the present paper, we consider differential evolution, an evolutionary algorithm introduced by Storn and Price [8], based on concepts from genetic algorithms and evolution strategies. We show that this algorithm is particularly suited for the registration of medical images using similarity measures, enabling an accurate (subvoxel) rigid registration of monomodal 3D images (Magnetic Resonance Images).

*Supported by the Pluri-Formations Program "Analyse et Synthèse Multi-images", Ministry of National Education, Research and Technology, France (1997/2000).

The paper is organized as follows. In section 2 we present two similarity metrics for single and multimodal image registration. The similarity-based registration algorithm is also described. In section 3 we give an overview on evolutionary algorithms, followed by a description of differential evolution. The fourth section is devoted to the experimental assessment of the performances of the differential evolution algorithm. A validation on a significant number of 3D image registration problems, with known ground truth, shows its efficiency and robustness.

2 Medical image registration using similarity measures

A complete review of standard registration techniques has been proposed by Brown [9]. A classification, in the framework of medical image analysis may be found in [10]. Similarity measure-based approaches rely on the minimization of cost functions that express the pixel or voxel similarity of the images to be aligned. They have been proposed for both single and multimodal image registration.

Registering images using similarity metrics consists in estimating the parameters Θ of the rigid transformation T_Θ minimizing a cost function C, that expresses the similarity between the single or multimodal image pair:

$$\Theta_{min} = \arg\min_{\Theta} \left[C(I(.), J(T_\Theta(.))) \right],$$

where (for 3D images):

$$\Theta = (t_x, t_y, t_z, \theta_x, \theta_y, \theta_z)^T$$

is a vector containing the 3D translation parameters (t_x, t_y, t_z) with respect to the $X, Y,$ and Z axis and the Euler rotation angles $(\theta_x, \theta_y, \theta_z)$. $I(.)$ represents the reference image and $J(.)$ the floating image (to be registered).

A classical similarity measure, widely used for registration of single modal images is the quadratic similarity measure [9]. This similarity metric assume that the two registered images differ only by an additive Gaussian noise, leading to the following cost function:

$$C(I(.), J(T_\Theta(.))) = \sum_s \left[I(s) - J(T_\Theta(s)) \right]^2 \quad (1)$$

where s designates the voxel (or pixel) coordinates.

For multimodal image registration, a metric yielding excellent results is the one based on the maximization of the mutual information (MI) [11, 12]. This measure is based on the fact that the mutual information $\mathcal{I}$ is maximized when the two images are correctly registered. The cost function to be minimized is then:

$$C(I(.), J(T_\Theta(.))) = -\mathcal{I}\left(I(s) - J(T_\Theta(s)) \right) \quad (2)$$

$$= -\sum_{g=1}^{G} \sum_{k=1}^{K} p(g,k) \log \left(\frac{p(g,k)}{p(g)p(k)} \right) \quad (3)$$

where G and K represents the number of grey levels of I and J. The joint probabilities $p(g,k)$ are the elements of the cooccurence matrix of $I(.)$ and $J(T_\Theta(.))$, and $p(g), p(k)$ are their respective marginal probabilities, both computed from the normalized histograms of the two images.

The two similarity measures presented above are highly non-linear, non-convex and have multiple local minima. Usually the minimization process is done by a local search algorithm that is initialized close to the optimal solution. In this work we consider differential evolution, a global search algorithm. To reduce the time complexity, the evolutionary optimization will be conducted on a sequence of multiresolution 3D grids, from coarse to fine resolution, until the full 3D image resolution is reached. In practice, the cost functions (1) and (3) are calculated by successively considering a voxel out of 81, 27, 9, 3 and finally every voxel in the MRI images ($128 \times 128 \times 128$ voxels). Multigrid algorithms are far less sensitive to local minima. Furthermore, it has been conjectured that multigrid analysis may smooth the "landscape" of the objective function to minimize [1].

