# Evolutionary intensity-based medical image registration: a review

Andrea Valsecchi - European Centre for Soft Computing, Mieres, Spain –
andrea.valsecchi@softocomputing.es

Sergio Damas - European Centre for Soft Computing, Mieres, Spain –
Sergio.damas@softcomputing.es

José Santamaria – University of Jaén, Jaén, Spain – jslopez@ujaen.es

**Abstract**

Metaheuristics are techniques that use approximate and intuitive strategies to quickly find near-optimal solutions of complex optimization problems. A number of outstanding examples belong to evolutionary computation, the class of methods inspired to biological and evolutionary phenomena. These techniques have been extensively and successfully applied to feature-based image registration in medicine. However, with the increase in computational power during the last decade, intensity-based (or voxel-based) image registration methods have been preferred in many medical imaging applications, due to their robustness, accuracy and applicability, in cases where landmarks or other features are not available or easy to detect. While traditional numerical optimization techniques are employed to solve the registration problem, a number of contributions in the literature support the use of metaheuristics to overcome the shortcomings of classic methods. The aim of the paper is to review the state of the art in the application of evolutionary computation and other metaheuristics to intensity-based medical image registration. The study considers both well-know techniques with a large number of references in the literature as well as recent, outstanding proposals. The analysis focuses on the design of the methods to highlight common and successful practices. In addition, recommendations and open research lines in the field are provided.

## 1. Introduction

In its most general formulation, image registration [1] is the task of aligning two or more images in order to establish a spatial correspondence of their common content. Such images usually have the same or a similar subject but have been acquired under different conditions, such as time and viewpoint, or by multiple sensors. In medical image analysis, IR is a key technology that allow to ``fuse'' visual information from different sources [2]. Applications include combining images of the same subject from different modalities, detecting changes before/after treatment, aligning temporal sequences of images to compensate for motion between scans, image guidance during interventions and aligning images from multiple subjects in cohort studies. The remarkable developments in medical imaging technology over the last decades determines a constant demand for better image processing and analysis techniques. Dealing with novel, more diverse, and increasingly accurate sources of imaging data is the main challenge in IR and it explains why it is still a very active research field.

The alignment between two images is specified as a spatial transformation, mapping the content of one image to the corresponding area of the other. A popular strategy among IR methods is to perform the alignment based on only salient and distinctive parts of the image, such as lines, corners and contours, called *features*, ignoring the rest of the image content. This approach, called *feature-based* [2], has the advantage of greatly reducing the complexity of the problem, but relies on the ability to precisely detect the features, either manually or automatically. Any error during the feature extraction stage will propagate into the registration and can hardly be recovered at a later stage. Moreover, this approach is limited to the cases in which features provide enough information to characterize the image content. To avoid these drawbacks, it is possible to use the image intensities directly without any feature extraction, an approach called *intensity-based* (or voxel-based). Though more expensive in computational terms, intensity-based methods achieve the superior

level of accuracy and robustness demanded by medical applications, and over the last decade they have become the method of choice in medical image analysis.

Regardless of the division, the core of every IR technique is an *optimization process* that explores the space of geometrical transformations. Two strategies are available. In *parameters*-based approaches the search is directly performed in the space of the transformation parameters. Hence, a solution is a vector of values for the parameters of the registration transformation. In *matching*-based approaches, features or regions of the image are matched through a search in the space of possible correspondences; once a suitable matching has been found, the transformation parameters are derived accordingly through numerical methods. In both cases the search is guided by a *similarity metric*, a function that measures the degree of resemblance between the input images after the alignment. This can be done either by comparing the whole images or just their corresponding features. Traditional parameters-based methods use classic numerical optimization algorithms, while matching-based methods use matching algorithms like iterative closest point (ICP) [3].

Many characteristics of the IR problem, such as noise, discretization and large differences in the order of magnitude of the transformation parameters still pose a challenge to traditional optimization methods. A number of alternative approaches based on evolutionary computation (EC) [4] and other *metaheuristics* (MH) [5] are often used to deal with complex real-world problems in computer vision and image processing. As for feature-based IR, approaches using metaheuristics have demonstrated to be promising solutions to overcome the drawbacks of traditional optimization algorithms in medical applications, as shown in a recent review [6]. This contribution, instead, focuses on intensity-based methods, for which no such study has been performed. We aim to review the state-of-the-art of the IR methods that lay their foundations on EC and other MHs, including the most established and relevant works. The design of the optimization component will be the central point of our analysis, which will highlight the most relevant components of each method and those that made them successful. This will allow us to provide recommendations for the design of new methods as well as to point out open research lines in the field.

The paper is organized as follows. Section 2 provides preliminaries on medical image registration, metaheuristics and evolutionary computation. The latter part includes a general overview, as well as a description of common techniques and their adaptation to IR. Section 3 reviews outstanding contributions in medical IR using MH and EC, while in section 4 trends and other patterns in the design of the methods are discussed. Finally, section 5 provides conclusions.

## 2. Image registration

This section is devoted to medical image registration based on metaheuristics. In section 2.1 we define the IR problem and its main components. Section 2.2 reviews applications of IR in medicine, while section 2.3 introduces metaheuristics and their adaptation to IR.

## 2.1 Definition and components

A typical IR problem involves two images, conventionally called *model* ($I_M$) and ($I_S$), with different roles in the registration process. The model is the reference (or target) image, while the scene is the image that is transformed to reach the geometry of the other. The registration aims to find a geometric transformation $T$ that aligns the scene to the model; in other words, $T$ is such that the model $I_M$ and the transformed scene $T(I_S)$ are as similar as possible. Image registration can be stated as a minimization problem, i.e.

$$\text{argmin}_T \text{ Similarity}(I_M, T(I_S))$$

A number of components characterize an IR method, but the main ones are just three: the kind of transformation used to relate the images, the similarity metric that measures the quality of the alignment and the optimization procedure that perform the search for a suitable transformation.

### 2.1.1 Transformation model

The transformation model determines which kind of geometrical transformation can be applied to the scene image to reach the model. This also controls which geometrical properties (e.g. size, shape, position, orientation, etc.) are preserved through the process. Common models include rigid transform, which allows translations and rotations, similarity transform, which also admits scaling, and affine transformation, which can also represent shearing. These are examples of global transformations having respectively 6, 7 and 12 degrees of freedom for 3D images. At the other end of the spectrum there are *non-rigid* (also called elastic) transformations, such as B-spline and thin-plate splines transformations, able to represent local deformations (warping) using hundreds or even thousands of parameters.
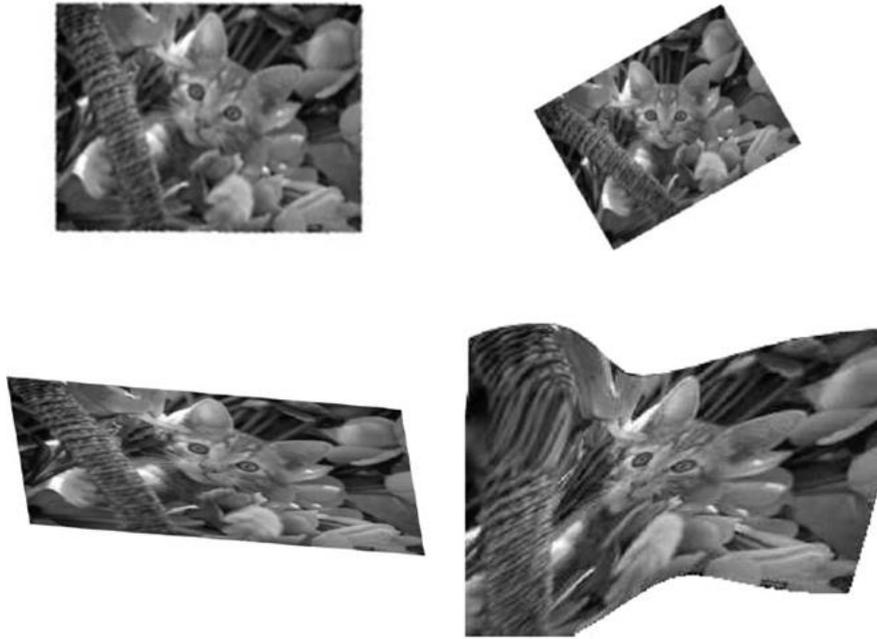
Figure 1. Images obtained from the same scene (top left) by applying different transformations: similarity (top right), affine (bottom left) and B-spline (bottom right).

The choice of the transformation model depends entirely on the needs of the application at hand. A simple 2D translation transform can be enough in certain contexts such as remote sensing, while registering images subject to respiratory motion requires a deformable model. On the other hand, a too flexible transformation model is not just more complex and computationally expensive to apply, but can also lead to undesired or anatomically implausible results, such as bones being bent or tissues growing at an unrealistic rate.

