An Image Registration Approach using Genetic Algorithms

Andrea Valsecchi and Sergio Damas
European Centre for Soft Computing
Mieres, Spain
{andrea.valsecchi, sergio.damas}@softcomputing.es

José Santamaría
University of Jaén
Jaén, Spain
jslopez@ujaen.es

***Abstract*—We introduce a new intensity-based image registration (IR) technique based on a modern, real-coded genetic algorithm. Our proposal is tested on 16 registration scenarios involving real-world MRI medical images. A novel methodological framework to compare heterogeneous IR algorithms is also described. Following such methodology, our algorithm is compared with four well-known IR techniques of different natures. The proposed method is able to improve the results of these techniques in the majority of the scenarios.**

I. Introduction

A large number of applications in image processing require the integration of information from multiple images of the same or similar subjects obtained under different conditions (time, viewpoint, sensor or any combination of the latter). Hence, the images need to be properly aligned in order to put in correspondence the common content. This task is called *image registration* (IR) [1]: given two images, image registration aims to find the geometric transformation leading to the best possible overlapping.

IR approaches usually fall into two categories: *intensity-based* (or *voxel-based*) and *feature-based* methods. The former make use of the entire images while the latter employ only salient, distinctive objects such as lines, corners and contours, detected in a preprocessing step. Feature-based techniques are faster, as they use only a fraction of the imaging data, but usually suffer from inevitable errors in the feature detection process. Independently of their nature, IR techniques involve an iterative *optimization procedure* that explores the space of possible transformations. Registration approaches based on *evolutionary algorithms* (EAs) have proven to be a promising solution to overcome the drawbacks of traditional gradient-based algorithms [2], [3]. In fact, they are considered global optimization approaches able to perform a robust search in complex search spaces like those arising in IR.

The contribution of this work is two-fold. First, we propose a methodological framework to provide an objective evaluation of the performance of heterogeneous IR methods according their nature, feature or intensity-based. Also, we introduce a new intensity-based IR method based on evolutionary computation. The proposed method will be validated on a number of realistic medical image registration scenarios, and the experimental results will be analyzed in detail and compared with those of state-of-the-art techniques, both feature- and intensity-based.

The paper is structured as follows. Section II introduces the IR problem in detail. Notable IR methods are reviewed and discussed in Section III. Section IV describes the proposed approach, whereas the comparison framework is discussed in Section V. Section VI introduces the test scenarios, the experiments performed and the analysis of their results. Finally, Section VII provides conclusions and directions for future work.

II. Image registration

In a typical problem instance we are provided with two images: a reference image, the *model*, and the image that will be transformed to reach the model geometry, called *scene* [1]. We will denote these two images by I_M and I_S respectively. The result of the registration process is a transformation f such that the model I_M and the transformed scene $f(I_S)$ are as similar as possible.

IR methods can be characterized by their three main components: the *transformation model*, the *similarity metric* and the *optimization process*. The transformation model determines what kind of transformation is used to align the images. For instance, a rigid transformation is a combination of translation and rotation operations, while similarity transformations also allow scaling. Their degrees of freedom for 3-D images are 6 and 7, respectively. B-splines and thin-plate splines are instead examples of *elastic* (or *non-rigid*) transformations models, able to represent local deformations (warping). In applications, the appropriate transformation model depends on both the nature of the images and the particular application involved.

A similarity metric is a function F that measures the quality of a solution of an IR problem. To evaluate a solution f, the scene image I_S is transformed according to f and then the degree of resemblance between the transformed scene image $f(I_S)$ and the model image I_M, denoted by Ψ, is computed, so $F(I_M, I_S, f) = \Psi(I_M, f(I_S))$. Several choices for Ψ can be found in the literature, depending on the nature of the considered images. In feature-based approaches, metrics are usually based on the distance between corresponding geometric primitives [4], such as mean square error (MSE), which is the average square distance between corresponding feature (in this case points) in the scene and the model images. To compute the MSE, each point of the model is assigned

to the closest point in the transformed scene, regardless of whether the latter had been already assigned to another model point. That is, MSE $= \frac{1}{r}\sum_{i=1}^{r}\|x_i - c_i\|^2$ where c_i is the point of $f(I_S)$ that is closest to x_i.

In intensity-based approaches, common choices are sum of squared differences, normalized correlation (NC) and mutual information (MI) [5]. In particular, MI is specially suited for multi-modal registration and other scenarios in which the images have different intensity distributions. It is defined as

$$\text{MI} = \sum_{s\in L_S, m\in L_M} p(m,s,f)\log_2\frac{p(m,s,f)}{p_M(m)\ p_S(s,f)}$$

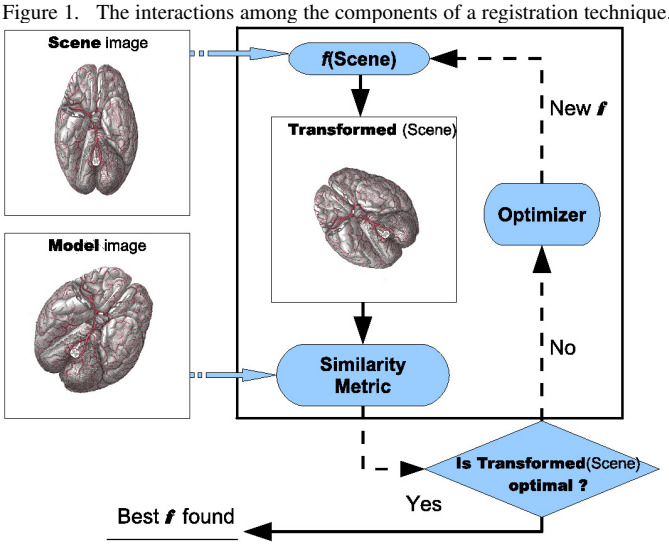
where L_M and L_S are sets of regularly spaced intensity bin centers, p is the discrete joint probability and p_M, p_S are the marginal discrete probabilities of the model and scene image. Normalized correlation is defined as the correlation coefficient between the images intensity distributions, so

