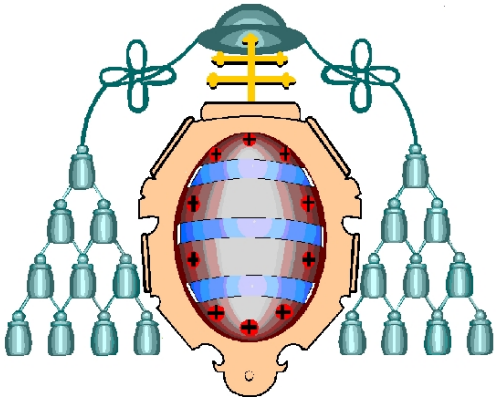
Departamento de Informática



Universidad de Oviedo

Tesis Doctoral

Intensity-based medical image registration using metaheuristics

Andrea Valsecchi

Summary

Remarkable developments in imaging technology over the last decades have brought us novel and more accurate ways to create images of the human body, such as magnetic resonance imaging and X-ray computed tomography. A key step in medical image analysis is the alignment of such images, in order to establish a spatial correspondence of their common content, a task called image registration. Applications include combining images of the same subject acquired by different sensors, detecting changes before and after treatment, motion compensation in temporal sequences and image guidance during interventions.

A common strategy in image registration is to base the alignment only on lines, corners and other features of the images. Intensity-based methods, instead, use the whole imaging data, which is computationally expensive but allows to achieve the superior level of accuracy and robustness demanded by most medical applications.

The core of every image registration method is an optimization process, which explores the space of geometrical transformations to determine the one leading to the best overlap. Metaheuristics are a family of optimization approaches, alternative to classic, gradient-based numerical optimization techniques, that are able to efficiently exploring the space of possible problem solutions in complex real-world problems.

This dissertation focuses on the application of these techniques to intensity- based medical image registration. First, by studying the state of the art in the field, we have identified the key, common element among the most outstanding existing methods: the ability to switch from a broad, global search at their earliest stage, to a more focused search once a coarse approximation of the solution has been found. Then, we have proposed a novel image registration method based on genetic algorithms. Restart and dynamic boundary mechanisms were introduced and integrated with a multiple resolution strategy. This problem-specific approach is the ultimate responsible for the excellent performance of the method. In a second proposal, further improvement has been obtained using Scatter Search for the optimization, due to its flexible, modular design and the reduced population. Our proposals have been validated through a thorough experimental comparison and statistical analysis. We designed two different experimental studies employing simulated and real-world brain magnetic resonance images, and we tested both the ability of the algorithms to perform image registration directly as well as to carry out atlas-based segmentation.

We studied the application of offline parameter tuning techniques to image registration algorithms. Through a stochastic process known as iterated racing, algorithms are

automatically tuned for a specific application and can deliver the best possible results without the burden of a manual or exhaustive tuning. In addition to demonstrating the applicability of this kind of techniques to image registration, we developed a comparison of tuned image registration algorithms. Through tuning, we removed any bias that could have affected the comparison due to of an improper setting of the algorithms' parameters. Two major conclusions were reached. On one hand, all algorithms in the comparison benefited from the tuning; on the other, our two proposals delivered the best performances in the comparison, proving the effectiveness of metaheuristic-based image registration also in this context.

Finally, we have we developed a novel segmentation algorithm based on level set. Our registration algorithm is used to perform atlas-based segmentation, resulting in a prior segmentation of the target object. The prior is used both as initial contour and as a force driving the evolution of the level set, together with other force terms that exploit information about edges and regional boundaries. We carried out a large comparison of segmentation methods on multiple medical datasets, including histological, computed tomography and magnetic resonance images, and our algorithm outperformed all seven competitors with a significant margin.

Resumen

Avances notables en la tecnología de imágenes en las últimos décadas nos han traído formas nuevas y más precisas para crear imágenes del cuerpo humano, como la resonancia magnética y la tomografía computarizada de rayos X. Un paso clave en el análisis de imágenes médicas es la superposición de este tipo de imágenes, con el fin de establecer una correspondencia espacial de su contenido común, una tarea llamada registrado de imágenes. Sus aplicaciones incluyen la combinación de imágenes del mismo sujeto adquiridas por diferentes sensores, la detección de cambios antes y después de un tratamiento, la compensación del movimiento en secuencias temporales y la guiado durante intervenciones quirúrgicas.

Una estrategia común en el registrado de imágenes es basar el solapamiento sólo en líneas, curvas y otras características de las imágenes. Los métodos basados en intensidad, en cambio, utilizan las imágenes en su totalidad, lo cuál es computacionalmente mas caro, pero permite conseguir el nivel de precisión y robustez exigido por la mayoría de las aplicaciones médicas.

La base de todos los métodos de registrado de imágenes es un proceso de optimización, que explora el espacio de las transformaciones geométricas para determinar la que conduce a la mejor superposición. Las metaheurísticas son una familia de métodos de optimización, alternativa al las técnicas numéricas clásicas basadas en el gradiente, que son capaces de explorar de manera eficiente el espacio de posibles soluciones en problemas complejos del mundo real.

Esta tesis se centra en la aplicación de metaheurísticas al registrado de imágenes basado en intensidades. En primer lugar, mediante el estudio del estado del arte en el campo, hemos identificado un elemento clave, común a los métodos más destacados: la capacidad de cambiar de una búsqueda amplia y global en su etapa más temprana, a una búsqueda más enfocada una vez que se ha encontrado una buena aproximación inicial. A continuación, hemos propuesto un nuevo método de registrado de imágenes basado en algoritmos genéticos.

En una estrategia de resoluciones múltiples se han integrado unos mecanismos de reinicio y de frontera dinámica. Este enfoque específico al problema es el mayor responsable del excelente rendimiento del método. En una segunda propuesta, se han obtenido más mejoras mediante el uso de búsqueda dispersa para la optimización, debido a su flexibilidad, diseño modular y el uso de una población reducida. Nuestras propuestas han sido validadas a través de una comparación experimental meticulosa y un análisis estadística de los resultados. Hemos diseñado dos estudios diferentes que

emplean imágenes de resonancia magnética simuladas y reales del cerebro, y hemos probado tanto la capacidad de los algoritmos para realizar el registrado de imagen directamente, así como para llevar a cabo la segmentación basada en un atlas.

Se ha estudiado también la aplicación de técnicas automáticas de ajuste de parámetros a algoritmos de registrado de imágenes. A través de un proceso estocástico conocido como carreras iteradas, los algoritmos son configurados automáticamente para una aplicación específica y pueden devolver los mejores resultados posibles sin la carga de un ajuste manual o exhaustivo. Además de demostrar la aplicabilidad de este tipo de técnicas al registrado de imagen, hemos desarrollado una comparación más objetiva de los algoritmos de registrado a consecuencia del ajuste automático de sus parámetros. Todos los algoritmos en la comparativa se han beneficiado del ajuste con respecto a sus configuración por defecto. Además, nuestras dos propuestas han dado los mejores resultados en la comparativa, lo que demuestra la eficacia del registrado de imágenes basado en metaheurísticas también en este contexto.

Por último, hemos desarrollado un nuevo algoritmo de segmentación basado en level sets. Nuestro algoritmo de registrado se utiliza para realizar una segmentación basada en un atlas, lo que resulta en una segmentación preliminar del objeto. Esta se utiliza como contorno inicial y como motor de la evolución del level set, junto con otros términos de fuerza que aprovechan la información de bordes y regiones. Hemos llevado a cabo una gran comparativa de métodos de segmentación en múltiples conjuntos de datos médicos, incluyendo imágenes histológicas, tomografías computarizadas e imágenes de resonancia magnética, en donde nuestro algoritmo supera a siete competidores de manera significativa.

Acknowledgments

I would like to thank many people I have met during my PhD. The first two are my supervisors, Sergio and José, for the confidence, help and guidance they gave me. Many thanks to the fellow members of the MIBISOC project, for the great time we have spent together across many places in Europe. Thanks to all the colleagues at the European Centre for Soft Computing, especially Nicola, Krzysztof, Arnaud, Óscar, Kevin, Ana and David, who became my friends and to whom I owe countless funny memories of my stay in Spain. Thanks to my teacher Anna and the fellow musicians Marc, Jaime and Wolfgang for being part of my small adventure in music. Thanks to my parents Alfonso and Giusy, and my sister Arianna, for being there for me. Thanks to Lorenza for being my companion and greatest friend during the last eight years.

Finally, I would like to express my gratitude to all the friends, far or near, who helped me through the last three years. Their list would not fit in this page, and, most importantly, this whole volume would not be enough to thank them for their friendship and support. This dissertation would have not been possible without them, and I would have not been the same person.

This work was supported by the European Commission under contract No. 238819 (MIBISOC Marie Curie ITN) and by the Spanish Ministerio de Economía y Competitividad under project TIN2012-38525-C02-01 (SOCOVIFI2).

Oviedo, September 2013

Andrea

Contents

List of Figures

List of Tables

List of Algorithms

*

Statement

Introduction

Perception is the organization and interpretation of sensory information in order to understand the environment. In humans, the single most important role in perception is played by images, and indeed vision is the most advanced of our senses. For machines, who are not limited to the visual band of the electromagnetic spectrum, there is an even wider class of signals that can be turned into imaging information. This broadens the perception of humans to include sources as different as ultrasounds and X-rays. Not surprisingly, then, image processing and analysis encompass a wide and varied field of technological applications.

Formally, an image is defined as function that associates a point, specified by a tuple of coordinates values, with the *intensity* value of the image in that point. When coordinates and intensity values are all finite, discrete quantities, the image is said to be a *digital image*. Such images are composed by a finite number of elements having a particular location and value. These elements are commonly called *pixels* (picture elements) or, for three-dimensional images, *voxels* (volume elements). The field of digital image processing refers to the processing of digital images by means of a computer.

Among common image processing tasks is the alignment of two or more images in order to establish a spatial correspondence of their common content. This process is called *image registration* (IR) [1]. Usually, the images to be registered 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 allows 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 determine 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 explains why it is still a very active research field.