### 2.1.2 Similarity metric

A similarity metric is a function $F(I_A, I_B)$ that measures the resemblance of two images. The quality of a transformation $T$ is assessed by computing the similarity metric over the model $I_M$ and the transformed scene $T(I_S)$. The actual evaluation mechanism depends on the nature of the registration approach. In feature-based methods the similarity metric usually measures the distance between corresponding features [7]. For instance, if the features are points, the alignment can be evaluated using the mean square error (MSE) between the position of a point in the model and that of the corresponding (or closest) point in the transformed scene, i.e.

$$\text{MSE} = \frac{1}{r}\sum_{i=1}^{r} \| x_i - c(T(x_i)) \|^2$$

In intensity-based approaches, similarity metrics are usually based on the resemblance of the intensity values in the two images. The subject of the images along with their modality determine what kind of the relationship is established between their intensity

distributions. For instance, if we assume this relationship is linear, we can assess the similarity between the images by computing the linear correlation coefficient; this is the approach of the similarity metric called normalized correlation (NC), which is defined as

$$NC(I_A, I_B) = \frac{\displaystyle\sum_{x\in\Omega}\left(I_A(x) - \overline{I_A}\right)\left(I_B(x) - \overline{I_B}\right)}{\displaystyle\sum_{x\in\Omega}\left(I_A(x) - \overline{I_A}\right)\sum_{x\in\Omega}\left(I_B(x) - \overline{I_B}\right)}$$

where $\overline{I_A}, \overline{I_B}$ are the average intensity value of the images and $\Omega$ is the common part of their domain. When two images have been acquired using different sensors, a scenario called *multi-modal* registration, the relationship between the intensity values in the images can be strongly non-linear. Metrics based on information theory, such as mutual information (MI), are better suited for this scenario. MI is defined as

$$MI(I_A, I_B) = \sum_{a\in I_A}\sum_{b\in I_B} p_{AB}(a,b)\log\frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)}$$

where $p_{AB}$ and $p_A, p_B$ are, respectively, the joint and marginal probability distributions of the intensity values of the images.

The final performance of any IR method depends on the accurate estimation of the alignment of the images, therefore the similarity metric is considered a crucial component [8].

### 2.1.3 Optimization procedure
The third main component of an IR method is the *optimizer*. It is responsible for finding the best transformation, in terms of similarity metric, among the transformations in our transformation model. Each optimizer has a different search strategy, which depends also on the nature of the algorithm. One approach is to perform the search directly in the space of the transformation parameters. This turns the registration in a continuous optimization problem, therefore classic numerical optimization algorithms can be used. Gradient descent, Newton's method, Powell's method and discrete optimization [9] are among the most common choices along with approaches based on evolutionary computation (EC) and other metaheuristics [10]. IR algorithms that follow this approach are called *parameter-based*. An alternative approach consists in searching for a matching between features, in feature-based methods, or areas of the image, in intensity-based ones. From the match, one can derive the parameters of the corresponding transformation using least squares estimation or other more robust model fitting techniques [13]. This class of algorithms is

called *matching-based*; the iterative closest point algorithm (ICP) is a notorious example of this second approach [14].

### 2.1.4 The complete IR procedure

IR is usually an iterative process. The optimizer computes a candidate transformation, which is then applied to the scene image. The similarity metric compares the model with the transformed scene image and returns a quality value that is sent back to the optimizer. Figure 2 shows a flow chart of the process. The loop ends when a suitable transformation has been found or the algorithm has performed a certain number of iterations.
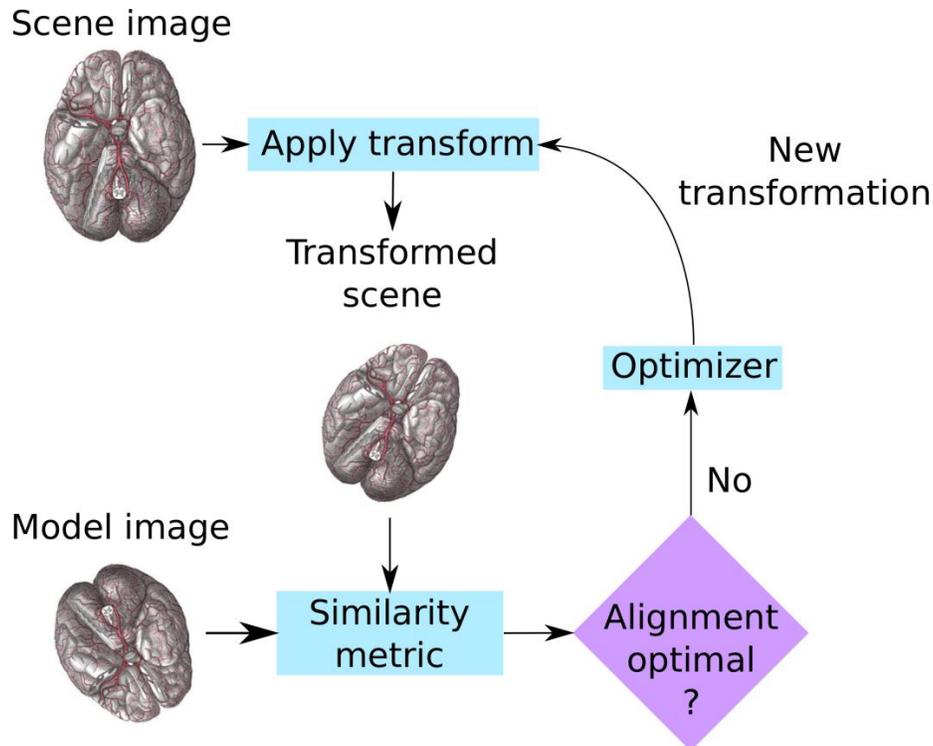


Figure 2. The interactions among the components of a registration technique.

A number of additional minor components are involved in the registration. For instance, after a transformation has been applied to an image, interpolation is used to compute the new intensity values. We mention two of these components that play a special role in the design of the optimization procedure. In intensity-based methods, computing the similarity metric on the whole image is usually unfeasible and unnecessary, therefore a *sampling strategy* determines how many and which voxels are actually used. Those are usually selected at random with uniform probability or sampled along a regular grid.

Second, following a hierarchical approach, it is common to perform the registration in multiple stages. The registration begins using a down-scaled, blurred version of the input images, so that small local optima are removed and the computation of the transformation is faster. Once the optimizer has found a suitable solution, the registration process enters the next stage. Part of the details of the input images is restored and the optimizer aims to adapt the solution of the previous phase to fit the new, more detailed data. The *multi-*

*resolution strategy* determines which kind of processing is performed on the images in each different stage of the registration; usually the procedure includes down-sampling and smoothing. The sequence of images used during the registration is called *pyramid*.

## 2.2 Applications in medical imaging

This section presents a taxonomy of IR scenarios that will illustrate the challenges and diversity of IR applications, in particular in medical imaging. Usually, image registration applications fall in one of the following four groups according to the way in which images have been acquired.

- In *multi-view registration*, images of the same scene are acquired from different viewpoints in order to gain a larger 2D view or a 3D representation of the scanned scene. Examples of applications are surgery planning and 3D object reconstruction.

- In *multi-temporal registration*, images of the same scene are acquired at different times, often on regular basis, with the aim of detecting the changes in the scene that occurred between consecutive acquisitions. This is the case of motion tracking in cardiovascular malfunctions and monitoring of tumor evolution in medical imaging.

- In *multi-modal registration*, images of the same scene are acquired by different sensors. The aim is to integrate the information obtained from different sources to gain a more complex and detailed representation of the scene. This approach is widely used in medical imaging, in which images from sensors recording the anatomical body structure such as MRI, CT or ultrasound are combined with images from sensors that monitor functional and metabolic activities like PET, SPECT or MRS.

- In *scene to model registration*, images of a scene and a model of the scene are registered. The model can be a computer representation of the scene, possibly with a different dimension, or a scene with a similar or canonical content. The aim is either to localize the acquired scene into the model or to compare them. The latter is usually the case in medical imaging, in which a patient's image is compare with digital atlases or ``average'' specimens.

## 2.3 Evolutionary and metaheuristics-based image registration

### 2.3.1 Metaheuristics and evolutionary computation

This section provides a general introduction to metaheuristics and evolutionary computation. An extensive survey of the matter is beyond the scope of this contribution; interested readers can find a large amount of literature reviewing the field [4].

A metaheuristic is an approximate and usually non-deterministic strategy that guides a search process, with the aim of efficiently explore a search space in order to find optimal solutions. MHs are among the most prominent and successful techniques to solve a large amount of complex and computationally hard combinatorial and numerical optimization problems arising in human activities, such as economics (e.g., portfolio selection), industry (e.g., scheduling or logistics), or engineering (e.g., routing), among many others. MHs can be seen as general algorithmic frameworks that require relatively few modifications to be

adapted to tackle a specific problem. They are a diverse group of optimization algorithms that includes simulated annealing (SA), tabu search (TS), multi-start methods, iterated local search (ILS), greedy randomized adaptive search procedures (GRASP), memetic algorithms (MAs) and scatter search (SS).