3 An evolutionary approach for optimization

3.1 Overview on evolutionary algorithms

Evolutionary algorithms [13] are based on the paradigm of a population of individuals evolving by means of stochastic operators, in order to favor emerging individuals of better fitness. Each individual represents a potential solution to the optimization problem. The fitness value of an individual gives a qualitative information reflecting his behavior in the environment (related to the function to be optimized) and is obtained by transformation of the objective function value. The start population is usually initialized at random or may use some specific knowledge on the problem to be solved. Natural evolution is simulated by three stochastic operators: crossover, mutation and selection. Crossover (or recombination) performs exchange of information accross several individuals and mutation introduces new information by disrupting them.These two operators ensure exploration of the search space, whereas selection guides the search by favoring the reproduction of the best adapted individuals.

Let $P(t)$ represent the parent population in generation t, containing μ individuals a_j, $j \in \{1,...,\mu\}$, and let Φ denote the fitness function. Following the description given above, an evolutionary algorithm may be outlined as following :

$t = 0$ /* **Generation counter** */
Initialization($P(t)$) /* $P(t) = \{a_1(t), ..., a_\mu(t)\}$ */
Evaluate($P(t)$) /* $\{\Phi(a_1(t)), ..., \Phi(a_\mu(t))\}$ */
While *termination criterion not met* **do**
  $t = t + 1$
  $P'(t) =$ **Crossover**($P(t-1)$) **(or recombine)**
  Mutate($P'(t)$)
  Evaluate($P'(t)$)
  $P(t) =$ **Select**($P'(t) \cup Q(t)$) , $Q(t) \in \{\emptyset, P(t-1)\}$
End while

Depending on the class of evolutionary algorithms, selection operates on the offspring population $P'(t)$ or takes into account the parent population too ($Q(t) = P(t-1)$). The last one is usually preferred as it ensure that the so far known best individual is conserved ("elitism"). A maximum number of generated populations is typically used as termination criterion.

3.2 Differential evolution

The key idea in differential evolution is to use the difference between two randomly chosen individuals in the population for disrupting a third one. Differential evolution uses a real-valued representation, working directly on the parameter vector to be optimized. Therefore for the medical image registration problem described in section 2, we take $a = \Theta$, and the fitness function is simply defined as the cost function C, resulting in $\Phi(a) = C(I(.), J(T_\Theta(.)))$. As shown by figure 1, there are few differences between differential evolution and other evolutionary algorithms. In fact, only the reproduction step is different. Indeed Storn and al. [8] distinguish first an operator generating a new population by disrupting $P(t)$ (Step 1), followed by a two point crossover of each parent with its disrupted form (Step 2). In the following we merge these two operators in a single one that we call reproduction operator.

Let us assume that the population contains μ individuals, each individual containing n components $a_j = (a_{j,1}, a_{j,2}, ..., a_{j,n}), j \in \{1, ..., \mu\}$. Denoting a'_j the offspring resulting from reproduction of a_j, we can give the formal description of the reproduction operator.

$\chi_1, \chi_2 \in \{1, ..., n\}$ are the two crossover positions:

$$\chi_1 = (\chi \bmod n)+1, \chi_2 = ((\chi+L-1) \bmod n)+1$$

where χ and L are two integers uniformly drawn in set $\{1, ..., n\}$, for each individual. χ defines the first position of the crossover, whereas L is the number of parameters to be exchanged. Integer L is chosen so that:

$$P(L \geq \nu) = (p_c)^{\nu-1}$$

where $\nu > 0$ and $p_c \in [0, 1]$ is the crossover probability given by the user.