In IR, 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 only on salient and

distinctive parts of the image, such as lines, corners and contours, called *features*. This approach, called *feature-based* [2, 1], has the advantage of greatly simplifying the problem. In the first place, the features are usually much fewer than all the pixels in the images. In addition, in most cases only the location of the features is employed, meaning that the intensity values are no longer used and can be discarded. However, feature-based methods rely on a precise and reliable detection the features. Many automatic feature extraction methods have been developed [3, 4], but, in a considerable amount of scenarios, manual or human-supervised selection is required. Any error during the feature extraction process will be propagated 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 previous feature extraction process, an approach called *intensity-based* (or voxel-based). Despite being more expensive in computational terms, intensity-based methods achieve the superior level of accuracy and robustness demanded by most medical applications, and over the last decade they have become the method of choice in medical image analysis.

Regardless of the division into feature- and intensity-based approaches, the core of every IR method is an *optimization process*, which explores the space of geometrical transformations to determine the best possible transformation to overlap the input images. Two main strategies are available. In *parameter*-based approaches, the search is performed directly in the space of the parameters that define a transformation. IR is thus turned into a continuous optimization problem, in which a solution is the vector of real values for the parameters of the registration transformation. In the other approach, called *matching*-based, 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. With this formulation, IR becomes a combinatorial optimization problem. In both parameter- and matching-based 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 the corresponding features.

Traditional parameter-based IR methods use classic numerical optimization techniques based on the gradient of the objective function, either exact or approximated. Gradient descent, quasi-Newton, nonlinear conjugate gradient and Robbins–Monro are the most eminent examples [5]. Convergence to a global optimal solution can be guaranteed only if the objective function satisfies a number of requirements, depending on the actual optimization method, such as continuity, diffentiability and Lipschitz continuity. These rather strong assumptions are in general not satisfied when dealing with IR. In addition, noise, discretization and large differences in the order of magnitude of the transformation parameters create even more complex scenarios. Likewise, traditional matching-based methods use matching algorithms such as iterative closest point (ICP) [6], which suffers from similar limitations to those of classic intensity-based techniques. As a result, what is delivered by traditional optimization methods is a *local* optimal solution, which may lead to a large misregistration.

Metaheuristics (MHs) [7] are a family of alternative optimization approaches that

combine basic heuristic methods in higher level frameworks, with the aim of efficiently and effectively exploring the space of possible problem solutions. MHs are often used to deal with complex real-world problems in which traditional optimization approaches are not reliable enough. In particular, MHs, and especially the biologically-inspired approaches belonging to evolutionary computation (EC) [8], have recently outperformed classic optimization techniques in computer vision and image processing problems [9]. As for feature-based IR, approaches using MHs have already demonstrated to be effective solutions to overcome the drawbacks of traditional optimization algorithms in medical applications, as shown in a recent review [10]. This dissertation, instead, focuses on intensity-based IR methods, which have not yet reached the same degree of maturity of the feature-based counterpart. Compared to the latter, the former approaches have just recently started to experience a larger interest from the computer vision community, mostly due to larger computing power requirement, which can now be satisfied by modern technologies.

Objectives

The main limitation of traditional intensity-based IR techniques is their inability to perform global optimization, resulting in poor performance when facing complex registration scenarios. The main objective of the dissertation is to propose novel medical IR methods, whose optimization procedure exploits the capabilities of MHs to perform a robust, global optimization and is therefore able to deal with the challenging IR applications arising in medical image analysis. Specifically, this objective is divided into the following ones:

- *Study the state of the art in intensity-based medical IR using MHs and EC computation.* By studying the most relevant and successful methods in the field, we aim to identify the key elements in their design along with their strengths and limitations.

- *Design new IR methods taking advantage of robust optimization algorithms based on EC or other MHs.* We aim to tailor one or more optimization techniques for the specific features of the IR problem, integrating expert knowledge about the structure of the solutions or any other element that can be exploited to improve the optimization. In addition to robustness and efficiency, we want the algorithms to be flexible enough to deal with a wide range of IR tasks, without limitations on dimensionality, modality or type of transformations.

- *Validate the performance of our proposals through a thorough experimental comparison and analysis.* We aim to assess the quality of our algorithms through large experimental studies. In particular, we want to employ publicly-available synthetic/benchmark imaging data, to easily allow for comparison with other algorithms, as well as real-world data, to tackle a specific medical application. In addition, we aim to compare our algorithms with an heterogeneous group

of techniques, not limited to a specific type of approaches. This includes both intensity- and feature-based methods using different optimization techniques.

- *Apply approaches for algorithm configuration to IR.* The performance of IR algorithms can depend strongly on their parameter settings, and manual parameter tuning is time consuming, error-prone and requires expertise. We aim to apply automatic techniques to deal with algorithm configuration to IR methods.

- *Integrate our algorithms into larger medical applications based on image registration.* IR is a fundamental preprocessing step in a large number of medical analysis tasks, including segmentation of anatomical structures, tissue classification and morphological analysis. In addition to direct IR applications, we aim to prove the effectiveness of our methods in other tasks which include or can benefit from high-quality registration.

Outline

The dissertation is organized in six chapters and the structure of each of them is briefly introduced as follows.

- Chapter 1 introduces the basis of IR and MHs. The chapter also presents a review of outstanding medical IR methods. The analysis focuses on intensity-based algorithm using MHs, but includes also classic and feature-based techniques.

- In Chapter 2 we present a novel IR method based on Genetic Algorithms. The algorithm is tested in two experimental studies, in which it is compared with other well-known IR methods. An extended statistical analysis of the performance of the algorithms is also carried out.

- Chapter 3 introduces a Scatter Search-based approach to IR, with the aim of providing a faster, more accurate and more robust algorithm than those in the literature and in our previous proposal. The new method is also validated on the two already mentioned experimental comparisons with a further statistical analysis of the results.

- In Chapter 4 we tackle the problem of algorithm configuration. The contribution of the chapter is two-fold. On one hand, we apply a technique for automatic parameter tuning to our IR methods. On the other, we carry out a comparison of tuned registration algorithms. This demonstrates the use of automatic tuning to perform comparisons that are not biased by the specific configuration of the algorithms.

- Chapter 5 presents a novel, advanced medical image segmentation algorithm in which our second IR proposal plays a fundamental role. The registration is used both to provide an initial contour of the object to be segmented and as one of the forces that drives the evolution of the system.

- Finally, the results achieved in this dissertation are summarized in Chapter 6, which also includes conclusions, future work and new research lines on the topic.

Chapter 1

Introduction

This chapter is devoted to image registration and metaheuristics. After providing a general introduction to the two fields, we study how MHs and classic optimization techniques are applied to IR. Finally, we analyze the design of the most important contributions in the literature. The chapter is organized as follows. Section 1.1 provides some preliminaries on medical image registration, while Section 1.2 introduces metaheuristics and evolutionary computation, including a general overview, a description of common techniques and their adaptation to IR. Section 1.3 reviews outstanding contributions in medical IR and discusses trends and other patterns in the design of the most successful methods.

1.1 Image registration

1.1.1 Problem definition

A typical IR problem involves two images, conventionally called *model* (I_M) and *scene* (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 according to the chosen similarity metric. In other terms, IR is a maximization problem over transformations, formally stated as

$$\underset{T \in \text{ Transformations}}{\text{argmax}} \quad \text{Similarity}\left(I_M, T(I_S)\right)$$

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. The optimization is usually an iterative process. The optimizer computes a candidate transformation, which is then applied to the scene image. Next, the similarity metric compares the model with the transformed scene image and returns a quality value that is sent back to the optimizer. Figure 1.1 shows

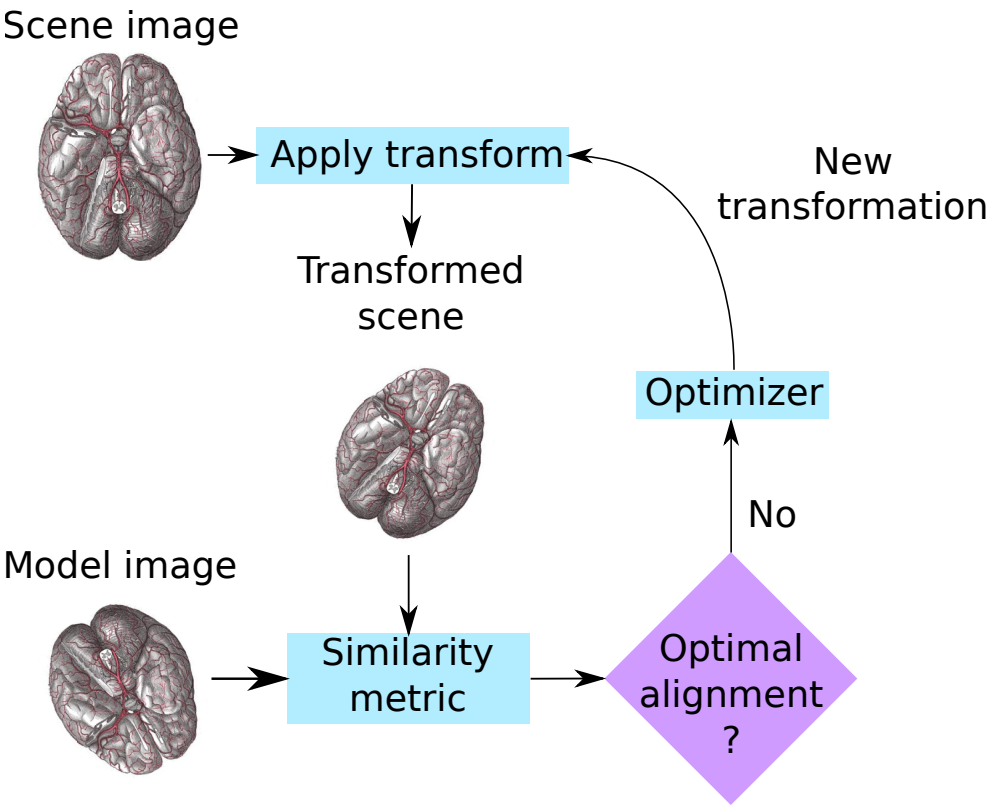


Figure 1.1: The interactions among the components of a registration technique.

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.