Evolutionary Computation (EC) is a remarkable family of MHs drawing their inspiration from nature, with particular emphasis on evolutionary models of computation. The field includes evolutionary algorithms (EAs) [4] such as genetic algorithms (GAs), evolution strategies (ES), genetic programming (GP), differential evolution (DE) and evolutionary programming (EP). In particular, GAs are probably the most used EA in the literature to face real-world optimization problems. Another important branch of EC is swarm intelligence (SI), the area of artificial intelligence that concerns the collective behavior of decentralized, self-organized systems, either natural or artificial. This includes ant colony optimization (ACO) and particle swarm optimization (PSO).

Nowadays, MHs have become an interdisciplinary research area intertwining disciplines such as computer science, operations research, engineering, etc. They have received enormous attention as witnessed by thousands of journal and conference papers, hundreds of authored and edited books published, and a large number of dedicated conference series.

### 2.3.2 Common metaheuristics for Image Registration

In this section we describe in detail the three metaheuristics used in most of the methods under review, namely genetic algorithms, particle swarm optimization and simulated annealing. These are also good representatives of different branches of metaheuristics, therefore this section also gives a flavor of MHs and the diversity of this kind of approaches.

*Simulated Annealing*

The name and inspiration of simulated annealing come from annealing in metallurgy, a technique involving heating and controlled cooling of a material. SA is similar to a hill-climbing algorithm. At each iteration, the algorithm considers a solution $r$ in the neighborhood of the current one $s$ and computes its quality. Just like in hill-climbing, if $r$ has a higher quality than $s$, $r$ becomes the current solution and the iteration ends. Otherwise, if $r$ is worse then $s$, $r$ is accepted with probability

$$e^{\frac{\text{quality}(r)-\text{quality}(s)}{t}}$$

The ability to accept worse solutions allows the algorithm to escape local minima and perform a global optimization. The temperature parameter $t$ controls the probability of accepting a worse solution, which also depends on the decrease in quality: the worse the new solution, the lower is the probability of its acceptance.

The temperature is initially set to a high value and then decreased at each iteration, according to a *schedule*. At high temperatures, a solution is likely to be accepted regardless of its quality, and the behavior of the algorithm is similar to that of random search. At

temperatures close to zero, instead, worse solutions are rarely accepted, so simulated annealing works like a hill-climbing algorithm. The slower the temperature is decreased, the more the algorithm explores the search space.

**Simulated annealing - pseudocode**

```
t = initial temperature
s = initial solution
repeat
    r = neighbor(s)
    p = random(0,1)
    if quality(r) > quality(s) or
    p < e^((quality(r)-quality(s))/t) then
        s = r
    decrease t
until stop condition
return s
```

### Genetic Algorithms

Like most algorithms in EC, genetic algorithms mimics the processes observed in natural evolution. A solution of the optimization problem is though as an *individual* and is represented as a string of values called *chromosome*. The quality of a solution becomes its likelihood of survival or *fitness*. The algorithm considers multiple solutions at the same time, organized in a set called *population*, which is *evolved* (improved) in an iterative process. First, the fittest individuals in the population are randomly selected for reproduction. Their chromosomes are recombined and undergo minor random variations; the resulting individuals are placed in a new population, which replaces the current one at the end of the iteration, forming a new *generation*.

**Genetic algorithm - pseudocode**

```
generate a random population
repeat
    evaluate the fitness of each individual
    repeat
        select two parents from the population
        create two offsprings by combining the chromosomes of the parents
        apply a small mutation to the chromosomes of the offspings
        add the offsprings to the new population
    until the new population is full
    replace the population with the new one
until stop condition
```

The term *crossover* is used for the operation of generating new individuals by combining chromosomes of other individuals, while *mutation* refers to applying a small variation to the chromosome of an individual. Other important components of a GA are those that create the initial population and select the individuals for reproduction. All these four operations are inherently stochastic, and both crossover and mutation are applied with certain probabilities.

The design of the components of a GA depends on the optimization problem and the encoding of a solution. For instance, consider an optimization problem over the set of integers $\{0, \dots, 255\}$. To represent a solution as a string, one could use the binary representation of the integer value. Here is an possible design of a GA to solve this problem. The chromosome is a binary string of length 8; the initial population is generated at random with uniform probability over $\{0,1\}^8$; parents are selected by sampling five elements from the population and taking the best and second best individuals in the sample; mutation flips a random bit; crossover choose a random position $i$ in the chromosomes and copies the bits up to $i$ from the first parent and those after $i$ from the second one.

## Particle Swarm Optimization

PSO is a population-based algorithm, but rather than a population of biological organisms, this group of solutions is though as a *swarm of particles* in the search space. For simplicity, assume we are dealing with a continuous optimization problem. The position of a particle is what actually encodes a solution of the optimization problem, while the velocity of a particle depends on the quality of the associated solution and on the interaction with the other particles. Through the movement of the particles the algorithm examines new solutions and the optimization is performed.

In a typical PSO algorithm, the particles are initially generated at random with uniform probability over the whole search space. Then, an iterative process begins. First, the quality of the solutions associated with particles is computed. Then, velocity and position of the particles are updated. The position of a particle $i$, $x_i$, is updated using the rule $x_i = x_i + v_i$, just like an object moving at constant speed $v_i$ for a unit time interval. The velocity update rule is

$$v_i = wv_i + c_1 r_1 (p_i - x_i) + c_2 r_2 (g_i - x_i)$$

where $w, c_1, c_2$ are called respectively the *inertia weight* and the *acceleration coefficients*, while $r_1, r_2$ are random vectors with uniform distribution over $[0,1]^n$. The term $p_i$ is the *personal* best position, meaning the best position found by the particle $i$ since the beginning of the algorithm, while $g_i$ is the *global* best known position, considering current and past positions of all the particles in the swarm.

The velocity update rule is what governs the evolution of the swarms, which is more complex than that of a physical system. In addition to the inertia component, indeed, there is a *memory* component, which moves a particle toward its best past position, and also a *social* component, based on the global best position shared across the whole swarm.

## Particle swarm optimization - pseudocode

```
generate a random population of particles
repeat
    evaluate the quality of the particles
    update global and personal best positions
```

```
    update the velocity of the particles
    update the position of the particles
until stop condition
```

### 2.3.3 Suitability of Evolutionary Computation and other Metaheuristics in Image Registration

There are different strengths and limitations that have been stated either to justify or to avoid the use of these methods when tackling complex optimization problems like IR. Some of the advantages are:

- Unlike classical gradient-based search methods, those based on EC and other MHs do not depend on the starting solution, thus being more robust approaches. Moreover, they provide specific strategies to escape from local optima. In particular, they can cope with multimodal functions to tackle IR [20].

- EC and MHs have been used in a wide variety of optimization tasks within IR including numerical optimization and combinatorial optimization problems, i.e. facing both the transformation parameters and the matching-based IR approaches, respectively.

- They are conceptually simple and easy to implement.

- They can handle arbitrary kinds of constraints and objectives easily.

- Unlike other numerical IR techniques (e.g. gradient-based) that are only applicable for continuous functions or other constrained sets, their performance is independent of the solution representation.

- They offer a framework wherein including prior knowledge about the problem is easy. Thus, the search process is more appropriate, yielding a more efficient exploration of the space of possible solutions. For instance, feature-based IR approaches [21] improved the design of the objective function to exploit information related to the geometry the images.

- They can also be easily combined with more traditional optimization techniques such as gradient-based methods [24]. An outstanding approach to exploit the benefits of both strategies is their hybridization in the well-known memetic computation paradigm [26]. Such scheme was successfully applied to the IR problem in [25]. Currently, this hybrid approach brings an outstanding performance due to the proper combination of the exploration and the exploitation capabilities of both stochastic and deterministic optimization schemes.

The most important shortcomings related to the use of EC and other MHs are shown as follows:

- They require a tuning of the control parameters, which is often a manual, error-prone, expert-based procedure. Approaches based on automatic parameter tuning [28] or MHs with an adaptive behavior [26, 61, 62] have been recently introduced to solve this problem.

- Typically, EC and MHs are time consuming, therefore they are usually avoided in real-time applications. Parallel and GPU implementations are increasingly more common [29].

- Some MHs lack a formal proof of convergence to the global optimum and there is hardly any theoretical result on the performance of MH. However, there is a very large amount of empirical results to support effectiveness.