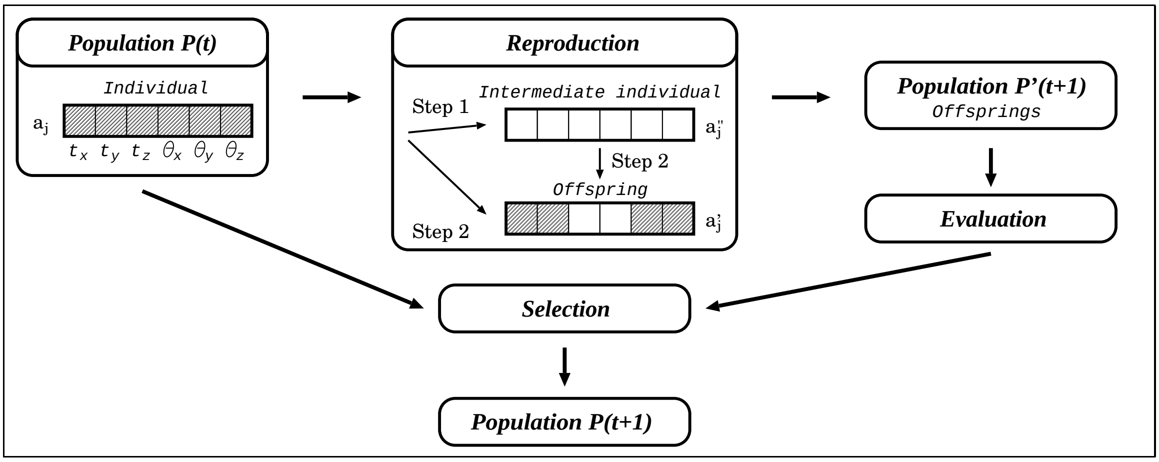

Figure 1: Generation of a new population.

The new individual $a'_j = (a'_{j,i})_{i \in \{1,...,n\}}$ is then formed according to the rule presented below:

$$a'_{j,i} = \begin{cases} a_{j,i} + \lambda \cdot (a_{min,i} - a_{j,i}) + F \cdot (a_{S,i} - a_{T,i}) \\ \quad \text{if } (\chi_1 \leq i \leq \chi_2) \vee \\ \quad (\chi_2 < \chi_1 \wedge (1 \leq i \leq \chi_2 \vee \chi_1 \leq i \leq n)) \\ \\ a_{j,i} \text{ otherwise} \end{cases}$$

with a_{min} denoting the current minimum; $S, T \in \{1, ..., \mu\}$ satisfying $S \neq T$, $S \neq j$, $T \neq j$; and $F, \lambda \in [0.1, 1]$ are factors controlling the amplification of the differential vectors. Usually to reduce the number of parameters, λ is set equal to F.

The new population is then obtained using a deterministic selection corresponding to a binary tournament between each parent and its respective offspring. More formally the selection operator can be described for a minimization task by:

$$\text{Selection}\left(P' \cup P\right) = \left\{s\left(a_j, a'_j\right), j \in \{1, ..., \mu\}\right\}$$

where $s(a_j, a'_j) = (1 - p_s(a'_j) \cdot a_j + p_s(a'_j) \cdot a'_j$ and $p_s(a'_j) = 1_{\mathbb{R}^*_+}(\Phi(a_j) - \Phi(a'_j))$. This kind of selection is called $\mu-$elitist. Resulting from several experiments with test functions and real problems, rules have been proposed [14] to guide the choice of the different parameters (μ, F, λ and p_c).

4 Experimental results

To evaluate the differential evolution algorithm on a representative set of 3D 128^3 MRI/MRI registration problems, we have applied twenty randomly generated rigid transformations on a reference MRI (figure 2) provided by the Institut de Physique Biologique (Strasbourg University Hospital). This set of registration problems, involve translation parameters between -20 and $+20$ voxels and rotations between -20 and $+20$ degrees. Let us notice that large rotations are generally difficult to handle, leading to objective functions with many local minima.

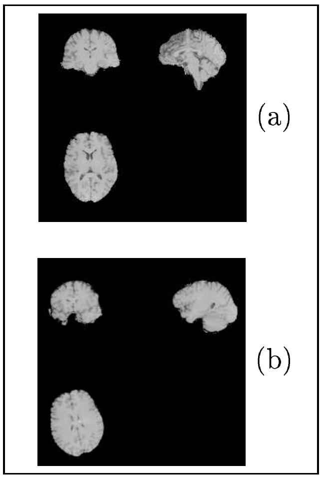

Figure 2: (a) Reference image and (b) one of the 20 floating images.