1.1.2 Components

1.1.2.1 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. 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 results that are undesired or implausible with respect to the phenomenon under study. Bones being bent and tissues growing at an unrealistic rate are examples of such unacceptable outcomes in the medical context.

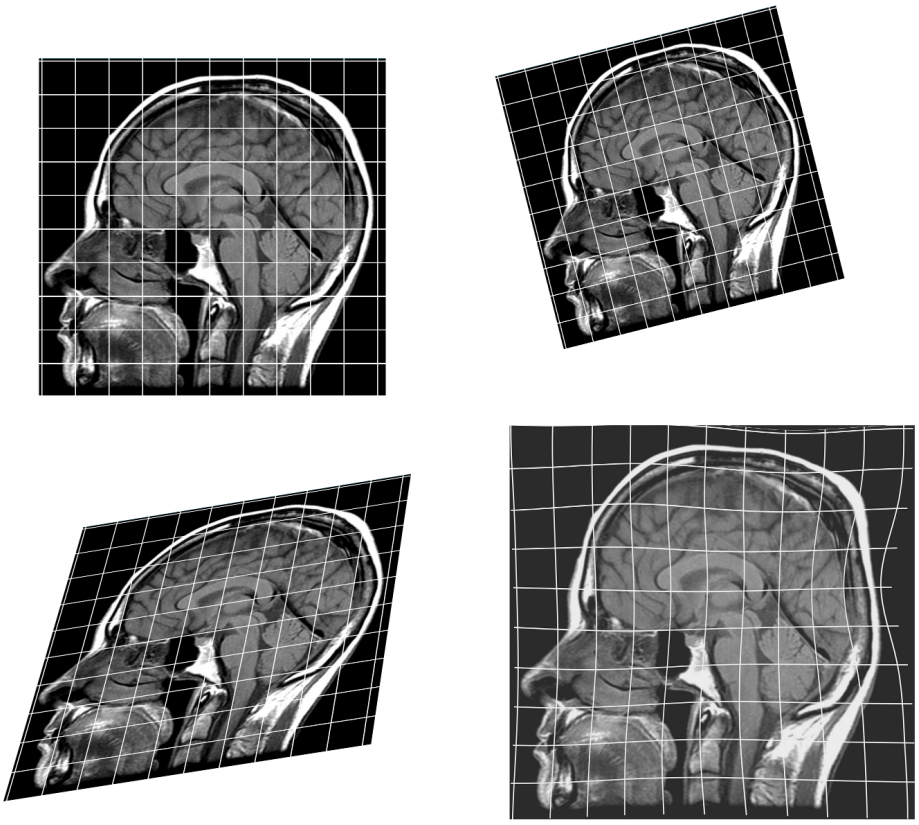


Figure 1.2: Images obtained from the same scene (top left) by applying different transformations: similarity (top right), affine (bottom left) and B-spline (bottom right).

Rigid transformation A common assumption in medical image registration is that both images are related by a rigid transformation. For example, for images of the head, the rigid-body assumption is normally justified as the skull is rigid and constrains the motion of the brain sufficiently. In three dimensions, a rigid transformation involves six degrees of freedom (DoFs): three rotations and three translations. The transformation can be conveniently represented in matrix form using homogeneous coordinates, and it can be specified as a combination of three rotation operations R_x, R_y, R_z (one for each axis) and a translation t, i.e.

$$T_{\text{rigid}} = R_x \cdot R_y \cdot R_z \cdot t$$

where

$$R_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha & 0 \\ 0 & -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad R_y = \begin{pmatrix} \cos\beta & 0 & -\sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R_z = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 & 0 \\ -\sin\gamma & \cos\gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad t = \begin{pmatrix} 0 & 0 & 1 & t_x \\ 0 & 1 & 0 & t_y \\ 1 & 0 & 0 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Similarity transformation In some cases, it is necessary to correct not only for rigid transformation but also for scaling. This adds a uniform scaling factor s and the corresponding matrix S'.

$$T_{\text{similarity}} = S' \cdot T_{\text{rigid}} \qquad S' = \begin{pmatrix} s & 0 & 0 & 0 \\ 0 & s & 0 & 0 \\ 0 & 0 & s & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Affine transformation In complex scenarios, it may also be necessary to correct for shears, for example caused by the gantry tilt of CT scanners, or for scaling on a per-axis basis. This means having two transformations H and S,

$$H = \begin{pmatrix} 1 & h_{xy} & h_{xz} & 0 \\ h_{yx} & 1 & h_{yz} & 0 \\ h_{zx} & h_{zy} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad S = \begin{pmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Combining the rigid transformation matrix with the scaling and shearing matrices yields an affine transformation with a total of 12 DoFs for 3D images.

$$T_{\text{affine}} = S \cdot H \cdot T_{\text{rigid}}$$

Projective transformation Projective transforms play an important role in applications involving the alignment of 3D volumes like CT or MRI to 2D images such as radiography and photography. Different types of projections including parallel or perspective projections can be used depending on the application. However in most cases, the transformation that relates the 3D and 2D images is a combination of a projective with a rigid transformation, which determines the pose of the 3D volume relative to the camera. Often it is possible to determine the perspective transformation parameters using either knowledge about the internal geometry of the camera, or by camera calibration techniques, in which case the problem is reduced to rigid registration.

Deformable transformations The transformation models discussed so far can be characterized by a small number of parameters, between six and twelve. While such transformation models are frequently used for the registration of anatomical structures like the brain or bones, they are not applicable in the case where significant deformation is expected, e.g., in soft tissues like the liver or breast. In these cases, deformable or non-rigid transforms are required.

In a typical approach, an object is deformed by manipulating an underlying mesh of control points. The resulting deformation controls the shape of the 3D object and produces a smooth and continuous transformation between control points. A popular choice is to use tri-variate B-spline tensor products [11], i.e.

$$T(x) = x + \sum_{x_k \in N_x} p_k \beta^3 \left(\frac{x - x_k}{\sigma} \right)$$

where x_k are the control points, β^3 is the cubic B-spline polynomial, p_k are the coefficient vectors (loosely speaking, the control point displacements), σ is the control point spacing and N_x is the set of all control points within the compact support of the B-spline at x.

1.1.2.2 Similarity metric

The performance of any IR method depends on an accurate estimation of the degree of alignment between the images, therefore the similarity metric is considered a crucial component [12]. Technically, a similarity metric is a real-valued function $F(I_A, I_B)$ that measures the resemblance of two images I_A, I_b. The quality of a transformation T is assessed by computing the similarity metric over the model I_M and the transformed scene $T(I_S)$, i.e. $F(I_M, T(I_S))$. 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 [13]. 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.

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

where r is the number of points and c is the correspondence criterion.

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 relationship is established between their intensity distributions. For instance, if we assume this relationship to be 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

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

where $\overline{I_A}, \overline{I_B}$ are the average intensity value of the images and Ω 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 their intensity values 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

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

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

1.1.2.3 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 the 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 into a continuous optimization problem, therefore classic numerical optimization algorithms can be used. Gradient descent, conjugated gradient descent, Newton's and quasi-Newton methods, Powell's method and discrete optimization [14, 5] are among the most common choices along with approaches based on evolutionary computation and other metaheuristics [15, 16, 17]. 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 [18]. This class of algorithms is called *matching-based*; the iterative closest point algorithm (ICP) is a famous example of this second approach [19, 20, 21].

1.1.2.4 Other minor components

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.

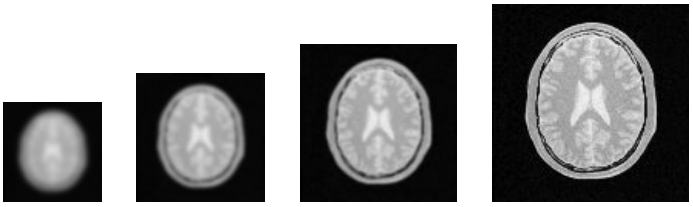


Figure 1.3: A Gaussian pyramid of a 2D MRI human brain image.

Second, following a hierarchical approach, it is common to perform the registration in multiple stages. The registration begins using a down-scaled, smoothed version of the input images. This serves a two purpouses. In the first place, minor image details disappear, removing a number of small local optima from the objective function. Furthermore, reducing the size of the images makes the computation of the transformation faster. Once the optimizer has found a suitable solution, the registration process enters the next stage. Some of the details of the input images are 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*. Figure 1.3 shows a four-resolution pyramid of a human brain image.

1.1.3 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 create 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 a regular basis, with the aim of detecting the changes that occurred between consecutive acquisitions. This is the case of motion tracking in cardiovascular malfunctions studies 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 get 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 dimensionality, 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 compared with digital atlases or ''average'' specimens.

1.1.4 Validation and comparison of registrations

Validation usually consists in showing that a registration algorithm, when applied to typical data in a given application, consistently succeeds with an acceptable maximum (or average) error. For feature-based approaches, a real-world error can be computed in the form of distance between corresponding features after the registration. Such an analysis is not generally possible for non-rigid techniques, because although the error in the feature position can be computed, that of other parts of the volume is dependent on the transformation model and must be estimated using other means. In intensity-based approaches, the similarity metric does not usually provide any measure of the real-word displacement, therefore validation is usually performed by making additional measurements post registration. A common approach is to identify corresponding landmarks or regions independently of the registration process and measure how well the registration brings them into alignment [22]. The segmentation of anatomical structures provides the means to measure the structure overall distances before and after registration.

1.2 Evolutionary and metaheuristics-based image registration

1.2.1 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 [8, 23, 24, 25].

Metaheuristics arise in the field of optimization. In the simplest case, an optimization problem consists of minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function. Formally, given a function $f: D \to \mathbf{R}$ from some domain D to the real numbers, the solution of the optimization problem is an element $x_o \in D$ such that $f(x_0) \leq f(x) \forall x \in D$. The function f is usually called objective or cost function, while x_0 is called a global optimal solution.

If the objective function and the domain are linear, the problem can be solved efficiently using the simplex method [26]. Although a large number of convex optimiza-

tion problems can be solved exactly, general optimization is **NP**-hard. Exact algorithms, such as the branch and bound method [27], have been developed, but in practice their application is computationally infeasible and approximate methods are to be used.