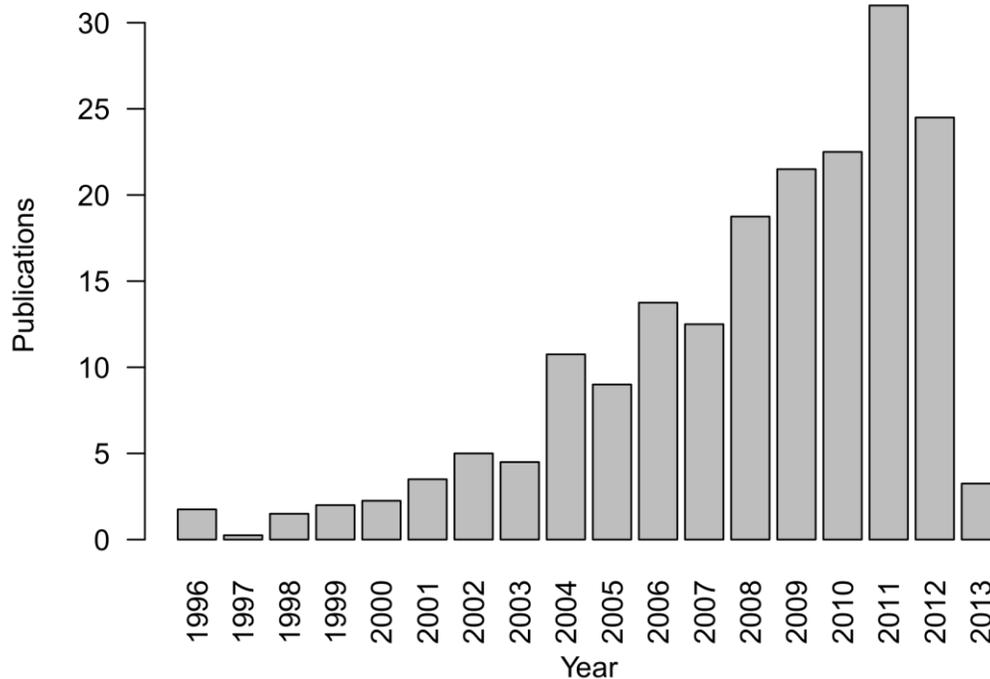### 2.3.4 Early evolutionary image registration methods



Figure 3. The number of publications in proposing MHs and EC to solve intensity-based medical image registration.[1]

The application of MHs and EC to intensity-based medical IR enjoyed a growing interest in the scientific community over the last fifteen years, as shown in Figure 3. The first attempts at IR can be found in the early eighties. The size of data as well as the number of parameters that are looked for prevent from an exhaustive search of the solutions. An approach based on a GA was proposed in 1984 for the 2D case and applied to angiographic images [30]. Later, in 1989, Mandava et al. [31] used a 64-bit structure to represent a possible solution when trying to find the eight parameters of a bilinear transformation through a binary GA. Brunnström and Stoddart [21] proposed a new method based on the

---

[1] The data for the graph was obtained from Thomson Reuter's Web of Science using the query (Title OR Abstract OR Keywords) = ("image registration" OR "image alignment" OR "image matching") AND ("evolutionary algorithm" OR "evolutionary computation" OR "genetic programming" OR "genetic algorithm" OR "evolutionary programming" OR "evolution strategy" OR "differential evolution" OR "swarm intelligence" OR "particle swarm" OR "ant colony" OR "bee colony" OR "scatter search" OR "tabu search" OR "simulated annealing" OR "grasp"). The literature corpus was then manually filtered to include only contributions in medical imaging using the intensity-based approach.

manual prealignment of range images followed by an automatic IR process using a novel GA that searches for solutions following the matching-based approach. Tsang [32] used 48-bit chromosomes to encode three test points as a base for the estimation of the 2D affine registration function by means of a binary-coded GA. In the case of Yamany et al. [11] and Chalermwat et al. [12] proposals, the same binary coding is found when dealing with 3D and 2D rigid transformations, respectively. Yamany et al. enforced a range of $\pm 31°$ over the angles of rotation and $\pm 127$ units in displacement by defining a 42-bit chromosome with eight bits for each translation parameter and six bits for each rotation angle. Meanwhile, Chalermwat et al. used twelve bits for the coding of the 2D rotation parameter to get a search scope of $\pm 20.48°$, therefore allowing the use of a precision factor for the discretization of the continuous rotation angle interval. Other ten bits stored each of the two translation parameters ($\pm 512$ pixels).

All the latter approaches showed several pitfalls from an EC perspective. On the one hand, they make use of the basic binary coding to solve inherently real coded problems, when it is well known that binary coding suffers from discretization flaws (as problem solutions of search space never visited) and requires transformations to real values for each solution evaluation. Moreover, the kind of GA considered is usually based on the old-fashioned original proposal by Holland [33]. In this way, a selection strategy based on fitness-proportionate selection probability assignment and the stochastic sampling with replacement, as well as the classical one-point crossover and simple bit flipping mutation, are used. On the one hand, it is well known that such selection strategy causes a strong selective pressure, thus having a high risk of premature convergence of the algorithm. On the other hand, it has also been demonstrated that it is difficult for the single-point crossover to create useful descendants as it is excessively disruptive with respect to the building blocks.

## 3. State of the art in intensity-based medical IR using EC and MHs

This section presents a selection of well-established and relevant intensity-based medical IR methods that use EC and MHs. We included in the study the most cited contributions in the literature, and in addition, we chose a number of outstanding proposals published between 2011 and 2013 that, being introduced only recently, have not yet reached a high number of citations.

### 3.1 Well-established proposals

#### Wachowiak et al.'s PSO

The proposal [34] tackles the registration of 2D slices to 3D volumes having different modality. The authors introduce a number of alternative designs for the optimization component based on PSO. The proposed modifications make use of the following components:

- the local best position instead of the global best position, meaning that the particles are attracted to the best particle in their neighborhood rather than in the whole swarm. This can prevent premature convergence, but also makes the optimization more susceptible to local optima;

- a crossover operator, which combines linearly both the position and the velocity of two particles;

- a constriction coefficient on the velocity of the particles, using the popular approach proposed in [35], trough which the amplitude of a particle's oscillation decreases over time and allow its convergence

- subpopulations, i.e. the particles are clustered in different groups, and the crossover operator is applied to particles in the same group or in different groups according to a probability $p$

- the use of Powell's direction set algorithm, a numeric optimization technique, to improve the final solution provided by PSO, as well as to refine the initial solution.

Furthermore, all the algorithm versions are designed to exploit an initial solution supplied by the user. This provides an extra term in the velocity of the particles, in addition to those based on the global or local best, the personal best and the current position.

A total of eight PSO variants are tested, and the authors compare the results with those obtained by Evolution Strategy with seven different configurations. The experimental study involves the registration of abdomen and head images across several modalities (histological images, ultrasound, CT, T1 and T2 MRI). Rigid registration model and normalized mutual information are used. Three PSO variants delivered a promising performance, however, the quality of the initial solution has a very pronounced effect on the final result. This limits the robustness of the algorithm even in the cases in which an initial solution is a available.

### Kagadis et al.'s GA

In [36], the authors propose a method for the registration of intra-subject SPECT and CT brain images of patients suffering from ischemia or hemorrhage. The optimization is performed by real-coded GA with two distinctive traits. First, the mutation probability is not global, but it is specific of each individual. The probability is encoded in the individual's representation and it undergoes the variation operators just like the rest of the chromosome. Second, after the GA has terminated, the best individual is improved using Powell's method, and the outcome of the process is the final solution provided by the algorithm.

In the experimental study, the GA is used as an optimizer for an intensity-based method, as well as for a feature-based one. The latter register the brain outer surface, whose extraction requires a substantial preprocessing step. Both algorithms use affine transformation and, respectively, mutual information and mean square error as similarity metric. The ICP method is also included in the comparison as a reference. The results show that the intensity-based method is significantly more accurate and, unlike its competitors, it does not require a fine-tuned feature extraction stage to work properly.

### Wang et al.'s GA

In [37], a memetic approach combines a GA with the Nelder-Mead simplex method. In addition to the simplex, the authors introduce the use of a novel immigration operator, which replaces the worse $n$ individuals in the populations with random solutions.

The proposed GA is designed around the concept of earlier/latter stage of evolution. The current stage is determined by the difference in fitness between individuals: large differences mean the population is far from convergence, while small differences are a sign that the population is about to converge. At the end of an iteration, if the population is in the earlier stage, the Nelder-Mead simplex is applied to the individuals before being moved into the next population. Otherwise, the immigration operator is used. This operation is meant to prevent the algorithm from converging prematurely.

In the experimental study, the proposed GA is tested over the task of registering 2D brain MRI with CT using similarity transformations. For comparison, a regular GA without simplex and immigration is also tested. The novel method delivered moderately better registration results, but the comparison appears to be unfair. The additional computation effort of the simplex was not taken into account, and the novel GA was run for almost twice as much time as the regular GA. It is thus unclear whether this improvement is due to the new operators or just the larger running time. There is also a lack of detail about the GA, and crucial information such as the coding scheme are missing.

### Xu and Dony's DE

In [38], the authors present an IR method using multiple resolutions. At the first, coarse resolution, a DE algorithm is used, while in the further resolutions the optimization is carried out by Powell's direction set method. This approach combines the ability of global optimization of the DE with the fast, local optimization provided by Powell's method.