$$\text{NC} = \frac{\sum_{i\in\Omega_M}\left(\ I_M(i) - \overline{I_M}\ \right)\left(\ I_{f(S)}(f(i)) - \overline{I_S}\ \right)}{\sum_{i\in\Omega_M}\left(\ I_M(i) - \overline{I_M}\ \right)\sum_{i\in\Omega_S}\left(\ I_S(f(i)) - \overline{I_{f(S)}}\ \right)}$$

where Ω_M, Ω_S are the images' voxels and $\overline{I_M}, \overline{I_{f(S)}}$ are average intensity values. Normalized correlation is conceptually simpler and faster to compute, but is suitable only when the images' intensity distributions are similar.

Finally, the optimization procedure is the component responsible for finding an appropriate transformation to carry out the registration. Figure 1 show the flow chart of the whole registration process. The search strategy adopted depends on the nature of the algorithm. In matching-based algorithms, once the images features have been detected, the optimizer looks for a matching between them and the transformation parameters are derived from the match. The process is iterated until reaching convergence within a tolerance threshold of the concerned similarity metric.

Figure 1. The interactions among the components of a registration technique.



Instead, in parameters-based methods the search is performed directly in the transformation parameters space. Classic numerical optimization algorithm like gradient descent, Newton's method, Powell's method and discrete optimization [6] are among the most common choices, together with approaches based on EC and other meta-heuristics [2], [3]. It is common to start the registration process using a "simpler" version of the images obtained through smoothing and down-sampling. The registration is divided in multiple stages, called *resolutions*, in which increasingly larger and more detailed versions of the input images are used.

III. Outstanding feature- and intensity-based IR approaches

This section describes a selection of the most remarkable IR methods, that later will be involved in the experimental study along with the IR approach presented in the next section.

A. I-ICP

I-ICP has been introduced by Liu in [7] as an improvement of the original ICP proposal. The latter is a well-known matching-based algorithm in computer-aided design, originally proposed to recover the 3D transformation of pairs of range images. The novelty in I-ICP compared to its predecessor is the use of both the collinearity and closeness constraints to evaluate the possible correspondences established by the traditional ICP criterion. The collinearity constraint minimizes the distance between the transformed points and the ray passing through the corresponding points, while the closeness constraint minimizes the distance between matching points. In order to improve the robustness of the algorithm, a tiny perturbation is added to the estimated motion parameters when the algorithm is about to terminate. As a result, the algorithm is forced to search for an optimal solution in the neighboring region of the estimated parameters.

B. SS*

SS* is a matching-based IR method based on Scatter Search (SS) [8]. It has been introduced in [9], in which SS* was found to be the best algorithm in a comparison of feature-based IR techniques. In the SS* proposal, the authors designed new improvement, combination and diversification-generation methods. Crest-lines points [10], [11] are used as features, and the algorithm exploits the knowledge of the local curvature of those points to perform the matching. The authors proposed an advanced coding scheme where a matching is represented as a permutation of points. Besides, they define a function m that evaluates the quality of a matching π using the mentioned curvature values as follows: $m(\pi) = \Delta k_1 + \Delta k_2$ where $\Delta k_j = \sum_{i=1}^{r}(k_j^i - k_j^{\pi_i})^2$ for $j = 1, 2$. The values Δk_1 and Δk_2 measure the error in terms of the first and second principal curvatures. To evaluate a solution, the information about the matching provided by a similarity metric g (MSE) and m are combined into F according to the formula $F(\pi) = w_1 \cdot g(\pi) + w_2 \cdot m(\pi)$ where the weighting coefficients w_1, w_2 define the relative importance of each term.

C. Dyn-GA

Dyn-GA [12] is a parameter-based IR techniques based on genetic algorithms. The algorithm deals with 3D rigid transformations, that are encoded as real-coded vector of six elements: three values encode the rotation (through Euler angles) and three parameters represent the translation component. The GA uses fitness-proportionate selection and two novel crossover and mutation operators. Crossover selects randomly a number of genes to be swapped between two individuals, while mutation replaces the value of a randomly selected gene. For the rotation components, the new value is drawn at random from a fixed range, while for the translation components, the range of the new value is dynamically computed from the fitness of the individual: the larger the fitness the larger is this range. This allows the GA to perform a small mutation if the individual is close to the optimum and large mutation otherwise.

A sophisticated restart mechanism, named *dynamic boundary* is also used. The rationale of this method is that once the GA has reach a stationary state (i.e. the fitness of the best individual have not improved for a number of generations), the individuals in the population are close to the optimum, and therefore we can concentrate the search in the area in which the current population lies. The valid ranges for the genes of an individual are thus restricted around the range of values already in the population.