Gradient-based methods can be applied if f is differentiable on D, but even in this case the convergence to a global optimal solution is not guaranteed. A metaheuristic, instead, makes few or no assumptions about the problem to optimized. A MH 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 [28]. 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) [29], tabu search (TS) [30], multi-start methods, iterated local search (ILS) [31], greedy randomized adaptive search procedures (GRASP) [32], memetic algorithms (MAs) [33] and scatter search (SS) [34].

Evolutionary Computation (EC) [31] 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) [8, 25] such as genetic algorithms (GAs) [35], evolution strategies (ES) [36], genetic programming (GP) [37], differential evolution (DE) [38] and evolutionary programming (EP) [8]. In particular, GAs are probably the most used EA in the literature to face real-world optimization problems. Another important branch of MHs is swarm intelligence (SI) [39], 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) [40] and particle swarm optimization (PSO) [41].

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.

1.2.1.1 Common metaheuristics for image registration

In this section we describe in detail the four MHs used in most of the methods under review, namely GAs, PSO, SA and SS. These are also good representatives of different branches of metaheuristics, thus this section also gives a flavor of MHs and the diversity of this kind of approaches.

Simulated Annealing The name and inspiration of SA [29] 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 than s, r is accepted with probability

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

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.

Algorithm 1: The Simulated Annealing procedure.

```
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
    end
    decrease t
until stop condition
return s
```

Genetic Algorithms Like most algorithms in EC, GAs [35] mimic the processes observed in natural evolution. A solution of the optimization problem is thought 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*.

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

Algorithm 2: A basic Genetic Algorithm procedure.

```
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 new population is full
    replace the population with the new one
until stop condition
```

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 $\{0, \ldots, 255\}$. 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 [41] is a population-based algorithm, but rather than a population of biological organisms, this group of solutions is thought 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 $w, c_1, c_2 \in \mathbf{R}$ 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.

Algorithm 3: A basic Particle Swarm Optimization procedure.

```
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
```

Scatter Search Scatter Search (SS) [42] is a population-based algorithm. Unlike other MHs such as GAs, SS uses a very small pool of solutions called *reference set*. New candidate solutions are created by combining small set of solutions and applying an *improvement method* to the result. The reference set stores the most high-quality solutions generated throughout the process, while the others are discarded.

The original SS proposal, as coined by Glover [43], has a very general template with five fundamental operations corresponding to five methods:

- a *diversification generation method*, that generates a set of diverse trial solutions;
- an *improvement method*, used to enhance a solution, usually applying an heuristic method;
- a *reference set update method* to build and maintain a reference set of solutions selected for their quality or diversity;
- a *subset generation method*, by which sets of solutions from the reference set are created;
- a *solution combination method*, that combines a set of solutions into one or more new solutions.

Algorithm 4: A basic Scatter Search procedure.

```
generate a set of initial solutions P using the DiversificationGeneration
method
apply the ImprovementMethod to the elements of P
update the ReferenceSet using the ReferenceSetUpdate method
repeat
    create Subsets using the SubsetGeneration method
    while Subsets ≠ ∅ do
        S ← pop(Subsets)
        x ← SolutionCombination(S)
        x′ ← ImprovementMethod(x)
        store x′ temporarely
    end
    update the ReferenceSet using the ReferenceSetUpdate method
until stop condition
```

1.2.2 Suitability of EC and other MHs 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 [44].
- 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 [45, 46, 47] 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 [48, 49]. An outstanding approach to exploit the benefits of both strategies is their hybridization in the well-known memetic computation paradigm [50, 51]. Such scheme was successfully applied to the IR problem in [49]. 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:

- They require a tuning of the control parameters, which is often a manual, error-prone, expert-based procedure. Approaches based on automatic parameter tuning [52] or MHs with an adaptive behavior [50] 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 [53, 54].
- 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 their effectiveness.

1.3 State of the art algorithms for medical IR

In this section we review outstanding proposals in medical IR. We include both classic optimization techniques and algorithm based on metaheuristic and evolutionary computation. The study includes the most cited contributions in the literature and, in addition, a number of outstanding proposals published between 2011 and 2013 that, being introduced only recently, have not yet reached a high number of citations.

1.3.1 Feature-based techniques

1.3.1.1 Classic algorithms

Iterative Closest Point ICP [6] is a well-known matching-based algorithm in computer-aided design, originally proposed to recover the 3D transformation of pairs of range images[1]. The method proceed as follows:

- We are given a point set P of N_p points $\vec{p}_i$ from the data shape (the scene) and the model image X of N_x. In the original contribution, the approach dealt with 3D rigid transformations stored in a solution vector $\vec{q} = [q_1, q_2, q_3, q_4, t_1, t_2, t_3]$. The

[1] In a range image, the intensity value of a pixel is the distance between the sensor and the corresponding point of the scene. The registration of range images is an important problem in computer vision [55, 56].

first four parameters correspond to a quaternion, determining the 3D rotation component, while the remaining three parameters store the translation vector.

- The procedure is initialized by setting $P_0 = P$, the initial registration transformation to $\vec{q_0} = [1, 0, 0, 0, 0, 0, 0]$ and the iteration counter k to zero. The next four steps are applied until reaching convergence within a tolerance threshold $\tau > 0$:

  1. Compute the matching between the scene and the model points using the closest assignment rule.
  $$Y_k = C(P_k, X)$$
  2. Estimate the registration by least squares.
  $$f_k = \rho(P_o, Y_k)$$
  3. Apply the registration transformation to the scene image:
  $$P_{k+1} = f_k(P_0)$$
  4. Terminate the procedure if the change in MSE falls below τ. Otherwise increase k and go to step 1.

Note that ICP is not directly guided by a similarity metric but rather by the computed matching, as other matching-based IR methods. In this strategy, the function MSE only plays the role of the stopping criterion. Moreover, the transformation estimator is a numerical method that depends on the good outcomes of the matching step. Thus, the better the choice of the correspondences that is performed, the more precise the estimation of the transformation f. Consequently, the value of the similarity metric will be more accurate, leading to a proper convergence.

The original ICP proposal has three main drawbacks. First, the algorithm is very dependent on the initial guess and it is likely to get trapped in local optima, which forces the user to manually assist the IR procedure. Also, the algorithm can not cope with non-overlapping regions, as outliers are never removed. This means one of the images needs to be contained in the other. Finally, ICP can only handle normally distributed observations. Many extensions of the original proposal have been presented, partially solving some of its main shortcomings [19, 20, 21].

1.3.1.2 Algorithms based on EC and MHs

Liu's ICP Liu's algorithm [20], called I-ICP, is an improvement of the original ICP proposal. ICP has several limitations: it is very dependent on the initial guess, it is likely to get trapped in local optima solutions and it can only handle normally distributed observations. The novelty in I-ICP 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.

To improve the robustness of the algorithm, the authors implemented a mechanism to escape local optima, making it similar to SA. When the algorithm has converged to a solution x, a small perturbation is applyied to x and the algorithm is resumed. As a result, the algorithm is forced to search for an optimal solution in the neighboring region of the estimated parameters.

Cordon et al.'s SS The algorithm is a matching-based IR method based on Scatter Search. It has been introduced in [57], in which it was found to be the best algorithm in a comparison of feature-based IR techniques.

In this proposal, called SS*, the authors designed new improvement, combination and diversification-generation methods. Crest-lines points [58, 59] 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 defined a function m that evaluates the quality of a matching π using the mentioned curvature values as follows

$$m(\pi) = \Delta k_1 + \Delta k_2 \quad \text{where} \quad \Delta k_j = \sum_{i=1}^{r} (k_j^i - k_j^{\pi_i})^2 \quad \text{for } j = 1, 2$$

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.

Chow et al.'s GA The proposal [60] 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. This search space reduction is repeated until no improvement in the best solution is detected.

Wachowiak et al.'s TS In this contribution [61], the authors introduced an adapted Tabu Search [62] method to tackle 2D to 3D IR problems. Specifically, they adopted a parameter-based approach and addressed medical IR problems considering 2D ultrasound scans (US) and 3D MRIs. They considered the maximum value of the overlap-invariant normalized MI [63] functional as similarity metric. In their work, once the TS-based IR method has identified a promising area of problem solutions, the affine shaker algorithm [64] is applied to locate the best point in each subarea. The shape of the promising area is thus adapted to include areas of the search space that may have been missed during diversification. Finally, an intensified search is conducted around the most promising point using the Nelder-Mead simplex algorithm [65].

Santamaría et al.'s SS In [49], the authors performed a large study of the performance of parameter-based memetic IR methods. The baseline algorithm were recent IR proposals based on SS [66], CHC [67], and DE [68], while the local search algorithms were XLS [69], Solis & Wets's method [70] and Powell's method [71]. Two different criteria (deterministic and probabilistic) for the application of the local search and different search intensity levels were tested. After comparing a total of fifty seven algorithms on a 3D reconstruction task, the authors recommended the combination of SS with XLS using the deterministic criterion. This algorithm was able to outperform all competitors, including the baseline algorithms, by a significant margin.

1.3.2 Intensity-based techniques

1.3.2.1 Classic algorithms

We discuss four well-established continuous optimization algorithms for IR: gradient descent (GD), quasi-Newton (QN), nonlinear conjugate gradient (NCG) and adaptive stochastic gradient descent (ASGD). In [5], the authors review and compare IR methods for the registration of follow-up chest CT scans and found these methods to yield the best results. All four algorithms are gradient-based; they consist of an iterative optimization process

$$\mu_{k+1} = \mu_k + a_k d_k$$

where d_k is the search direction at iteration k, and a_k is a gain factor that controls the step size along the search direction. The search directions and the gain factors are chosen such that the sequence μ_k converges to a local minimum of the similarity metric. The difference between the four optimizers lies in the way the search direction and the gain factor are computed.

Gradient descent The gradient descent method takes steps in the direction of the negative gradient of the cost function, i.e.

$$\mu_{k+1} = \mu_k - a_k g(\mu_k)$$

where g is the derivative of the cost function and the gain factor is the decaying function $a_k = a/(k + A)^\alpha$ with $a > 0$, $A \geq 1$ and $0 \leq \alpha \leq 1$.