Another novelty of the proposal is the similarity metric, which combines normalized MI with *gradient information* [39]. The second quantity measures the similarity of the gradient vectors between the model and the transformed scene images. For each point $x$, the length and the angle $\alpha$ between the gradients in the two images is computed. The gradient information between the images $A, B$ is the sum

$$G(A, B) = \sum_{x \in \Omega} w\,(\alpha_x) min(\mid \nabla A(x) \mid, \mid \nabla B(x) \mid)$$

where $w$ is a weighting function that favors angles that are either very small or close to $\pi$. The final similarity metric of the algorithm is the product between the gradient information and the normalized MI.

The experiments are run on 2D brain MRIs that have been transformed using known rigid transformations. By comparing the results those obtained using Powell's method alone, the authors can highlight the benefits of using DE in the first resolution. Registration scenarios are created applying increasing larger rotations and translations, and it is shown that the

performance of Powell's method drops significantly as the magnitude of the transformation increases. In the proposed method, instead, DE is able to recover the initial transformation with enough precision that in the further resolutions Powell's method can then converge to a high quality solution. Quantitatively, the ratio of suitable registrations found by the full method stays above 90% in all scenarios, while that of Powell's method alone drops below 20% when large rotations are applied to the images.

### Talbi and Batouche's PSO

In [40], a hybrid PSO / DE algorithm is proposed. A standard, local-best PSO design is augmented with a mutation operator inspired by DE. The key idea is to use the difference between two individuals in the population to disrupt another one. What is actually used is the difference between the personal best position of two randomly-chosen particles, called $\delta$. Each component of a solution undergoes mutation with a fixed probability. When a mutation occurs, the $i$-th position of a solution $x$ is replaced by $p_{g,i} + \delta$, where $p_g$ is the best neighbor solution of $x$. The use of $p_g$ provides the social learning capability that speeds up the convergence. In the proposed algorithm, regular particle evolution is alternated with applying mutation to all particles.

The algorithm was tested on rigid registration of 2D bran MRI with CT using MI. The results of the registration seems visually accurate, but there is no quantitative evaluation other than the similarity metric values. Moreover, there is no comparison with other algorithms or different designs, so it is difficult to assess the performance of the proposal and the effect of the different novel components.

### Winter et al.'s CMA-ES

The proposal in [41] is a surface to volume registration algorithm for pedicle screw insertion during spinal surgery. The surface of the vertebra is extracted from a CT scan and registered to a 3D ultrasound using a rigid 3D transformation. The similarity metric exploits the fact that the tissue-bone interface is the brightest part of the image, so the registration tries to maximize the sum of gray values of the voxels covered by the bone surface.

The proposed optimization method is a covariance adaptation matrix evolution strategy (CMA-ES), which the author turned into a multi start algorithm. The design of the actual CMA-ES component and even the parameters values were taken from a generic setup for multimodal continuous functions [42].

The author carried out a thorough experimental study comparing CMA-ES with nonlinear conjugate gradient, the Broyden-Fletcher-Goldfarb-Shanno algorithm and resilient backpropagation. All algorithms were tested using multiple starting positions to avoid introducing any bias in favor of the proposed algorithm. The authors found CMA-ES to be the best algorithm in the comparison, with large differences in the results of different algorithms. This applies to both the rate of successful registrations and the size of the misalignment after registration. Impressively, the multi start CMA-ES failed to deliver an acceptable registration in just four of 12,000 trials.

### He and Narayana's GA

In [43], the authors combine a real-coded GA with DIRECT, a deterministic, global optimization method based on *branch and bound* [44]. In a two-resolution strategy, the GA is used in the first resolution, while DIRECT is applied in the second. This exploits the ability of the GA to explore a large search space and the features of DIRECT, which is able to find a globally optimal solution in a small search space.

The GA uses arithmetic crossover and a custom mutation operator. A "perturbation" vector $\sigma$ determines the magnitude of the change produced by a mutation. An individual component $x_i$ is replaced with $x_i + y$, where $y$ is sampled from a normal distribution with mean 0 and standard deviation $\sigma_i$. The value of the perturbation vector was set to a fraction of the transformation parameter ranges, whose value was determined experimentally.

The algorithm was tested on registering brain MRI with 3D rigid transformations. The registration is actually multi modal, as different MR setups (echo times, pre/post contrast) were used to acquire the images. Accordingly, mutual information was employed. The results show that the GA is able to converge in the neighborhood of the global solution, and DIRECT is able to locate it more accurately. A comparison with the AIR medical software [45] shows that the proposed method is just slightly more precise when a small transformation is required, but it is greatly more robust when the transformation is larger.

### Castellanos et al.'s GA

The work presented in [46] tackles deformable registration. The authors propose a novel transformation model which is a composition of circular warpings of different size. The warpings are applied in a hierarchically; first, there is a large, central warping that covers the whole image, then the image is divided in four squares and for each of them, a warping of the size of the square is placed at its the center. The process continues in a similar fashion increasing the number of squares, until the desired level of granularity is reached.

Each warping is defined by five real parameters. The optimization focus on one warping at the time, starting from the largest one, so the optimizer is run multiple times and it uses just the part of the image that is below the warping being optimized. The authors used a real-coded GA with two main novelties. First, when the reproduction operator is employed on a solution $x$, the algorithms performs an arithmetic crossover of $x$ with the current best solution $b$. The weight of this linear combination depends on the ratio between the fitness of $x$ and $b$, so that bad solutions are moved toward the current best, while good solutions undergo a negligible change. Second, the mutation rate depends on the current iteration, so it can be raised when the algorithm is approaching convergence.

The authors carried out an experimental study on both synthetic and real medical images, using normalized mutual information as similarity measure. A multi resolution strategy is used, so that during the registration, the smaller the warping, the bigger and more detailed the images. The test are performed over brain MRIs, chest Rx scans and a pair PET-CT of chest. The results show that the combination of an original transformation model and the proposed optimizer is able to deal with deformable registration with visually acceptable results. What the study is missing is a comparison with other, more established approaches, both in terms of transformation model and optimization procedure.

### Li et al.'s SA

In [47], the authors study the repeat radiosurgery of trigeminal neuralgia. This process requires the registration of two brain MRI taken before and after the first radiotheraphy treatment, but only a limited number of slices can be acquired, e.g. 28 in the experimental study. This can be a challenge due to the small volume on which the quality of the registration is assessed.

The registration is rigid and mutual information is used. The proposed algorithm, based on simulated annealing, was able to register the volumes correctly in all 41 cases considered in the study, while Nelder-Mead simplex and Powell's method were unsuccessful in 11 and 9 cases, respectively. The factor that affects the results the most is the initial overlap between the volumes to be registered, and when this amount is small, SA is the only algorithm which is still able to deliver accurate results.

### Chen et al.'s PSO

In [48], a deformable 2D IR algorithm based on PSO is developed. The algorithm uses rigid and B-Spline transformation, while the PSO design is essentially that of [34]. A brief experimental study compares the proposal with gradient descent on two registration of MRI and CT images of liver. When the rotation applied to the images is small, both algorithms deliver accurate results, but if a large rotation is used, PSO is able to recover the pose of the liver, while GD fails. The description of the experiments on deformable registration is not informative enough to make any conclusion, although it shows PSO is able to deliver a visually acceptable result.

### Li and Sato's PSO

The study in [49] compares different PSO designs for multimodal rigid IR. The method aims to be applicable in different fields, so the experimentation includes images of very different subjects and modalities, such as administrative maps, satellite images as well as proton density and MR images of brain. The analysis of the PSO design focuses on the velocity update equation. The study found that the best rule includes an additional term, which is based on the difference between the global best solution $p_g$ and the personal best solution $p$ of the particle, i.e.

$$v = wv + c_1 r_1 (p - x) + c_2 r_2 (p_g - x) + c_3 r_3 (p_g - p)$$

where $w, c_1, c_2, c_3$ are used-defined weights and $r_1, r_2, r_3$ are random vector whose components have uniform distribution over $[0,1]$.

### Loeckx et al.'s SA

The study [50] presents a nonrigid registration algorithm for temporal subtraction of 2D thorax X-rays. The algorithm is rather complex and relies on machine learning to exploit previous knowledge on the problem. First, there is an initialization phase, in which the lung field is segmented using an active contour model approach. This requires a deformable model of the lungs, which is learned through principal component analysis on training data. The segmentation provides the ROI for the actual registration step. The transformation

model is based on B-Splines with nine control points (which very small for the large area covered by the lungs). By using PCA on registered training data, the algorithm can compute the principal components of the transformation. Then, instead of optimizing the parameters of the spline, the algorithm optimizes the coefficients of the principal components.

The similarity metric of choice is called Pattern Intensity (PI), which has been specifically designed to minimize subtraction artifacts. Pattern intensity is similar to MSE, but artifacts smaller than the threshold $r_{max}$ are considered noise. The actual formula is

$$\text{PI}(x) = \sum_{|x-y|<r_{max}} \frac{(I_{diff}(x) - I_{diff}(y))^2}{\sigma^2 + (I_{diff}(x) - I_{diff}(y))^2}$$

where $I_{diff}$ is the difference image and $\sigma$ acts as a sort of threshold over the intensity difference in an artifact, meaning differences smaller than $\sigma$ have a small contribution to PI.