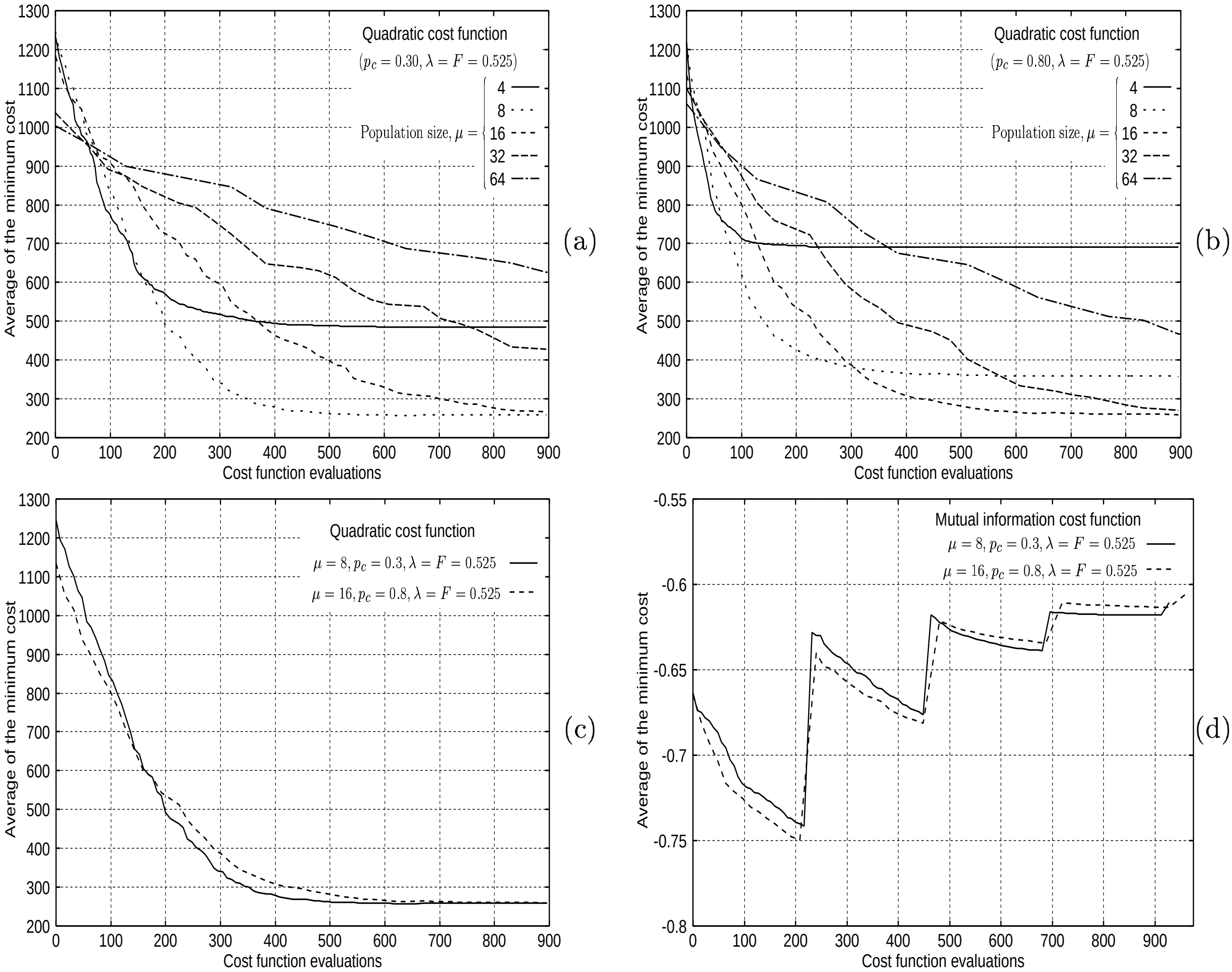

Figure 3: Average evolution of the minimum cost: in the case of the quadratic cost function considering different population size ((a) $p_c = 0.3$ and (b) $p_c = 0.8$; $\lambda = F = 0.525$); the two optimal curves for the quadratic cost function (c) and for the mutual information cost function (d).