Quasi Newton The quasi-Newton method also moves along the negative gradient direction. The gain factor is an approximation of the inverse Hessian matrix $[H(\mu_k)]^{-1}$, computed using the Broyden–Fletcher–Goldfarb–Shanno method.

Nonlinear Conjugate Gradient In the nonlinear conjugate gradient method, the search direction is a linear combination of the gradient and the previous search direction, i.e. $d_k = -g(\mu_k) + \beta_k d_{k-1}$. Several expressions for gradient type β_k have been proposed in the literature. The gain factor is determined by an inexact line search routing, the Moré–Thuente algorithm.

Adaptive Stochastic Gradient Descent The adaptive stochastic gradient descent [72] is based on the Robbins–Monro stochastic optimization procedure, which is designed to deal with noisy observation of the objective function. It follows the same scheme as the regular gradient descent, but implements an adaptive step size mechanism and an automatic estimation procedure for the parameters a and A. ASGD considers the solutions

$$\mu_{k+1} = \mu_k - \gamma_k(t_k)g_k$$

where

$$t_{k+1} = \max(0,\ t_k + \text{sigm}(-g_k \cdot g_{k-1}))$$

and $\gamma_k = \frac{a}{t_k + A}$. The ''time'' t_k is adapted depending on the inner product between the current and the previous gradients. If the gradients have the same direction, the time is reduced, leading to a larger step size.

1.3.2.2 Algorithms based on EC and MHs

Wachowiak et al.'s PSO The proposal [73] 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 [74], through 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 [75], 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 [76], 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 population 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 [77], 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* [78]. 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 α between the gradients in the two images is computed. The gradient information between the images A, B is the sum

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

where w is a weighting function that favors angles that are either very small or close to π. 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 to 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 increasingly 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 [79], 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 δ. 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 seem 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 [80] 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 image 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 [81].

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 [82], the authors combine a real-coded GA with DIRECT, a deterministic, global optimization method based on *branch and bound* [83]. 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 σ 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 σ_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 [84] 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 [85] 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 focuses on one warping at the time, starting from the largest one, so the optimizer is run multiple times and 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 [86], 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 [87], 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 [73]. 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 [88] 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 [89] 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

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

where I_{diff} is the difference image and σ acts as a sort of threshold over the intensity difference in an artifact, meaning differences smaller than σ 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 [90], 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 [91].

Two resolution are used. The algorithm uses simulated annealing at the coarser resolution and Powell's method at the finer 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 focuses 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.

Zhou et al.'s PSO The study [92] proposes an advanced quantum-behaved PSO algorithm (QPSO) [93]. 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 $u \sim \mathrm{Unif}(0,1)$, 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 larger search scope and viceversa. The update equation can be rewritten as

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

where $r \sim \text{Norm}(0,1)$. The parameters α and β 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 α and β 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 α and β. 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 α, β 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 α and β. 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 in all but one of the scenarios.

1.3.2.3 Discussion

The aim of this section is to analyze trends and other patterns that occur in the group of intensity-based algorithms described in the previous section. We will focus on the main IR components mentioned in Section 1.1.2, namely the transformation model, the similarity metric and the optimizer. Minor IR components are considered as well.

We begin by remarking that all algorithms under review follow a 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 [94].

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

	Ref.	Year	Strategy	Optimizers	Transformation Model	Multiple resolutions	Multi modal	Metric	Modalities	Target
Kagadis et al.	[75]	2002	Parameters	GA, Powell	Affine 3D	No	Yes	MI	CT, SPECT	brain
He and Narayana	[82]	2002	Parameters	GA, DIRECT	Rigid 3D	Yes, 5, gaussian	Yes	MI	MRI	brain
Loeckx et al.	[89]	2003	Custom	SA	Affine + B-splines 2D	Yes, 2	No	Patter Intensity	Rx	thorax
Wachowiak et al.	[73]	2004	Parameters	PSO, Powell	Rigid 3D	No	Yes	NMI	Histological, MRI, US, CT	brain, abdomen
Xu and Dony	[77]	2004	Parameters	DE, Powell	Rigid 3D	Yes, 2, wavelet	Yes	MI and Gradient Information	MRI, CT	brain
Talbi and Batouche	[79]	2004	Parameters	PSO	Rigid 2D	No	Yes	MI	MRI, SPECT, CT	brain
Castellanos et al.	[85]	2004	Parameters	GA	Composition of local warpings, 2D	Yes, 3, shrinking	Yes	NMI	MRI, Rx, PET, CT	brain, thorax
Li et al.	[86]	2005	Parameters	SA	Rigid 3D	No	No	MI	MRI	brain
Du et al.	[90]	2006	Parameters	SA, Powell	Rigid 3D	Yes, 2	Yes	DRPIU	MRI, PET	brain
Li and Sato	[88]	2007	Parameters	PSO	Rigid 2D	No	Yes	MI	PD, MRI	brain
Winter et al.	[80]	2008	Parameters	CMA-ES	Rigid 3D	No	Yes	Application specific	CT, 3D US	spine
Chen et al.	[87]	2008	Parameters	PSO	Rigid 2D	No	Yes	MI	MRI, CT	liver
Wang et al.	[76]	2011	Parameters	GA, Nelder-Mead	Similarity 2D	No	Yes	NMI	MRI, CT	brain
Zhou et al.	[92]	2011	Parameters	Quantum PSO	Rigid 3D	No	Yes	MI	MRI, CT, PET	brain

Table 1.1: The characteristics of the methods under review.

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 (one algorithm), 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. Seven algorithms use MI directly, three algorithms use normalized MI and one algorithm combines MI with gradient information. The remaining three methods use an application specific similarity metric.

The next component is the optimization procedure, which is the focus of our study. For the MH part, the most prevailing choices are PSO, GAs and SA, with respectively 5, 3 and 3 contributions over 14. This is not unexpected, as 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 [95]. Also, GAs are among the most established algorithms in MHs and they have been applied successfully to IR in the past. 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 [80] are quite impressive. However, in a different study [5], 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 allows the algorithm to adapt the solutions to the new level of detail.

The experimental comparison in the various studies shows the advantages of these two-component or two-stage approaches. With respect to using a single local optimizer, the MH part improves 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 improving the final solution. Nevertheless, a clever design of the MH can make this operation needless, as shown by the most modern approaches.

Five 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 thought 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 (11 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 11 proposals out of 14 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 [96].

Chapter 2

Intensity-based medical image registration using Genetic Algorithms

In this Chapter, we introduce our first proposal to tackle medical IR using MHs. We used a GA for the optimization component, a popular choice among MHs and, in particular, in the context of IR. However, as a result of insights we gathered during the review of the state-of-the-art in this kind of algorithms, we employ a modern, modular design that takes advantage of the specific features of the IR problem.

Most of the Chapter is devoted to the validation of our algorithm. We have developed two experimental studies using simulated and real-world brain MRIs. Our proposal is compared with a large and heterogeneous group of outstanding IR algorithms, and the results are evaluated through a quantitative, statistical analysis.

The Chapter is structured as follows. Section 2.1 analyses previous works in the application of GA to IR. Section 2.2 describes the proposed approach, whereas Section 2.3 introduces the test scenarios, the experiments performed and the analysis of their results. Finally, Section 2.4 provides conclusions.

2.1 Genetic algorithms in image registration

The first examples of algorithms tackling IR using GAs were proposed in the eighties. In 1984, Fitzpatrick et al. introduced a binary-coded GA for the registration of 2D angiographic images [97]. In 1989, Mandava et al. [98] used a 64-bit structure to represent the eight real parameters of a bilinear transformation in a binary GA. Brunnström and Stoddart [45] proposed a method based on a manual prealignment of range images followed by an automatic IR process. In the latter, the registration is matching-based and performed by a novel GA. Tsang [99] used a binary-coded GA with a 48-bit chromosomes to encode three test points as base for computing a 2D affine registration. In Yamany et al.'s [16] and Chalermwat et al.'s [17] proposals, 3D and 2D rigid transformations are encoded using respectively 42 and 22 bits. Rotation

angles and translation components, therefore, can only assume a discrete set of values, depending on their range and the corresponding number of bits in the chromosome.

All the latter approaches show severe pitfalls from an EC perspective. First, they use binary encoding to represent solutions of a continuous optimization problem, an approach that suffers from discretization issues and requires transformation to real values at each evaluation of a solution. Moreover, the GAs considered are based on the old-fashioned original proposal by Holland [35], which uses fitness-proportionate selection, one-point crossover and simple bit-flipping mutation. Such selection strategy causes a strong selective pressure, thus having a high risk of premature convergence of the algorithm. In addition, single-point crossover and bit-flipping mutation are not suited for chromosomes representing real values. In such case, their effect is excessively disruptive, thus it is difficult for the operators to create useful descendants.

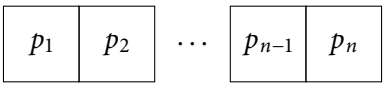
2.2 r-GA$^+$, an evolutionary approach to IR

This section describes our GA-based proposal to tackle intensity-based medical IR. The registration is carried out through a search in the space of transformation parameters. As reviewed in Section 1.1, every IR method has three major components: a transformation model, a similarity metric and an optimizer. In this proposal, the optimization is carried out by a GA. To avoid the shortcomings discussed in the previous sections, we used a real-coded representation and operators. In addition, to make our method as general as possible, we designed the optimizer to be able to handle different transformation models and similarity metrics, so that their actual values can be chosen by the user depending on the registration application at hand.

2.2.1 Generic components

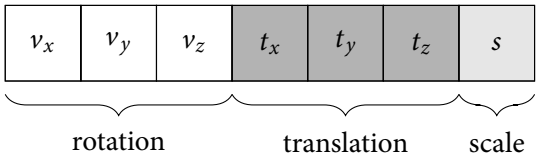
This section describes the components of the GA in detail, following the general structure introduced in Section 1.2.1.1.

Coding scheme A solution of our optimization problem is a geometric transformation specified by a sequence of real-valued parameters $p_1, \ldots, p_n$ depending on the actual transformation model used for the application. Therefore, an individual of the GA is simply a real vector with n elements.



The range of each parameter depends on the specific information it encodes, while other constraints can be added according to the actual registration scenario, to incorporate knowledge about the solution of the problem.