The optimization procedure uses simulated annealing with a two-resolution strategy. The range of the parameters as well as the initial and final temperatures are changed between resolutions, so that in the second resolution the search is more focused around the solution found at the first resolution. The algorithm was tested on 26 pairs of images and in the 85% of the cases the results were rated adequate for clinical use. Although no formal comparison is developed in the paper, the author praise the quality of the optimization obtained through SA and highlight its reproducibility despite the stochastic nature of the method.

### Du et al.'s SA
In [51], the authors introduce a method for multi-modal image registration in noisy scenarios. The algorithm is based on a novel similarity metric called double directional partition intensity uniformity (DRPIU), which is designed to be less affected by outliers and noise through the use of a robust estimator [52].

Two resolution are used. The algorithm uses simulated annealing at the coarse resolution and Powell's method at the fine one. This approach aims to provide Powell's method with a starting point near the global optimum and, at the same time, it does not require a large computation time.

The experimental study focus on the rigid registration of T1, T2 MRIs and PET brain scans. Strong salt-and-pepper and speckle noise is added to the images to highlight the robustness of the method. Compared to MI and PIU, the use of DRPIU yielded much more accurate results and allowed correct registration in all the test scenarios.

## 3.2 Recent contributions (2011-2013)

### Valsecchi et al.'s GA

In [53], the authors designed an IR method based on a genetic algorithm. The GA is real-coded and uses BLX-crossover, random mutation, elitism and tournament selection. The design is flexible enough to support multiple similarity metrics and the transformation models, include deformable ones.

The main novelty of the algorithm is that it combines a two-resolution strategy with a restart and a search space adaptation mechanism. The key idea is that the first resolution is devoted to finding a coarse approximation of the desired transformation, while the second resolution is meant to improve it rather then finding a completely new one. Therefore, at the end of the first resolution, the best solution found $x'$ is tested on quality. If the quality is high enough, the registration continues with the second resolution, otherwise the process restarts and the algorithm performs the first resolution again. In the original proposal [54], the test on $x'$ was a simple comparison with a threshold. However, the authors found that appropriate threshold values were strongly dependent on the actual pair of images, even when the content was similar, therefore they proposed a different policy. The first resolution is repeated a fixed number of times, and at the end of the process the population containing the best solution overall is used for the second resolution. This approach does not guarantee a certain quality level, but the probability of a bad solution reaching the second resolution decreases exponentially at each restart. As the first resolution uses downsampled images, the computational cost of performing a restart is small compared to that of the whole process. In addition, at second resolution the search space is restricted around the transformation parameters values in $x'$. This focuses the search on the area of the search space that is supposed to contain the global optimum.

The algorithm is thoroughly tested on the registration of brain MRIs and an atlas-based segmentation application. The experimental comparison includes other methods based on evolutionary algorithms, both intensity- and feature-based, as well as classic gradient-based technique from a medical toolkit. The GA is able to outperform all competitors in almost all test scenarios, proving its applicability.

### Zhou et al.'s PSO

The study [55] proposes an advanced quantum-behaved PSO algorithm (QPSO) [56]. Regular PSO is not a globally convergence-guaranteed algorithm as the particles are restricted to a finite search space in each iteration, which weakens the global search capability of the algorithm. In QPSO, each individual particle exhibits quantum behavior and it is able to cover the whole search space. The position of a particle is updated using the rule

$$x = q \pm \frac{L}{2} ln\left(\frac{1}{u}\right)$$

where $u \sim$ Unif(0,1), $q$ is the basin of attraction and $L$ regulates the extent of the particle movement. In this proposal, the particles are attracted toward the mean personal best of all particles, named $C$, and $L$ depends on the distance of the particle with respect to its personal best $p$. This way particles far from the swarm have a lager search scope and viceversa. The update equation can be rewritten as

$$x = p \pm \alpha \mid p - x \mid ln(\frac{1}{u}) + \beta \mid C - x \mid r$$

where $r \sim$ Norm(0,1). The parameters $\alpha$ and $\beta$ balance the exploration and exploitation behavior of the algorithm.

Just like regular PSO, QPSO faces the problem of premature convergence due to the fast flow of information between particles. The second novelty of the proposal is that $\alpha$ and $\beta$ are adjusted dynamically, depending on the *diversity* of the swarm, hence the name diversity-controlled QPSO (DQPSO). The algorithm uses the average distance between a particle and the center of the swarm, which is then normalized by dividing by the longest diagonal in the search space. At every iteration, the diversity is computed and a threshold triggers the adaptation of $\alpha$ and $\beta$. In this proposal, both parameters decreased linearly from their initial value to a final value, while the threshold is set to the maximum number of iterations times 0.8. Each time the diversity declines, the decrease in $\alpha, \beta$ triggers an "explosion" of the swarm that allows it to escape local optima. The size of the explosion decreases through time, so that eventually the algorithm focuses on the promising area of the parameter space.

The authors developed an experimental study on classic numeric benchmark functions to fine-tune the algorithm, focusing mainly on the initial and final values for $\alpha$ and $\beta$. In comparison with PSO variants and a few other metaheuristics, DQPSO had better results in terms of precision and it exhibited a much faster convergence.

Last, using the results of the tuning, the algorithm is incorporated in a multi-modal IR approach for brain MRI, CT and PET images. Rigid 3D transformations and mutual information are the other main components. The method is tested over six MRI-CT pairs and six MRI-PET pairs. The results show DQPSO is more precise than regular and quantum PSO, as well as Powell's method. DQPSO is also more robust than its competitor, as it always delivered correct registrations but in one of the scenarios.

## 4. Discussion

The aim of this section is to analyze trends and other patterns that occur in the group of algorithms described in the previous section. A classification is provided in Table 1.

| | Strategy | Optimizers | Transformation Model | Multiple resolutions | Multi modal | Metric | Modalities | Target |
|---|---|---|---|---|---|---|---|---|
| Kagadis 2002 | Parameters | GA, Powell | Affine 3D | No | Yes | MI | CT, SPECT | brain |
| He 2002 | Parameters | GA, DIRECT | Rigid 3D | Yes, 5, gaussian | Yes | MI | MRI | brain |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Loeckx 2003 | Custom | SA | Affine + B-splines 2D | Yes, 2 | No | Patter Intensity | Rx | thorax |
| Wachowiak 2004 | Parameters | PSO, Powell | Rigid 3D | No | Yes | NMI | Histological, MRI, US, CT | brain, abdomen |
| Xu 2004 | Parameters | DE, Powell | Rigid 3D | Yes, 2, wavelet | Yes | MI and Gradient Information | MRI, CT | brain |
| Talbi 2004 | Parameters | PSO | Rigid 2D | No | Yes | MI | MRI, SPECT, CT | brain |
| Castellanos 2004 | Parameters | GA | Comp. of local warpings 2D | Yes, 3, shrinking | Yes | NMI | MRI, Rx, PET, CT | brain, thorax |
| Li 2005 | Parameters | SA | Rigid 3D | No | No | MI | MRI | brain |
| Du 2006 | Parameters | SA, Powell | Rigid 3D | Yes, 2 | Yes | Double directional partition PIU | MRI, PET | brain |
| Li 2007 | Parameters | PSO | Rigid 2D | No | Yes | MI | PD, MRI | brain |
| Winter 2008 | Parameters | CMA-ES | Rigid 3D | No | Yes | Application specific | CT, 3D US | spine |
| Chen 2008 | Parameters | PSO | Rigid 2D | No | Yes | MI | MRI, CT | liver |
| Wang 2011 | Parameters | GA, Nelder-Mead | Similarity 2D | No | Yes | NMI | MRI, CT | brain |
| Zhou 2011 | Parameters | Quantum PSO | Rigid 3D | No | Yes | MI | MRI, CT, PET | brain |
| Valsecchi 2013 | Parameters | GA | Affine 3D | Yes, 2, gaussian | No (yes) | MI (any) | MRI | brain |

Table 1. The characteristics of the methods under review.

We begin by remarking that all algorithms follow the parameter-based strategy. The alternative approach of matching areas of the image is rather complex, therefore those kind of algorithms consider a very specific type of images and use an additional parameter-based registration step, so that corresponding areas are already very close to each other [57].

The next component is the optimization procedure, which is the focus of our study. For the MH part, the most prevailing choices are GAs, PSO and SA, with respectively 5, 5 and 3 contributions over 15. This is not unexpected, as GAs are among the most established algorithms in MHs and they have been applied successfully to IR in the past (check with introduction to avoid repetitions). Also, PSO is the most prominent example of techniques based on swarm intelligence, and contributions in computer vision constitute a very large share of its application bibliography [58]. SA appears less frequently in recent publications, but its applicability in spite of a conceptually simple mechanism still makes it a popular choice. The remaining two algorithms use DE and CMA-ES. It is worth to mention that the results of the algorithm based on CMA-ES ([41]) are quite impressive. However, in a different study [59], CMA-ES delivered one of the worst performance among a large group of IR algorithms despite being granted a larger time limit.