D. GD, an IR approach based on gradient descent

In [13] the authors introduced *elastix*, a toolbox for intensity-based medical image registration. They also presented an experimental study over 50 clinical MRI scans of prostates, in which they compared several gradient-based methods. Following the guidelines resulting from the study, we designed a representative for this kind of methods and named it GD. MI is used as similarity metric and the transformation model is similarity transformations. The optimizer is an adaptive stochastic gradient descent (ASGD) [14], which is used to minimize a cost function C with parameter x. From an initial solution x_o, ASGD considers the solutions $x_{k+1} = x_k - \gamma_k g_k$ where $t_{k+1} = \max(0,\ t_k + \text{sigm}(-g_k \cdot g_{k-1}))$, $g_k = \frac{dC}{dx}(x_k)$ and $\gamma_k = \frac{a}{t_k + A}$. The values $a > 0$ and $A \geq 1$ are user-specified constants, while sigm is the sigmoid function. GD uses a random image sampler and a multi-resolution strategy. The images in the pyramids are obtained by applying both downsampling and Gaussian smoothing to the original images.

IV. GA$^+$, an evolutionary approach to IR

This section describes the methodology proposed in this work. Following the intensity-based approach, our method tackles IR using the images in their whole, rather than considering only some of their features. Moreover, the registration is carried out through a search in the space of transformation parameters. We have chosen similarity transformations as our transformation model, so our method deals with transformations composed by translation, rotation and uniform scaling, although the approach can be easily adapted to other models.

In three dimensions a transformation can be represented by seven real numbers: three to specify the versor of the rotation v, three for the translation t and one for the scaling factor s. Therefore, an individual of the GA is simply a real vector with seven elements. Valid solutions require $v_x, v_y, v_z \in [-1, 1]$ and $s > 0$. Note that the translation component is specified in spatial units (e.g, millimeters), rather than in number of voxels.

The operators used in our method are common choices for real-coded genetic algorithms: blend crossover (BLX-α) [15] and random mutation [16]. The fitness value of a solution f is simply the similarity between the two input images when registered using f, i.e. $f \rightarrow \Psi(I_M, f(I_S))$, where Ψ is a similarity metric; we considered both MI and NC. Finally, the parents selection is performed following the tournament approach.

Finally, the method also adopt a *restart* strategy. For a number of causes, it may happen that at the end of the first resolution the best transformation found by the optimizer has a very low quality, i.e. it does not even provide a coarse registration. Thus, it may not be appropriate to proceed to refine this solution in the further resolutions, but rather to perform again the search for a suitable initial registration. The algorithm proceeds as follows: if the metric value of the best solution has not reached an appropriate amount at the end of the first resolution, the current state of the optimization process is discarded and the whole procedure is executed again from the beginning. As the optimization in the first resolution deals with a low-resolution version of the input images, this stage of the registration is the cheapest in terms of computational effort, thus performing a restart increases the computational cost of the whole method only by a fraction of its total.

V. A methodological framework for image registration

In order to provide an objective evaluation of the performance of different IR methods, a common methodological framework is required. Indeed, as we reviewed in Section II, most IR methods differ in their optimization procedure and in particular in the similarity metric, which measure the quality of a solution. The question arises on how to measure the performance of heterogeneous IR methods, given one cannot simply compare the similarity values of their output. In particular, we aim to compare the performance of feature-based IR methods (typically guided by MSE) and intensity-based IR methods (usually guided by either MI or NC).

Our approach is to evaluate all solutions using a common similarity metric. For instance, if MSE is chosen, all solutions are finally evaluated computing the MSE between the transformed scene's features and the model's features. This happens irregardless of whether these solutions has been produced by an intensity-based algorithm or a feature-based algorithm that uses a different metric. This approach is just part of the final comparison of all methods and it does not affect the original behavior of each particular IR method, that is implemented according to the proposal of their authors.

VI. Experimental study

In this section we present a number of experiments that have been developed on a medical image data set in order to study the performance of our proposed method. The most important challenge is that the goal of the IR process is to register two *different images of similar objects*, instead of two images corresponding to the same object, thus reflecting a more realistic situation in medical IR.

We made use of a data set from the BrainWeb repository [17]. It contains four simulated real-world magnetic resonance images (MRIs) of four human brains with noise, anatomical lesions, and a certain degree of occlusion; see Table I. We considered the same images and registration scenarios that have been used in [9], in which the authors developed a comparison among state-of-the-art feature-based IR techniques. The use of the same scenarios will provide useful reference points for the assessment of the quality of the results of our proposal. The original comparison included all the feature-based algorithms we review in Section III. We extended this work by testing our approach and gradient-based algorithm GD on the same scenarios. The resulting comparison thus comprises algorithms having different nature and search strategies, as well as representatives of various optimization techniques: gradient-based, evolutionary and other metaheuristics.

A. Experimental setup

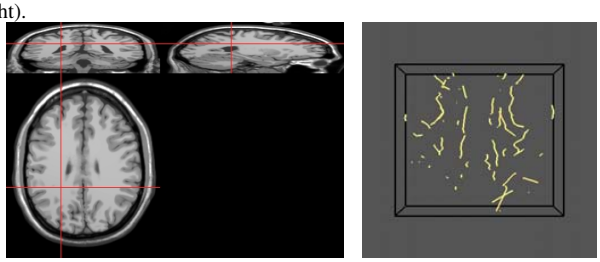
The results presented in this section correspond to a number of registration problems with four different simulated real-world MRIs. These images have been obtained from the BrainWeb database at McGill University [17]. Its purpose is to provide researchers with ground truth data for image-analysis techniques and algorithms. BrainWeb has been widely used by the IR research community [18]. Different levels of noise were added to three of the four images, to model the noisy conditions in which such images are usually acquired. Moreover, two of the images include a multiple sclerosis lesion. The influence of these two factors allow to design a set of experiments with different complexity levels. The images' characteristics are reported in Table I; an example image is shown in Figure 2. All images have size $60 \times 181 \times 217$.