For instance, a 3D similarity transformation can be represented by seven parameters: three to specify the versor of the rotation v, three for the translation t and one for the scaling factor s, i.e.



Valid solutions require $v_x, v_y, v_z \in [-1, 1]$ and $s > 0$.

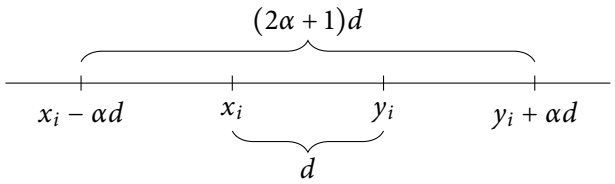
Genetic operators The genetic operators are selection, crossover and mutation, forming the core of the GA. The selection mechanism is responsible for choosing the set of individuals whose offspring will constitute the population at the next generation. According to the inspiring principle of genetic algorithm, the better is the fitness value of an individual, the higher is its chance of being selected. The proposed approach uses tournament selection: k individuals are drawn from the population using uniform probability, and the best individual of the group (the winner of the tournament) is selected.

After selection, variation operators are used to create new individuals based on the ones already in the population. The traditional approach uses a crossover operator, that combines the genetic material of two parents to generate two offspring, and a mutation operator, that given one individual returns a slightly altered version of it.

The operators used in our method are common choices for real-coded genetic algorithms: blend crossover (BLX-α) [100] and random mutation [8]. The random mutation operator picks randomly one the of individual's genes and replace it with a random value in the gene's range. Both random choices are made using uniform probability. Blend crossover is more complex. Given to two individuals x and y, called ''parents'', for each position i of the parents' coding, the algorithm computes the value $d = |x_i - y_i|$ and then randomly generates two values a, b in the interval

$$[\ \min(x_i, y_i) - \alpha d,\ \max(x_i, y_i) + \alpha d\ ]$$

with uniform probability. The values a and b are assigned to the i-th positions of the two offspring. The value α is a positive value controlling the width of the ranges in which the new genes' values are drawn.



Fitness function In general, the fitness function assigns a quality value to each solution. The fitness value of a solution t is simply the similarity between the two input images when registered using t, i.e. $f \to \Psi(I_M, t(I_S))$, where Ψ is a similarity metric and I_M, I_S are the scene and the model images ; in this study we considered MI.

Population The population of the GA has a fixed size. Initially, the individuals in the population are generated at random: the value of each gene is chosen from its valid range with uniform probability. At the end of a generation, the current population is entirely replaced by new one, except for the best individual that is never discarded (elitism).

2.2.2 Specific components

r-GA$^+$ also includes a number of specific components that exploit the features of the IR problem and are essential to improve the performance of our GA.

Image sampler So far, we assumed that all voxels of the images are used in the computation of the metric value. In general, this is not necessary and should be avoided due to its computational costs in large images. A subset is usually enough; commonly, the actual voxels are selected at random with uniform probability or taken along a regular grid. In the former case the subset can be either resampled at regular intervals or kept constant throughout the whole registration process. We adopted random sampling using the latter approach.

Multi-resolution strategy Before the registration is started, the input images are processed applying both down-sampling and gaussian smoothing in order to create two image pyramids. A few operations in the optimizer allow it to handle the change from one resolution to the next. The individuals in the population are left untouched; as we encoded transformations using physical units rather than voxels, the solutions are not affected by the change of the image's size. However, the quality of a solution changes between one resolution and the next due to the new details that are now visible in the images, therefore we need to compute the fitness of the solutions again.

Restart mechanism The motivation behind the use of restart is simple. At the end of the first resolution, the algorithm might have found a very low-quality transformation. Refining such transformation is unlikely to produce a good final solution, therefore is more appropriate to perform again the search for a suitable initial registration by restarting the algorithm. To check whether the best solution obtained at the end of the first resolution is acceptable or not, one might consider to set a threshold on its fitness value (i.e. its similarity metric value). However, the fitness value of an appropriate solution depends on the actual content of the input images and it is hard to predict. We adopted an alternative approach. The first resolution is performed a fixed number of times, independently of its outcome. At the end of this process, the best solutions found are considered for the second resolution.

The computational cost of performing a restart at the first resolution is quite low, as the images involved are still small. Note that if p is the probability of obtaining a low quality solution during the first resolution, performing n restarts reduces this probability to p^{n+1}, i.e. an exponential decrease. This observation provides a rule of

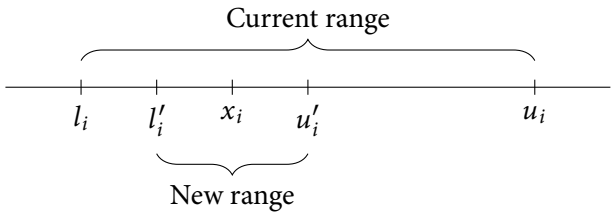


Figure 2.1: The dynamic boundary mechanism. At each resolution the transformation parameters ranges are restricted around the values in the best solution.

thumb to estimate an appropriate number of restarts. To bound to the probability of obtaining an unappropriate solution to a value r, just set

$$n \geq \frac{\ln(r)}{\ln(p)} - 1$$

Dynamic boundary This component is used to further take advantage of the use of multiple resolutions. The initial range of the transformation parameters are highly dependent on the application, therefore they should be provided by the user. However, as we change from one resolution to the next one, we can assume that the optimal solution for the next resolution lies in the same area of the search space as the best solution found at the current resolution. Therefore, we can improve the intensification of the search by limiting the range of the transformation parameters inside this area. This approach is called *dynamic boundary* [60]. Let x be the best solution found at the current resolution and l, u be respectively the vector of the lowest and highest values in the current range. Then, the range of the transformation parameters for the next resolution is $[l', u']$ where

$$l'_i = \max\{x_i - (u_i - l_i)/\gamma,\ l_i\}$$
$$u'_i = \min\{x_i + (u_i - l_i)/\gamma,\ u_i\}$$

and $\gamma > 1$ is the shrinking factor. Figure 2.1 shows an example of the process for $\gamma = 3$. Basically, dynamic boundary restricts the parameter ranges around the best solution, i.e. to a γ-times smaller interval centered around x. Note that even for modest values of γ, this can drastically reduce the size of the search space, depending on the number of transformation parameters. For instance, affine transformation has 12 parameters, meaning that for $\gamma = 2$, dynamic boundary reduces size of the search space by factor of $2^{12} = 4{,}096$.

2.3 Experimental Study

The aim of the experimentation is to carry out an objective comparison of the r-GA$^+$ proposal and other state-of-the-art IR methods. As competitors, we considered an heterogeneous group of algorithms to represent a wide range of approaches to the IR

Table 2.1: The IR algorithms included in the experimental study.

	Nature	Strategy	Optimizer	Ref.
I-ICP	feature	matching	Gradient Descent	[20]
Dyn-GA	feature	parameters	Genetic Algorithm	[60]
SS*	feature	matching	Scatter Search	[57]
m-SS	feature	parameters	Scatter Search	[49]
ASGD	intensity	parameters	Robbins–Monro	[72]

problem; they are listed in Table 2.1. Note that the algorithms differ in nature (feature- or intensity-based) as well as in the search strategy (based on matching or transform parameters). Also, different kind of optimization process are used: classic gradient-based techniques, evolutionary computation, metaheuristics.

For comparison purposes, we also tested a preliminary version of r-GA$^+$, called GA$^+$, which does not use dynamic boundary and has a fixed threshold for the restart policy. This test will highlight the role played by these two components in the performance of our approach.

We designed two experiments involving synthetic and real-world medical images. To make the comparison as objective as possible, the effectiveness of each method is assessed using a *quantitative* validation measure specific to each experiment. Furthermore, as most of the algorithms involved are of non-deterministic nature, we carried out a number of independent runs on each scenario. Our analysis investigate several aspects of the results. First, we measure the performance of the algorithms on each scenario by computing mean and standard deviation of the validation measure and ranking the algorithms accordingly. Next, we assess the overall performance of the algorithms in two ways: by computing the per-scenario mean rank of each algorithm and by counting the number of scenarios in which one outperforms another, called *wins*.

In the last part of the analysis, statistical tests are performed to determine which results are significantly different. We used the tests and the procedures recommended in [101] for comparing algorithms over multiple problems. We used non-parametric tests to avoid making (or testing) any assumption about the distribution of the results. The performance of r-GA$^+$ is compared with that of the remaining algorithms (i.e., a multiple comparison against a control method), a procedure that has more power than a pairwise comparison of all the algorithms. The test we used is Nemenyi's test [102], which is a post hoc procedure of Friedman's rank sum test [103] and is based on the ranks of the algorithms. As multiple comparison are performed, the p-values of the tests have been adjusted using Holm's method [104] in order to control the family-wise error rate.

For all algorithms, we used the original implementation by the authors. I-ICP, Dyn-GA and SS* are written in C, while r-GA$^+$ and ASGD are implemented in C++. Their running times can be fairly compared, as no bias due to the programming language or the execution environment is introduced. r-GA$^+$ has been written in C++

Table 2.2: The characteristics of the four brain MRI images used in the first experimental study (Figure 2.2). The noise value represents the percent ratio of the standard deviation of the white gaussian noise versus the signal of the brightest tissue. The number of crest line points, used as features, is also reported.

Image	Lesion	Noise	# of features
I_1	No	None	583
I_2	No	1%	393
I_3	Yes	1%	348
I_4	Yes	5%	248

and integrated in Elastix [105], a toolbox for intensity-based medical IR. Elastix is free, open-source and it has been used in over one hundred publications in medical imaging [106]. The software is built on top of the popular Insight Segmentation and Registration Toolkit (ITK) [107].