As for the complete optimization process, in six cases a single MH is often combined with a local optimization technique, i.e. Powell's method (four algorithms), Nelder-Mead simplex

(one) and DIRECT (one). This approach exploits the ability of the MH to explore a large search space, while the local optimizer is used to refine a single, high-quality solution in narrow solution space. In alternative, a single algorithm can be designed to exhibit an exploratory behavior in one phase of the processes, and then focus on the neighborhood the solution found so far. Both approaches are easily combined with a multiple resolution strategy. With a scaling pyramid, the first resolution offers a smoother search space in which exploration is cheap but effective. Once a suitable area of the search space has been located, switching to a higher resolution allow the algorithm to adapt the solutions to the new level of detail.

The experimental comparison in the various studies show the advantages of these two-component or two-stage approaches. With respect to using a single local optimizer, the MH part improve the robustness of the algorithm and, in general, provides the ability to deal with a wider range of scenarios. Indeed, as classic numeric optimization techniques are local optimizers, starting the optimization far from the global optimum is likely to result in a low quality solution. In practice, this means that the larger the transformation to be found, the higher the chance of a considerable misregistration. As a result, an algorithm can suddenly be made inadequate by a change in the way the images are acquired or processed. Using multiple resolutions can mitigate the problem to some extent, but does not guarantee the optimizer will not get stuck along the way due to rugged fitness landscape.

On the other hand, a single MH with a static (as opposed to dynamic) behavior may lack in precision. Given that the computational cost of using a local optimization technique is usually quite small compared to that of a global one, refining a solution found by a MH with a local approach is a simple, cheap and effective way of improve the final solution. Nevertheless, a clever design of the MH can make this operation needless, as shown by the most modern approaches.

Ten methods use rigid transformations, either in two or three dimensions. This is a consequence of the fact that those methods are not designed for a specific application. Instead, they offer a more general IR solution, and therefore the authors chose a common transformation model in applications in medical imaging as well as outside this field. Rigid transformation is the ideal candidate for a wide range of applications, as it can be used to alter the position of the image subject while keeping its shape and size unchanged. More flexible models, especially the deformable ones, are more difficult to handle and can lead to undesired or anatomically unrealistic transformations. The remaining methods use affine transformation (two algorithms), similarity transformation (one) or a combination of affine and deformable transformation (two).

A very similar rationale is behind the prevailing use of mutual information and its variants as similarity metric. MI is popular in medical IR, but most importantly it is also the most flexible among "general" similarity metrics, meaning it is able to handle both single and multiple modality scenarios, making an ideal candidate for a general registration method. Eight algorithms use MI directly, three algorithms use normalized MI and one algorithm combines MI with gradient information. The remaining two methods use an application specific similarity metric.

Six methods use multiple resolutions. Unlike local optimization approaches, global optimizers do not require multiple resolution to deliver good results, however, significant speedup can be obtained without any major drawback. Using of subsampling in the calculation of the similarity metric is an alternative or complementary solution, and it has a similar effect on the computational cost. It can be though as a tradeoff between the speed of computing the fitness function and its precision, which can be exploited to save time during the initial phase of the optimization, in which high accuracy in evaluation a solution is not required.

The methods in this review use a variety of image modalities, with the most common being MRI (12 cases), CT (9) and PET (3). This is not likely to indicate any particular aptitude of MHs in dealing with these type of images, but it rather reflects the higher availability of this kind of data as well as the higher interest in the medical community. The fact that the brain is by far the prevailing subject, with 12 proposals out of 15 including brain images in their experiments, has a similar explanation: human brain MRI images are publicly available in a number of repositories, such as BrainWeb [60].

## 5. Conclusions

Image registration is a fundamental step in medical image analysis and a very active research field due to the steady improvement of medical imaging technology. Being based on an optimization process, IR can be tackled by metaheuristics, such as those belonging to evolutionary computation, creating a global optimization approach that is able to overcome the limitations of classic numerical optimization techniques. While metaheuristic-based approaches performing registration based on features (i.e. salient points or lines) have been recently reviewed, to the best of our knowledge this paper is the first study to analyze intensity-based techniques using metaheuristics. The review considered established, well-know techniques as well recent, outstanding contributions. We classified the approaches based on the basic components common to every IR method: the search strategy, the optimization component, the transformation model, the similarity metric, the use of multiple resolutions, the modalities and the content of the images being registered.

Having reviewed the state-of-the-art in the field, we make the following recommendations for the design of new, generic IR methods as well as for possible research lines on the topic.

- follow the parameter-based strategy and use a metaheuristic with a real-coded design and appropriate operators for continuous optimization problems. Matching-based algorithms are complex to design even for specific applications and they have no clear advantage over parameter-based ones.

- design the optimization component to have a flexible or adaptable behavior. IR requires the optimizer to quickly explore a large search space, locate the most promising area and then focus on it. This can be accomplished either by explicitly diving the optimization in multiple stages or by smoothly transitioning from one to the other. In the first case, one can exploit multiple resolutions and combine the MH with a local numerical optimization technique.

- take advantage of multiple resolutions. The ability of gradually increasing the complexity of the registration has several advantages and virtually no drawbacks. Reducing the size of the images makes the optimization faster in its early stages, and it also makes any numerical optimization technique involved more effective and less prone to get stuck in local optima. In addition, a pyramid including smoothing can reduce the influence of noise and make the algorithm more robust.

- use subsampling to speed up the calculation of the similarity metric. Unless the registration involves a very fine-grained deformable transformation, not every voxel is needed to measure the quality of the alignment.

- use mutual information or a measure based on it. Outside specific applications, MI has become the de-facto standard similarity measure and it allows the algorithm to deal with different modalities. Compared to other metrics, the disadvantage in terms of accuracy and speed is negligible.

- support multiple transformation models, as the right model depends on the application at hand. Affine transformation is flexible enough for a large number medical application, and it can easily turned into a similarity or rigid transformation model by setting simple constrains on the transformation parameters. For deformable registration, B-splines is the method of choice.

Finally, we remark that existing proposals focused on a very small group of MHs. It is likely that different MHs can improve current methods; this holds especially for those using SA, which has a basic and rigid structure that is difficult to tune properly. A possible research line, therefore, concerns the application of other, advanced MHs to intensity-based IR. To this regard, we suggest using MHs that allow a high degree of flexibility in their design. Among others, we find scatter search to be an ideal candidate. Its structure makes it easy to adapt to the specific features of the optimization problem at hand, and it has already been applied to feature-based medical IR with very good results [6]. An additional suggestion is to consider a memetic algorithm, in which the optimizer takes systematically advantage of a local optimization technique. This was one of the key features of several methods under review, although it was employed to a smaller extent.

## Bibliography

[1] Zitová B, Flusser J. Image registration methods: a survey. Image Vision Comput. 2003;21:977–1000.

[2] Wang XY, Eberl S, Fulham M, Som S, Feng DD. Data registration and fusion. In: Feng DD, editor. Biomedical information technology. Burlington, USA: Academic Press; 2008. pp. 187–210.

[3] Besl PJ, McKay ND. A method for registration of 3D shapes. IEEE Trans. Pattern Anal. Mach. Intell. 1992;14:239–56.

[4] Bäck T, Fogel DB, Michalewicz Z. Handbook of Evolutionary Computation. Bristol, UK: IOP Publishing Ltd and Oxford University Press; 1997.

[5] Glover F, Kochenberger GA, editors. Handbook of Metaheuristics. Boston, USA: Kluwer Academic Publishers; 2003.

[6] Damas S, Cordón O, Santamaría J. Medical Image Registration Using Evolutionary Computation: An Experimental Survey. IEEE Computational Intelligence Magazine. 2011 nov.;6:26–42.

[7] Audette MA, Ferrie FP, Peters TM. An algorithmic overview of surface registration techniques for medical imaging. Med. Image Anal. 2000;4:201–17.

[8] Svedlow M, Mc-Gillem CD, Anuta PE. Experimental examination of similarity measures and preprocessing methods used for image registration. In: Swain PH, Morrison DB, Parks DE, editors. Symposium on Machine Processing of Remotely Sensed Data. Indiana, USA; 1976. pp. 9–17.

[9] Maes F, Vandermeulen D, Suetens P. Comparative evaluation of multiresolution optimization strategies for image registration by maximization of mutual information. Med. Image Anal. 1999;3:373–86.

[10] Rouet JM, Jacq JJ, Roux C. Genetic algorithms for a robust 3-D MR-CT registration. IEEE T. Inf. Technol. B. 2000;4:126–36.

[11] Yamany SM, Ahmed MN, Farag AA. A new genetic-based technique for matching 3D curves and surfaces. Pattern Recogn. 1999;32:1817–20.

[12] Chalermwat P, El-Ghazawi T, LeMoigne J. 2-phase GA-based image registration on parallel clusters. Future Gener. Comp. Sy. 2001;17:467–76.