Table I
THE NOISE LEVEL AND THE PRESENCE OF LESION IN THE FOUR BRAIN MRI IMAGES USED IN THE EXPERIMENTAL STUDY. THE NUMBER OF CREST LINE POINTS, USED AS FEATURES, IS ALSO REPORTED.

Image	Lesion	Noise (%)	# of features
I_1	No	0	583
I_2	No	1	393
I_3	Yes	1	348
I_4	Yes	5	248

To provide an input for the feature-based algorithms considered in the comparison, the appropriate kind of feature need to be extracted. In this case, all the algorithms use points; in the original comparison, the authors computed the isosurfaces and extracted the crest line points with relevant curvature information [19] (see Figure 2). It is important to remark this difference. The actual input of intensity-based methods consists of the whole images data; in this concrete case, two images made of $60 \times 181 \times 217 = 2356620$ voxels having a 8-bit intensity value. The input of feature-based approaches, instead, is a set of composed of a few hundred points (Table I).

Figure 2. An MRI brain image (left) and the corresponding crest line points (right).



B. Registration scenarios

In the IR problem instances considered in this study, different 3D images have been transformed using four global similarity transformations T_1, T_2, T_3 and T_4, shown in Table II. As mentioned in Section II, similarity transformations involve rotation, translation, and uniform scaling. In Section IV we described a way to represent such a transformation using seven real parameters, but here we used a more easily understandable alternative which consists of eight parameters: the rotation magnitude (λ), the rotation axis (a_x, a_y, a_z), the translation vector (t_x, t_y, t_z) and the uniform scaling factor s. To achieve a good solution, every algorithm must estimate these eight parameters accurately. Values in Table II have been selected within the appropriate ranges so that important transformations have to be estimated. Both rotation and translation vectors represent a strong change in the object location. In fact, the lowest rotation angle is 115°. Meanwhile, translation values are also high. Likewise, the scaling factor ranges from 0.8 (in the second transformation) to 1.2 (in the fourth one). In this way, complex IR problem instances were generated.

Table II
PARAMETERS OF THE SIMILARITY TRANSFORMATIONS WE USED IN THE EXPERIMENTS.

	λ	a_x	a_y	a_z	t_x	t_y	t_z	s
T_1	115	-0.863	0.259	0.431	-26	15.5	-4.6	1
T_2	168	0.676	-0.290	0.676	6	5.5	-4.6	0.8
T_3	235	-0.303	-0.808	0.505	16	-5.5	-4.6	1
T_4	276.9	-0.872	0.436	-0.218	-12	5.5	-24.6	1.2

In the experiments, we considered 16 image registration scenarios. From lower to higher complexity, these are I_1 versus $T_i(I_2)$, I_1 versus $T_i(I_3)$, I_1 versus $T_i(I_4)$ and I_2 versus $T_i(I_4)$, for $i = 1, 2, 3, 4$.

C. Parameter settings

We tested GA$^+$ using different configurations. For the similarity metric, we considered both normalized correlation and mutual information. We also tested the algorithm with and

without the restart capability. In the first case, the algorithm performs a single restart if the similarity metric value of the best solution is lower than 0.9 for NC and 0.4 for MI at the end of the first resolution. These values, as well as the one used for other parameters that will be introduced later, were determined through a preliminary study performed over the registration scenario I_1 versus $T_1(I_1)$, that is not included in the subsequent experimentation. The registration was performed in two resolutions. During the first one, the algorithm used a smoothed (Gaussian smoothing, $\sigma = 4$), downsampled (by a 4 factor) version of the input images, and the similarity metric was computed over 5000 random spatial samples. For the second resolution the algorithm, no preprocessing were applied to the images, and the number of spatial samples was increased to 20000. When the restart was disabled, the GA used the same configuration for both resolutions: population size of 500 individuals, 100 generations, mutation probability of 0.1, crossover probability of 0.5, blend factor (α) 0.3 and tournament size equal to 3. When restart occurred the number of generations in the second resolution were reduced to 75 to keep the computational cost roughly constant. Finally, as we are dealing with a non-deterministic algorithm, 15 independent runs were performed for each test scenarios.

The algorithm is also provided with the ranges of the transformation parameters; taken into account the scenario were are dealing with, we used $[-30, 30]$ for the translation components (t_x, t_y, t_z) and $[0.75, 1.25]$ for the scaling one. No restrictions were placed over the rotation, so the actual ranges are $[-1, 1]$ for each component of the versor.

For GD we used the parameters values suggested in [13]: the registration was performed in four resolutions in which 2000 iterations of the optimization process were performed. The parameter settings used for SS*, I-ICP and DynGA are those of the original comparison [9]. We performed a total of 15 runs (with different random seeds) for each of the problem instances.

Finally, we used running time as stopping criteria. As we are dealing with both feature and intensity-based techniques, we let the algorithms run for different amounts of time according to their nature. Indeed, recall that while intensity-based methods use the whole images data (or at least a large subset), only a small fraction of the input data is actually used in feature-based techniques. Also, the time needed to extract the features used by the latter algorithms is not counted. Taken into account these differences, we used a time limit of 20 seconds for the feature-based approaches (I-ICP, SS* and Dyn-GA) and one of 20 minutes for the intensity-based methods (GA$^+$ and GD). These amounts were determined by considering the proportion between the size of the data used by algorithms of two classes. The number of spatial samples of GA$^+$ and GD is 25000, while the average number of feature points in the test images is 393. The proportion is 63.6, which roughly approximates the proportion between the amounts of time, 60.