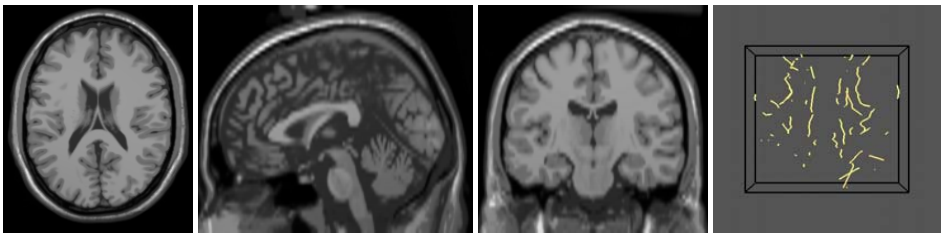
2.3.1 First experiment: registration of simulated brain MRIs

The first experiment is similar to the ones carried out in [57]. We used four simulated brain magnetic resonance images (MRIs) from a public database. A total of sixteen registration scenarios were artificially created by applying to the images a set of four large transformations. On those IR instances, we performed a comparison considering a large, heterogeneous group of IR algorithms.

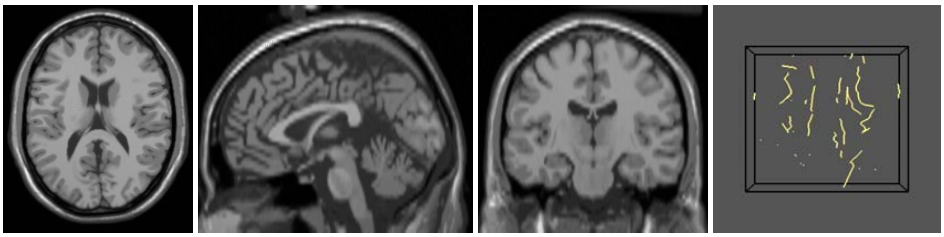
2.3.1.1 Setup

Images The images used in this experiment were obtained from the BrainWeb database at McGill University [108]. BrainWeb provides *simulated* brain MRI along with ground-truth data, therefore it can be easily used to evaluate the performance of various image analysis methods. Indeed, Brainweb has been frequently used by the IR research community [109]. To create scenarios with different difficulties, we added noise and multiple sclerosis lesions to some of the images, as detailed in Table 2.2. The images are shown in Figure 2.2; each image has size $60 \times 181 \times 217$ voxels.

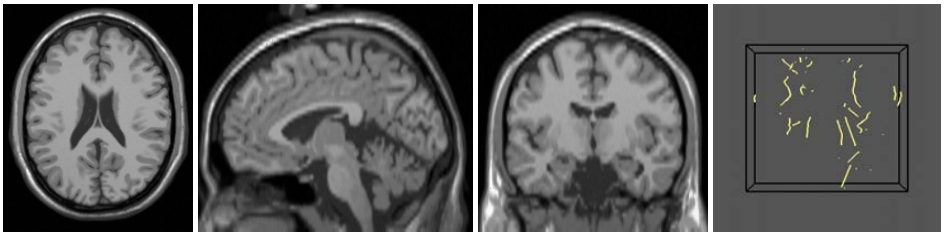
This experiment compares both feature- and intensity-based algorithms, thus some features need to be extracted from the images to provide an input for feature-based algorithms. As in the original comparison, we used crest line points, i.e. points where the surface normal has a sharp variation, detected through the approach described in [110]. SS* also uses the principal curvatures of the crest line points as heuristic information to guide the feature matching. Therefore, this information has been computed and provided to the algorithm. It is important to remark this difference: while the input of intensity-based methods consists of the whole images data (in case, two images made of $60 \times 181 \times 217 = 2,356,620$ voxels having an 8-bit intensity value), that of feature-based approaches is a set of just a few hundred of points (Table 2.2).



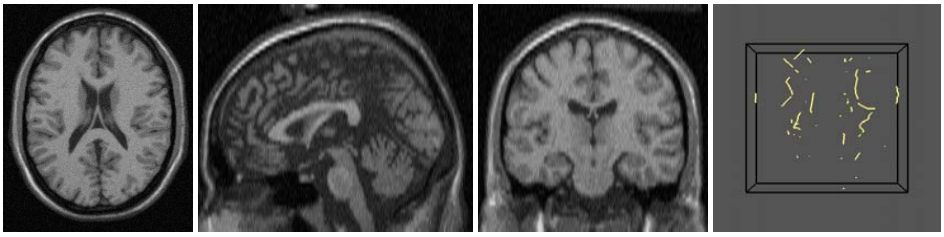
(a) I_1



(b) I_2



(c) I_3



(d) I_4

Figure 2.2: The four MRI brain images used in the first experiment. From left to right, axial, sagittal and coronal views, along with the corresponding crest line points used as features.

Table 2.3: Parameters of the similarity transformations we used in the experiments: rotation angle (λ), rotation axis (a_x, a_y, a_z), translation vector (t_x, t_y, t_z) and uniform scaling factor s.

	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

Registration scenarios Sixteen IR problem instances were created by choosing pairs of different images among the four available and applying one of the four similarity transformations shown in Table 2.3. Similarity transformations (Section 1.1.2) involve rotation, translation, and uniform scaling. The parameters values of the transformations were chosen to obtain large changes in the object location, orientation and scale. Changes of such magnitude are usually challenging for IR algorithms. The scenarios we considered in the experiments 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 $i = 1, 2, 3, 4$.

Algorithms and parameters settings The algorithms SS*, I-ICP and Dyn-GA were tested on this dataset in [57], therefore we used the parameter settings that delivered the best results in the original study. For m-SS, the parameters were manually adjusted for this experiment, starting from the recommended/default values provided by authors. As for r-GA$^+$, the parameters values were adjusted through a preliminary study that was performed using an additional registration scenario not included in the experimentation. The registration is performed in two resolutions; at the first resolution the images are smoothed (Gaussian smoothing, $\sigma = 4$) and downsampled by a factor of 4 in each dimension. The first resolution is repeated five times (i.e. four restart) independently of the results. The population was evolved for 50 generations in the first resolution and 25 in the second. The rest of the configuration was the same configuration in both resolutions: population size of 500 individuals, mutation probability of 0.1, crossover probability of 0.5, blend factor (α) 0.3 and tournament size equal to 3.

For all algorithms, the transformation model is similarity transform, and the transformation parameters ranges are $[-30, 30]$ for the translation component and $[0.75, 1.25]$ for the scaling factor. No restriction was applied to the rotation axis or to its magnitude. In r-GA$^+$, this transform is encoded by a real vector with seven elements representing versor of the rotation v_x, v_y, v_z, the translation component t_x, t_y, t_z and the scaling factor, with $v_x, v_y, v_z \in [-1, 1]$ and $s > 0$.

Validation procedure A suitable validation procedure also allows for comparison of different IR algorithms. However, if intensity- and feature-based methods are to be compared, the question arises on how to set up the experiments, in particular the time limit (or any other resource limit), for the comparison to be fair. Indeed, most basic

operations in an IR algorithm, such as applying a transformation or computing the similarity metric, are a lot faster in a feature-based method because the algorithm only needs to operate on a reduced version of the imaging data, e.g. a set of hundreds of points instead of full-resolution volumes having millions of voxels. On the other hand, to achieve that speed, feature-based methods benefited from a feature detection process, which might have been computationally expensive or even manually performed by an expert.

We designed a methodological framework for the comparison of feature- and intensity-based methods. The first step consist in choosing a common validation measure. As we mentioned earlier in Section 1.1.4, a similarity metric computed on features is preferable as it provides a real-word error measure. This might introduce some bias towards feature-based method using the same similarity metric, but as we will validate novel intensity-based algorithms, a bias for the competitors seems more appropriate than one in favor of the novel proposal. Then, we choose a time limit t_{stop} for the algorithms; the key idea is to consider the size of the imaging data used by the two classes of algorithms and compute their ratio

$$\alpha = \frac{\|\text{Image}\|}{\|\text{Features}\|}$$

This will give a rough approximation of the ratio between the costs of evaluating a solution for algorithms of different nature. Finally, we assign the original time limit t_{stop} to feature-based algorithm and $\alpha \cdot t_{\text{stop}}$ to intensity-based ones. This is the only part of the comparison procedure that changes according to the nature of the methods.

In this experimental study, intensity-based methods use a random subset of 25,000 voxels to compute the similarity metric, while the average number of features in the images is 393. This gives us a proportion α of roughly 60, which is used to set two time limits: 20 seconds for the feature-based approaches and 20 minutes for the intensity-based ones. In addition, we studied the behavior of the algorithms over time, to ensure these time limits allow them to reach convergence and deliver representative results. Figures 2.3 and 2.4 show examples of this analysis for the feature- and intensity-based approaches, respectively.

For each registration scenario we performed 15 independent runs of each algorithm. We used the MSE over the crestline points as validation measure. For the feature-based algorithms in the comparison, this is simply the similarity metric used by the algorithms. The solutions found by intensity-based algorithms were evaluated in the same way, i.e. by applying the obtained transformation to the scene's features and computing the MSE with respect to the model's features.

2.3.1.2 Analysis of results

Table 2.4 reports the results of the first experiment. For each scenario, we reported mean and standard deviation of the MSE values obtained by the algorithms along with their ranks. The average ranks (Table 2.5) and the count of wins (Table 2.6) provides another view of the results of the comparison.