As we have explained in section 2, the differential evolution algorithm is applied on a sequence of resolutions. We generate the same number of populations at all resolutions, except at the final one, in which case the population will be simply reevaluated, as it is the case when forwarding a population from a resolution to a finer one. This step is necessary because the cost associated to an individual vary with the resolution. Another point to be considered is the underlying interpolation model, since it influences the accuracy of the estimated parameters. We have adopted a trilinear interpolation model in the following experiments.

Suitable tuning parameters for differential evolution were found in some preliminary experiments, and we use a maximum number of generated populations (t_{max}) as a termination criterion. $\frac{t_{max}}{4}$ populations are generated by the evolutionary algorithm on the four first level. The total number of populations to be evaluated is $t_{max} + 5$ (evaluation of $P(0)$ plus the reevaluations needed when increasing the resolution).

Figures 3 presents the average evolution of the minimum cost, for the twenty registration problems, during the optimization process. The two first figures show for the quadratic cost function, the influence of the population size according to the crossover probability ((a) $p_c = 0.3$, (b) $p_c = 0.8$; λ and F are fixed to 0.525; t_{max} is chosen in order to evaluate a little more than 900 individuals). We notice that a population of less than 20 individuals is sufficient: 8 individuals for a crossover probability near 0 and 16 individuals for a value near 1. The two following figures (3(c) and 3(d)) show the minimum cost evolution for these two optimal cases, respectively for the quadratic cost function and the mutual information cost function. The four peaks that appear in the last figure are due to the change of resolution which induce a change in the value of the mutual information objective function.

To assess the accuracy of the registration, we compare the estimated registration parameters to the true parameters. Besides computing statistics on the registration errors, we have used the Root Mean Square error. This error metric corresponds to the average misregistration between voxels in the proposed solution (the floating image registered by the algorithm) and the optimal one. It can be seen from statistics gained for the set of twenty images (Table 1), that the algorithm achieves subvoxel accuracy. Moreover, the RMS errors computed on a sampling of voxels are on average below 0.50 voxels for both metrics.

As far as execution time is concerned, our method compares favourably to other approaches. For the quadratic cost function, differential evolution takes approximatively 11 minutes (MIPS R12000 processor, 300MHz), whereas for the multimodal metric case it increase to 16 minutes 30 seconds (3D 128^3 images). Cost function (3) requires more computations to evaluate the join and marginal probabilities. To obtain an additional gain in time complexity, the parallelization of the global optimization step can be considered. Thus, we have proposed a data-parallel implementation [15] that gives an execution time of less than one minute when using as many processors as individuals in the population (here for a size of 16 individuals).

Table 1: Single Modal (MRI/MRI) Registration (3D) $\{\lambda = F = 0.525\}$: error statistics (mean $\pm$ std. dev.).

Cost function ; Parameters	
Quadratic ; $\mu = 8, p_c = 0.3$ and $t_{max} = 108$	
Δt_x	0.19 ± 0.15
Δt_y	0.14 ± 0.10
Δt_z	0.16 ± 0.09
$\Delta \theta_x$	0.005 ± 0.004
$\Delta \theta_y$	0.004 ± 0.003
$\Delta \theta_z$	0.009 ± 0.006
Quadratic ; $\mu = 16, p_c = 0.8$ and $t_{max} = 52$	
Δt_x	0.20 ± 0.15
Δt_y	0.15 ± 0.10
Δt_z	0.16 ± 0.09
$\Delta \theta_x$	0.030 ± 0.040
$\Delta \theta_y$	0.020 ± 0.030
$\Delta \theta_z$	0.040 ± 0.080
MI ; $\mu = 8, p_c = 0.3$ and $t_{max} = 112$	
Δt_x	0.19 ± 0.15
Δt_y	0.14 ± 0.10
Δt_z	0.17 ± 0.09
$\Delta \theta_x$	0.020 ± 0.020
$\Delta \theta_y$	0.020 ± 0.020
$\Delta \theta_z$	0.030 ± 0.030
MI ; $\mu = 16, p_c = 0.8$ and $t_{max} = 56$	
Δt_x	0.23 ± 0.17
Δt_y	0.17 ± 0.14
Δt_z	0.16 ± 0.09
$\Delta \theta_x$	0.040 ± 0.050
$\Delta \theta_y$	0.040 ± 0.040
$\Delta \theta_z$	0.070 ± 0.070

Note. The translation errors are given in voxels and the rotation errors in degrees.