D. Analysis of results

In accordance to the methodology proposed in Section V, we evaluate the solutions provided by the different IR techniques using a common similarity metric. We chose the MSE over the crest line points we provided as input to feature-based methods. Table III show the results of the experiments on the first nine scenarios; the results obtained on the remaining seven scenarios are not reported due to lack of space. For each IR method, the table provides the minimum, maximum, average and standard deviation of the MSE values obtained tackling the IR problems. We printed in bold the best minimum and average MSE values for each scenario. We also compared each pair of IR methods and computed the number of scenarios in which one performed better than the other with respect to the minimum (Table IV) and the average MSE (Table IV). Figure 3 offers a visual comparison of some of the results.

1) Comparison among the variants of GA$^+$: We will undertake a detailed comparison of the results of the four different variants of GA$^+$ we tested. To support the discussion, we have drawn the boxplot to the distribution of MSE obtained by the algorithms over the 16 IR scenarios in Figure 4.

a) Effect of the similarity metric: The minimum MSE values of GA$^+$-MI and GA$^+$-NC are extremely close in all scenarios, with the largest difference being 1. However, the latter performed consistently better, having a smaller minimum value in 13 cases over 16. When we consider the average MSE values, we note that GA$^+$-NC has a smaller $\bar{x}$ value in 14 scenarios and the difference is often really large, reaching 6038 on I_1 vs $T_3(I_3)$. A similar picture is provided by the comparison between the restart versions of GA$^+$. With respect to the minimum, GA$^+$-NC-Re has a better MSE than GA$^+$-MI-Re in 13 scenarios over 16, but the difference in value is again extremely small. The difference average MSE, instead, varies from less than 1 (several scenarios) up to 3909 on I_2 vs $T2(I4)$. GA$^+$-NC-Re outperformed GA$^+$-MI-Re in terms of average MSE in 13 out of 16 scenarios. This allows us to conclude that normalized correlation lead to more accurate registrations on the average in comparison with mutual information.

b) Effect of the restart policy: By considering the boxplots in Figure 4, one can note that the distributions of the errors are usually extremely close to the bottom part of the graph, but they contains also a few really large values. This is also confirmed by the position of the medians and the third quartiles, which are generally close to the minimum. Thus, for all scenarios, the algorithms converges to a good solution at least in half of the times. However, the solutions provided have a really low quality when convergence does not occur, and actually their quality of is so low that even one of them is able to increase the average MSE value of more than one order of magnitude. This phenomenon affects also the version of GA$^+$ using normalized correlation, but we did not report the boxplot due to lack of space. This behavior motivate the use of the restart mechanism: basically we trade some time used for refining our solution to decrease the probability of one "bad" run to happen.

Table III
THE EXPERIMENTAL RESULTS OVER THE FIRST 9 IR SCENARIOS IN TERMS OF MSE. MINIMUM (m), MAXIMUM (M), AVERAGE($\bar{x}$) AND STANDARD DEVIATION (s) VALUES ARE REPORTED.

	I_1 vs $T_1(I_2)$				I_1 vs $T_2(I_2)$				I_1 vs $T_3(I_2)$			
	m	M	$\bar{x}$	s	m	M	$\bar{x}$	s	m	M	$\bar{x}$	s
Dyn-GA	101.3	263.7	194.9	50.5	43.7	283.6	107.5	52.1	86.6	678.4	211.0	137.3
GD	58597.7	61316.2	59956.9	1922.3	24207.4	29140.2	26673.8	3488.0	11154.0	109799.0	60476.5	69752.5
I-ICP	344.4		344.4		130.7		130.7		894.3		894.3	
GA$^+$-MI	36.1	50928.4	8197.8	18383.2	35.4	31256.7	4661.2	10970.5	39.2	25046.2	6665.4	10756.0
GA$^+$-MI-Re	36.1	49631.5	725.5	5844.8	**35.3**	37.6	**36.7**	0.4	39.2	24162.1	1374.2	5521.6
GA$^+$-NC	35.8	50008.4	3446.1	12570.4	35.8	32233.8	1355.4	6369.4	39.2	25435.5	6715.7	10844.8
GA$^+$-NC-Re	35.8	36.9	**36.4**	0.3	35.8	37.9	36.7	0.4	**39.1**	24992.1	1736.5	6216.7
SS*	**34.6**	39.7	37.0	1.5	36.8	49.8	43.4	3.6	57.0	67.2	**63.2**	2.9

	I_1 vs $T_4(I_2)$				I_1 vs $T_1(I_3)$				I_1 vs $T_2(I_3)$			
	m	M	$\bar{x}$	s	m	M	$\bar{x}$	s	m	M	$\bar{x}$	s
Dyn-GA	139.4	600.3	302.0	121.4	132.2	740.9	299.3	144.1	55.7	534.1	154.0	114.2
GD	33.3	35.6	34.4	1.7	60286.3	62012.5	61149.4	1220.6	25895.3	29263.1	27579.2	2381.4
I-ICP	631.7		631.7		517.7		517.7		330.3		330.3	
GA$^+$-MI	32.7	33.2	32.9	0.1	51.1	51144.7	4124.5	13607.1	43.4	30722.7	1718.5	6955.0
GA$^+$-MI-Re	32.7	33.1	32.9	0.1	51.1	52.3	51.7	0.3	43.3	44.1	**43.7**	0.2
GA$^+$-NC	32.3	33.1	32.7	0.2	50.9	59639.0	2437.9	11544.8	43.3	31328.7	478.3	3687.0
GA$^+$-NC-Re	**32.2**	33.1	**32.6**	0.2	**50.8**	51.9	**51.4**	0.2	**43.2**	44.2	43.8	0.2
SS*	48.8	58.8	53.9	2.6	89.8	131.7	112.2	12.4	49.5	65.8	56.7	4.5