[13] Zambanini S a, Sablatnig R a, Maier H b, Langs G c d. Automatic image-based assessment of lesion development during hemangioma follow-up examinations. Artificial Intelligence in Medicine. 2010;50:83–94.

[14] Zhang Z. Iterative point matching for registration of free-form curves and surfaces. Int. J. Comput. Vision. 1994;13:119–52.

[15] Liu Y. Improving ICP with easy implementation for free form surface matching. Pattern Recogn. 2004;37:211–26.

[16] Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm. In Third International Conference on 3D Digital Imaging and Modeling (3DIM'01). Quebec, Canada; 2001. pp. 145–52.

[17] Jong KAD. Evolutionary Computation. The MIT Press; 2002.

[18] Eiben AE, Smith JE. Introduction to Evolutionary Computation. Springer, Berlin; 2003.

[19] Fogel DB. Evolutionary Computation: Toward a New Philosophy of Machine Intelligence. Wiley-IEEE Press; 2005.

[20] Pascale G, Troiano L. A Niche Based Genetic Algorithm for Image Registration. In Ninth International Conference on Enterprise Information Systems (ICEIS 2007). 2007. pp. 342–7.

[21] Brunnström K, Stoddart A. Genetic algorithms for free-form surface matching. In International Conference of Pattern Recognition. Vienna, Germany; 1996. pp. 689–93.

[22] Cordón O, Damas S. Image Registration with Iterated Local Search. Journal of Heuristics. 2006;12:73–94.

[23] Cordón O, Damas S, Santamaría J, Martí R. Scatter Search for the 3D Point Matching Problem in Image Registration. INFORMS Journal on Computingyear. 20:55–68.

[24] Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Medical Image Analysis. 2001;5:143–56.

[25] Santamaría J, Cordón O, Damas S, García-Torres JM, Quirin A. Performance evaluation of memetic approaches in 3D reconstruction of forensic objects. Soft Comput. 2009;13:883–904.

[26] Ong YS, Lim M, Zhu N, Wong K. Classification of adaptive memetic algorithms: a comparative study. IEEE Transaction on Systems, Man, and Cyberneticsnumber. 2006;36:141–52.

[27] Ong YS, Lim MH, Chen X. Memetic Computation — Past, Present & Future. IEEE Computational Intelligence Magazine. 2010;5:24–31.

[28] Valsecchi A, Dubois-Lacoste J, Stützle T, Damas S, Santamaría J, Marrakchi-Kacem L. Evolutionary Medical Image Registration using Automatic Parameter Tuning. In IEEE Congress on Evolutionary Computation. 2013.

[29] Robertson C, Fisher RB. Parallel Evolutionary Registration of Range Data. Computer Vision and Image Understanding. 2002;87:39–50.

[30] Fitzpatrick JM, Grefenstette JJ, Gucht DV. Image registration by genetic search. In IEEE Southeast Conference. Louisville, EEUU; 1984. pp. 460–4.

[31] Mandava VR, Fitzpatrick JM, Pickens DR. Adaptive search space scaling in digital image registration. IEEE Transactions on Medical Imaging. 1989;8:251–62.

[32] Tsang PWM. A genetic algorithm for aligning object shapes. Image and Vision Computing. 1997;15:819–31.

[33] Holland JH. Adaptation in Natural and Artificial Systems. Ann Arbor: The University of Michigan Press; 1975.

[34] Wachowiak MP, Smolíková R, Zheng Y, Zurada JM, Elmaghraby AS. An approach to multimodal biomedical image registration utilizing particle swarm optimization. IEEE Transactions on Evolutionary Computation. 2004;8:289–301.

[35] Clerc M, Kennedy J. The particle swarm - explosion, stability, and convergence in a multidimensional complex space. Evolutionary Computation, IEEE Transactions on. 2002;6:58–73.

[36] Kagadis GC, Delibasis KK, Matsopoulos GK, Mouravliansky NA, Asvestas PA, Nikiforidis GC. A comparative study of surface- and volume-based techniques for the automatic registration between CT and SPECT brain images. Medical Physics. 2002;29:201–13.

[37] Wang Q, Li X. Application of improved genetic algorithm in practical medical image registration. International Journal of Digital Content Technology and its Applications. 2011;5:60–7.

[38] Xu X, Dony RD. Differential evolution with powell's direction set method in medical image registration. In Arlington, VA; 2004. pp. 732–5.

[39] Pluim JPW, Maintz JBA, Viergever MA. Image registration by maximization of combined mutual information and gradient information. Medical Imaging, IEEE Transactions on. 2000;19:809–14.

[40] Talbi H, Batouche M. Hybrid particle swarm with differential evolution for multimodal image registration. In Hammamet; 2004. pp. 1567–72.

[41] Winter S, Brendel B, Pechlivanis I, Schmieder K, Igel C. Registration of CT and intraoperative 3-D ultrasound images of the spine using evolutionary and gradient-based methods. IEEE Transactions on Evolutionary Computation. 2008;12:284–96.

[42] Hansen N, Kern S. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. In: Yao X, Burke E, Lozano J, Smith J, Merelo-Guervós J, Bullinaria J, et al., editors. Parallel Problem Solving from Nature - PPSN VIII. Springer Berlin Heidelberg; 2004. pp. 282–91.

[43] He R, Narayana PA. Global optimization of mutual information: Application to three-dimensional retrospective registration of magnetic resonance images. Computerized Medical Imaging and Graphics. 2002;26:277–92.

[44] Jones DR, Perttunen CD, Stuckman BE. Lipschitzian optimization without the Lipschitz constant. Journal of Optimization Theory and Applications. 1993;79:157–81.

[45] Woods RP, Grafton ST, Watson JDG, Sicotte NL, Mazziotta JC. Automated Image Registration: II. Intersubject Validation of Linear and Nonlinear Models. Journal of Computer Assisted Tomography. 1998;22:153–65.

[46] Castellanos NP, Angel PLD, Medina V. Nonrigid medical image registration technique as a composition of local warpings. Pattern Recognition. 2004;37:2141–54.

[47] Li X, Zhang P, Brisman R, Kutcher G. Use of simulated annealing for optimization of alignment parameters in limited MRI acquisition volumes of the brain. Medical Physics. 2005;32:2363–70.

[48] Chen Y-W, Lin C-L, Mimori A. Multimodal medical image registration using particle swarm optimization. In Kaohsiung; 2008. pp. 127–31.

[49] Li Q, Sato I. Multimodality image registration by particle swarm optimization of mutual information. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Qingdao; 2007;4682 LNAI:1120–30.

[50] Loeckx D, Maes F, Vandermeulen D, Suetens P. Temporal subtraction of thorax CR images using a statistical deformation model. IEEE Transactions on Medical Imaging. 2003;22:1490–504.

[51] Du J, Tang S, Jiang T, Lu Z. Intensity-based robust similarity for multimodal image registration. International Journal of Computer Mathematics. 2006;83:49–57.

[52] Meer P, Mintz D, Rosenfeld A, Kim D. Robust regression methods for computer vision: A review. International Journal of Computer Vision. 1991;6:59–70.

[53] Valsecchi A, Damas S, Santamaría J, Marrakchi-Kacem L. Genetic Algorithms for Voxel-based Medical Image Registration. In 2013.

[54] Valsecchi A, Damas S, Santamaría J. An image registration approach using genetic algorithms. In Brisbane, QLD; 2012.

[55] Zhou D, Sun J, Lai C-H, Xu W, Lee X. An improved quantum-behaved particle swarm optimization and its application to medical image registration. International Journal of Computer Mathematics. 2011;88:1208–23.

[56] Sun J, Feng B, Xu W. Particle swarm optimization with particles having quantum behavior. In Evolutionary Computation, 2004. CEC2004. Congress on. 2004. pp. 325–31.

[57] Inaba T, He L, Suzuki K, Murakami K, Chao Y. A Genetic-Algorithm-Based Temporal Subtraction for Chest Radiographs. JACIII. 2009;13:289–96.

[58] Poli R. Analysis of the publications on the applications of particle swarm optimisation. J. Artif. Evol. App. New York, NY, United States; 2008 jan;2008:4–1.

[59] Klein S, Staring M, Pluim JPW. Evaluation of Optimization Methods for Nonrigid Medical Image Registration Using Mutual Information and B-Splines. Image Processing, IEEE Transactions on. 2007;16:2879–90.

[60] Collins DL, Zijdenbos AP, Kollkian V, Sled JG, Kabani NJ, Holmes CJ, et al. Design and construction of a realistic digital brain phantom. IEEE Transactions on Medical Imaging. 1998;17:463–8.

[61] Santamaria J, Damas S, Cordon O, Escamez, A. Self-Adaptive Evolution: Towards New Parameter Free Image Registration Methods. IEEE Transactions on Evolutionary Computation. 2012; PP:99. doi: 10.1109/TEVC.2012.2209890

[62] Santamaría J, Damas S, García-Torres JM, Cordón O: Self-adaptive evolutionary image registration using differential evolution and artificial immune systems. Pattern Recognition Letters. 2012:33:16:2065-2070