5 Conclusion and future work

We have shown that differential evolution produces subvoxel registrations with an execution time that compares favourably to other approaches (it can be further reduced using a data-parallel implementation). The experiments show that differential evolution is a robust and flexible optimization method. As it only requires evaluation of the cost function (and no derivatives computation), it is well adapted to the minimization of the highly non-linear, eventually discontinuous objective functions that are considered in multimodal image registration. Furthermore, from work presented in [16] and [17] it appear that genetic algorithms and evolution strategies are also good appealing candidate for medical image registration. So, we plan to make a comparative study of different evolutionary algorithms.

References

[1] F. Heitz, P. Perez, and P. Bouthemy. Multiscale minimization of global energy functions in some visual recovery problems. *Comput. Vis. Graphics Image Proc. Image Understanding*, 59(1):125–134, 1994.

[2] O. Faugeras. *Three-dimensional computer vision : a geometric viewpoint.* MIT Press, 1993.

[3] S.T.C. Wong, R.C. Knowlton, R.A. Hawkins, and K.D. Laxer. Multimodal image fusion for noninvasive epilepsy surgery planning. *IEEE Computer Graphics and Applications*, 16(1):30–38, 1996.

[4] C. Nikou, F. Heitz, J.-P. Armspach, I.-J. Namer, and D. Grucker. Registration of mr/mr and mr/spect brain images by fast stochastic optimization of robust voxel similarity measures. *Neuroimage*, 8:30–43, 1998.

[5] J.H. Holland. *Adaptation in natural and artificial systems.* University of Michigan Press, 1975.

[6] I. Rechenberg. *Evolutionsstrategie : optimierung technischer systeme nach prinzipien des biologischen evolution.* Frommann-Holzboog Verlag, Stuttgart, 1973.

[7] H.-P. Schwefel. *Numerische optimierung von computer-modellen mittels der evolutionsstrategies*, volume 26 of *Interdisciplinary Systems Research.* Birkhäuser, Basel, 1977.

[8] R. Storn and K. Price. Minimizing the real functions of the icec'96 contest by differential evolution. In *Int. Conf. on Evolutionary Computation*, Nagoya, Japan, 1996.

[9] L.G. Brown. A survey of image registration techniques. *ACM Computing Survey*, 24(4):325–376, 1992.

[10] P.A. van den Elsen, E.J.D. Paul, and M.A. Viergever. Medical image matching - a review with classification. *IEEE Engineering in Medicine and Biology*, 12(4):26–39, 1993.

[11] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximisation of mutual information. *IEEE Trans. on Medical Imaging*, 16(2):187–198, 1997.

[12] W. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multimodal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):33–51, 1996.

[13] T. Bäck. *Evolutionary algorithms in theory and practice.* Oxford University Press, 1996.

[14] R. Storn. On the usage of differential evolution for function optimization. In *NAFIPS'96*, pages 519–523, Berkeley, 1996.

[15] M. Salomon, G.-R. Perrin, and F. Heitz. Parallelizing differential evolution for medical image registration. Technical Report 00-06, LSIIT, Université Louis Pasteur, Strasbourg, September 2000.

[16] J.-M. Rouet, J.-J. Jacq, and C. Roux. Genetic algorithms for a robust 3-d mr-ct registration. *IEEE Trans. on Information Technology in Biomedicine*, 4(2):126–136, 2000.

[17] D. Fischer, P. Kohlhepp, and F. Bulling. An evolutionary algorithm for the registration of 3-d surface representations. *Pattern Recognition*, 32:53–69, 1999.