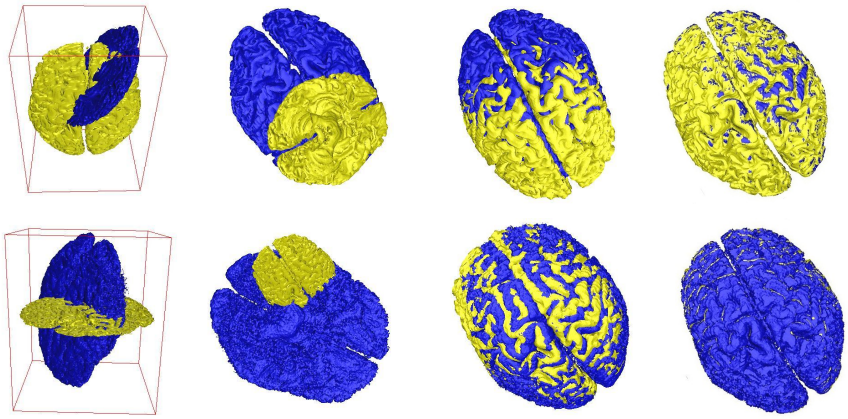
	I_1 vs $T_3(I_3)$				I_1 vs $T_4(I_3)$				I_1 vs $T_1(I_4)$			
	m	M	$\bar{x}$	s	m	M	$\bar{x}$	s	m	M	$\bar{x}$	s
Dyn-GA	138.8	839.0	326.5	174.0	221.3	841.3	354.3	146.9	123.7	1083.7	255.4	228.2
GD	10535.0	107814.0	59174.5	68786.6	45.2	45.8	45.5	0.4	59026.2	59126.7	59076.4	71.1
I-ICP	437.8		437.8		478.0		478.0		704.3		704.3	
GA$^+$-MI	55.3	26148.8	10789.9	11974.3	44.3	45.7	45.0	0.3	52.7	50293.6	5521.6	15577.8
GA$^+$-MI-Re	55.3	25387.8	6379.7	10635.1	44.3	45.7	45.0	0.3	52.7	54.0	53.3	0.3
GA$^+$-NC	54.3	25461.2	4752.2	9618.7	43.9	45.3	44.6	0.3	**52.5**	53949.8	4252.2	14061.0
GA$^+$-NC-Re	54.3	25461.2	1091.8	4965.8	**43.8**	45.3	**44.5**	0.3	**52.5**	53.5	**52.9**	0.3
SS*	**43.7**	235.1	**63.8**	46.2	112.3	143.2	122.7	8.2	149.3	269.0	183.6	33.0

Table IV
THE COMPARISON BETWEEN THE ALGORITHMS INCLUDED IN THE EXPERIMENTAL STUDY. THE TOP TABLE REPORTS THE NUMBER OF SCENARIOS IN WHICH THE ALGORITHM ON THE ROW HAS A BETTER MINIMUM MSE VALUE THAN THAT ON THE COLUMN. IN THE BOTTOM TABLE THE AVERAGE MSE IS USED.

	GA$^+$-MI	GA$^+$-MI-Re	GA$^+$-NC	GA$^+$-NC-Re	SS*	I-ICP	Dyn-GA	GD
GA$^+$-MI	-	2	3	3	13	16	16	16
GA$^+$-MI-Re	14	-	3	3	13	16	16	16
GA$^+$-NC	13	13	-	0	13	16	16	16
GA$^+$-NC-Re	13	13	16	-	13	16	16	16
SS*	3	3	3	3	-	16	13	12
I-ICP	0	0	0	0	0	-	0	12
Dyn-GA	0	0	0	0	3	16	-	12
GD	0	0	0	0	4	4	4	-

	GA$^+$-MI	GA$^+$-MI-Re	GA$^+$-NC	GA$^+$-NC-Re	SS*	I-ICP	Dyn-GA	GD
GA$^+$-MI	-	0	2	0	3	4	3	14
GA$^+$-MI-Re	16	-	10	3	9	10	9	15
GA$^+$-NC	14	6	-	0	4	4	4	16
GA$^+$-NC-Re	16	13	16	-	11	12	11	16
SS*	13	7	12	5	-	16	15	12
I-ICP	12	6	12	4	0	-	0	12
Dyn-GA	13	7	12	5	1	16	-	12
GD	2	1	0	0	4	4	4	-

Figure 3. Visual comparison of the best registration results for the scenarios I_1 versus $T_2(I_3)$ (top) and I_2 versus $T_4(I_4)$ (bottom). From left to right: the original setup, the best registrations obtained by I-ICP, SS* and GA$^+$.



The benefits of using restart is clear in both the boxplots and looking at the MSE values. For both GA$^+$-NC and GA$^+$-MI the average values decrease in all scenarios. Moreover, the frequency in which good solutions were delivered also increased, as we can note by the decrease of third quartiles in the boxplots; this happens in all IR problems for GA$^+$-NC and in all but one case for GA$^+$-MI.

2) Comparison of GA$^+$ with previous methods: The results of GA$^+$ are compared with other state-of-the-art techniques included in the study. In the original comparison, SS* emerged as the best algorithm overall, so we expect it to be the strongest opponent for our methods.

We begin by considering the minimum MSE values. We already noted that the difference among GA$^+$ variants are minimal (less than 1), with GA$^+$-NC and GA$^+$-NC-Re being the best. All four GA$^+$ variants scored better than I-ICP, Dyn-GA and GD in all 16 IR scenarios. GD has remarkably large error values in most of the cases and really small values in the others. This suggests that the algorithm is actually able to reach a global optimum but often get stuck in a local one. Instead, Dyn-GA found intermediate quality solutions in almost every scenarios. I-ICP delivers a less stable level of quality and the solutions are inferior to that of Dyn-GA. Finally, GA$^+$ variants outperformed SS* in 13 cases and in the remaining ones the difference in MSE is really small (less than 10). All GA$^+$ ranks as the best algorithms in the comparison.

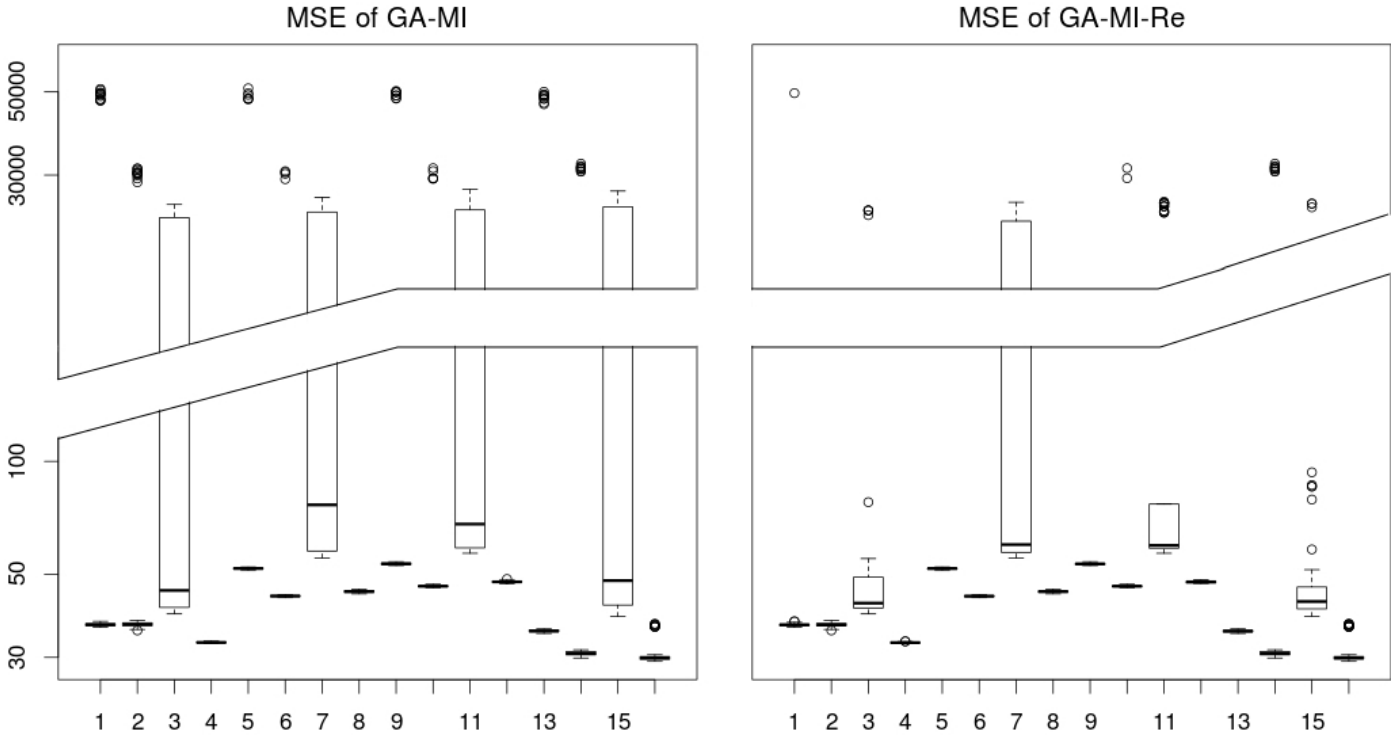
According to the mean MSE values of the algorithms, GA$^+$-NC and GA$^+$-MI are outperformed by all other methods but GD. When we consider the restart version of the latter approaches, the picture changes completely. GA$^+$-MI-Re has a smaller MSE value than those of SS*, I-ICP, Dyn-GA and GD in 9, 10, 9 and 15 cases, respectively. For GA$^+$-NC-Re, those values are even better since they increase to 11, 12, 11 and 16. Therefore GA$^+$ restart variants outperformed all other methods in most of the cases. Compared with SS*, the advantage of those methods can be noticeable. For instance, the MSE of SS* is five times higher than that of GA$^+$-NC-Re and GA$^+$-MI-Re on I_2 vs $T_4(I_4)$. However, this difference was extremely large in the scenarios where SS* performed better than GA$^+$. In particular, SS* reaches an average MSE of 82 on I_1 vs $T_3(I_4)$, while GA$^+$-NC-Re and GA$^+$-MI-Re values correspond to 2824 and 5202, respectively. Finally, the good overall performance of GA$^+$ restart variants is complemented by a quite robust behavior. In 8 scenarios the standard deviation of the MSE value of GA$^+$-MI-Re was below 1, and this value improve to 11 for GA$^+$-NC-Re.

VII. Conclusions

IR is the task of aligning two or more images in order to combine the information they contain. It is considered a fundamental task in image processing because it is used in a wide variety of applications from different domains. Despite the efforts of the scientific community over the last decades, there is still need for improvement of current IR techniques. One of the most critical components of these methods is the optimization procedure, which is responsible for the actual search of the best alignment of the images. Metaheuristics and in particular those methods based on evolutionary computation have proven their worth tackling IR as well as other image processing problems, avoiding the typical drawbacks of classical gradient-based numerical optimization techniques. Nevertheless, a number of evolutionary IR contributions in the literature did not follow an appropriate evolutionary design.

In this work, we introduced an intensity-based IR technique based on genetic algorithms, the most prominent family of evolutionary techniques. We used a modern, real-coded design for representing the solutions and to define the genetic operators. The algorithm also combines a multi-resolution strategy, that allows the registration to be performed in multiple stages with increasing complexity, with a restart procedure, that increases the reliability of the algorithm. In addition, we proposed a novel methodology to compare the performance of different feature and intensity-based IR approaches. Such framework allows the undertaking of broad experimental studies in which methods from both classes are considered, that so far have been almost absent in the literature. Using this methodology, we developed an experimental study to compare

Figure 4. Boxplot of the distribution of mean square error values obtained by the two MI variants of GA$^+$.



our technique with other well-known IR methods, both feature and intensity-based. Furthermore, we included in the comparison four algorithms having different optimization procedures which lay their foundations on classic gradient-descent, metaheuristics and evolutionary computation to represent all current approaches to IR. Brain MRI real-world image data has been used to generate 16 inter-patient registration scenarios of different complexity. Factors such as presence of noise and lesions have been represented in the most difficult scenarios. We proved that our genetic approach performs really well and it is competitive with the best IR techniques. Despite some convergence issues that have been properly addressed after the initial investigation stage, our method actually outperformed all other algorithms on the majority of the scenarios.

REFERENCES

[1] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image Vision Comput.*, vol. 21, pp. 977–1000, 2003.

[2] J. M. Rouet, J. J. Jacq, and C. Roux, "Genetic algorithms for a robust 3-D MR-CT registration," *IEEE T. Inf. Technol. B.*, vol. 4, no. 2, pp. 126–136, 2000.

[3] S. M. Yamany, M. N. Ahmed, and A. A. Farag, "A new genetic-based technique for matching 3D curves and surfaces," *Pattern Recogn.*, vol. 32, pp. 1817–1820, 1999.

[4] M. A. Audette, F. P. Ferrie, and T. M. Peters, "An algorithmic overview of surface registration techniques for medical imaging," *Med. Image Anal.*, vol. 4, no. 3, pp. 201–217, 2000.

[5] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE T. Med. Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[6] F. Maes, D. Vandermeulen, and P. Suetens, "Comparative evaluation of multiresolution optimization strategies for image registration by maximization of mutual information," *Med. Image Anal.*, vol. 3, no. 4, pp. 373–386, 1999.

[7] Y. Liu, "Improving ICP with easy implementation for free form surface matching," *Pattern Recogn.*, vol. 37, no. 2, pp. 211–226, 2004.

[8] F. Glover, "Heuristic for integer programming using surrogate constraints," *Decision Sci.*, vol. 8, pp. 156–166, 1977.

[9] O. Cordón, S. Damas, J. Santamaría, and R. Martí, "Scatter search for the point-matching problem in 3d image registration," *INFORMS Journal on Computing*, vol. 20, no. 1, pp. 55–68, 2008.

[10] O. Monga, R. Deriche, G. Malandain, and J. P. Cocquerez, "Recursive filtering and edge tracking: two primary tools for 3D edge detection," *Image Vision Comput.*, vol. 9, no. 4, pp. 203–214, 1991.

[11] J. P. Thirion and A. Gourdon, "Computing the differential characteristics of isointensity surfaces," *Comput. Vis. Image Underst.*, vol. 61, no. 2, pp. 190–202, 1995.

[12] C. K. Chow, H. T. Tsui, and T. Lee, "Surface registration using a dynamic genetic algorithm," *Pattern Recogn.*, vol. 37, pp. 105–117, 2004.

[13] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[14] S. Klein, J. Pluim, M. Staring, and M. Viergever, "Adaptive stochastic gradient descent optimisation for image registration," *International Journal of Computer Vision*, vol. 81, pp. 227–239, 2009.

[15] L. J. Eshelman, "Real-coded genetic algorithms and interval schemata," in *Foundations of Genetic Algorithms 2*, L. D. Whitley, Ed. San Mateo, USA: Morgan Kaufmann, 1993, pp. 187–202.

[16] T. Bäck, D. B. Fogel, and Z. Michalewicz, *Handbook of Evolutionary Computation*. IOP Publishing Ltd and Oxford University Press, 1997.

[17] D. L. Collins, A. P. Zijdenbos, V. Kollkian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE T. Med. Imaging*, vol. 17, pp. 463–468, 1998.

[18] P. Rogelj and S. Kovacic, "Validation of a Non-Rigid Registration Algorithm for Multimodal Data," in *SPIE in Medical Imaging*, M. Sonka and J. M. Fitzpatrick, Eds., 2002, pp. 299–307.

[19] O. Monga, S. Benayoun, and O. Faugeras, "From partial derivatives of 3-D density images to ridges lines," in *Computer Vision and Pattern Recognition*. Champaign, Illinois, USA: IEEE, 1992, pp. 354–389.