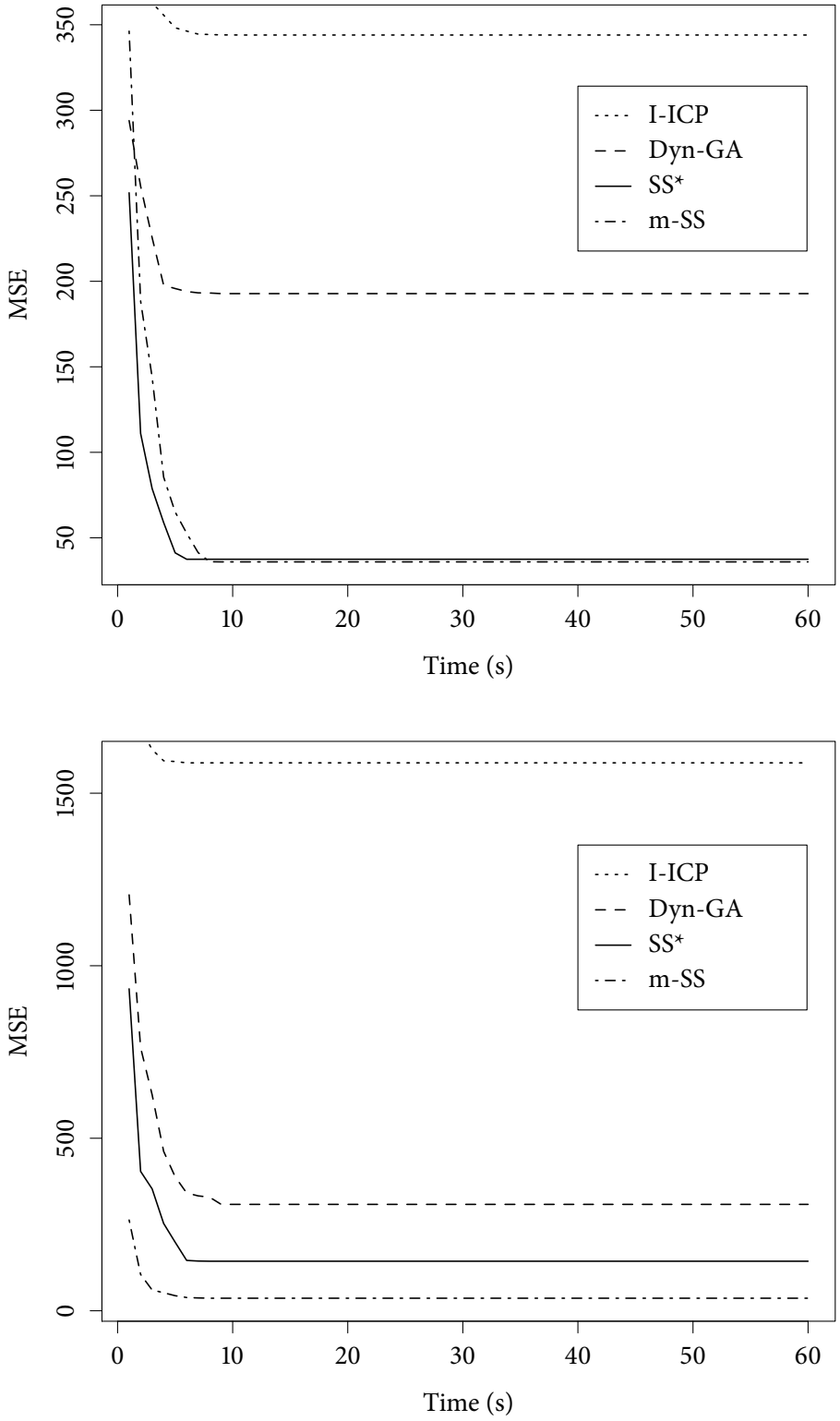


Figure 2.3: The MSE scored by the feature-based methods against running time in the first experimental study. Results are averaged over 15 runs and refer to scenarios I_1 vs $T_1(I_2)$ (top) and I_2 vs $T_4(I_4)$ (bottom).

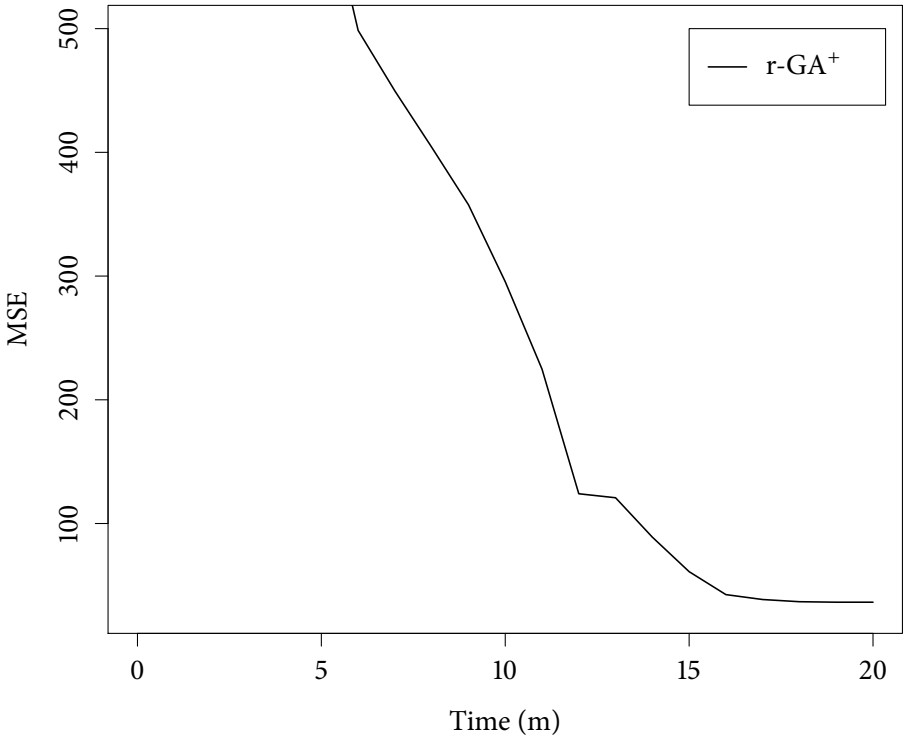


Figure 2.4: The results of ASGD and r-GA$^+$ against running time in the first scenario of the first experimental study. Results are averaged over 15 runs. The score of ASGD is above 50,000 and it has been omitted.

From the highest to the lowest average rank is ASGD, I-ICP, Dyn-GA, SS*, GA$^+$, m-SS and r-GA$^+$. ASGD scored the largest MSE values in all but one of the scenarios. Its performance varies greatly depending on the scenario, but in general the mean MSE is at least one order of magnitude away from the best solutions. I-ICP delivered a better, more steady performance, but still with very large MSE values. Dyn-GA scored better than I-ICP in all scenarios, with less variability between different scenarios, but the gap with the best results is large nevertheless. SS* scored constantly quite close to the best results, ranking third or second in 15 over 16 scenarios. GA$^+$ exhibits an inconsistent behavior. In 11 scenarios, it either scored best or extremely close to the best, while in the remaining 5 scenarios the mean MSE value is really large (>1000). As shown in Figure 2.5, the high average MSE is due to a few solutions having extremely high MSE. This points to the convergence problems we addressed with the new restart and the dynamic boundary mechanisms in r-GA$^+$. m-SS has a very stable behaviour: it ranked second or third in all but one scenario. m-SS was often outperfomed by GA$^+$ (9 cases) but its stability earned it the second best ranking (2.5). r-GA$^+$ consistently got either the best mean MSE (11 scenarios) or came really close (i.e. less than 1.0 from the best one). Also, the standard deviation values is always less than 3.0, confirming r-GA$^+$ is robust

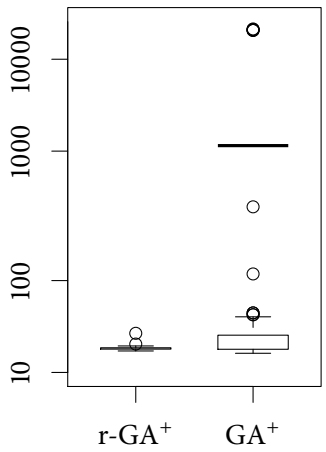


Figure 2.5: First experiment: boxplots of the results in the third scenario. While the majority of GA$^+$'s results are in an acceptable range, some solutions have very high MSE, explaining the high average MSE (the thick line) scored by the algorithm.

and no run of the algorithm has produced a low quality solution. r-GA$^+$ obtained the best ranking (1.38) and outperfomed all competitors in at least 11 scenarios, an amount that increases to 15 if we do not count GA$^+$.

Table 2.5 reports the p-value of Nemenyi's test comparing r-GA$^+$ against m-SS, GA$^+$, SS* and Dyn-GA. We included only the best ranking algorithms to avoid lowering the power of the test. In all four cases the test confirms the performance of r-GA$^+$ is significantly better than those of the competitors, with the highest p-value being that of m-SS, 0.0339.

2.3.2 Second experiment: atlas-based segmentation of real-world MRIs

In the second experiment we used real brain MRI images without applying any transformation. The registration is used to perform atlas-based segmentation of deep brain structures [111]. The quality of the segmentation obtained in this phase is used to assess the effectiveness of the registration methods.

Atlas-based segmentation is a procedure that aims to automatically delineate a region of an image using an atlas (or an ''average'' image) of a similar subject in which the desired region has been already segmented. The first step is to register the atlas (the scene) to the input image (the model). The transformation resulting from this phase is then used to overlap the segmented region of the atlas to the scene. The region of the scene that overlaps the segmented region of the atlas is the result of the segmentation process. Figure 2.6 illustrates the process. Often, atlas-based segmentation is used as

Table 2.4: Detailed results of the first experiment. For each scenario, the table reports the average MSE, standard deviation and ranking of the algorithms in the comparison.

	Algorithm	MSE mean	MSE sd	Rank		Algorithm	MSE mean	MSE sd	Rank
	ASGD	61816.4	795.7	7		ASGD	58146.8	661.0	7
	Dyn-GA	194.9	50.5	5		Dyn-GA	255.4	228.2	5
	r-GA$^+$	36.3	0.5	2		r-GA$^+$	53.1	0.2	2
1	GA$^+$	36.4	0.3	3	9	GA$^+$	52.9	0.3	1
	I-ICP	344.4		6		I-ICP	704.3		6
	m-SS	36.2	0.6	1		m-SS	59.9	1.9	3
	SS*	37.0	1.5	4		SS*	183.6	33.0	4
	ASGD	34773.9	238.0	7		ASGD	35695.3	2465.3	7
	Dyn-GA	107.5	52.1	5		Dyn-GA	163.1	57.5	4
	r-GA$^+$	36.5	0.5	1		r-GA$^+$	46.5	0.2	1
2	GA$^+$	36.7	0.4	3	10	GA$^+$	476.4	3648.3	5
	I-ICP	130.7		6		I-ICP	1493.2		6
	m-SS	36.6	0.6	2		m-SS	47.4	1.1	2
	SS*	43.4	3.6	4		SS*	89.2	40.8	3
	ASGD	111870.0	154.6	7		ASGD	111384.4	574.2	7
	Dyn-GA	211.0	137.3	4		Dyn-GA	224.9	87.3	4
	r-GA$^+$	41.7	1.6	1		r-GA$^+$	58.6	2.4	1
3	GA$^+$	1736.5	6216.7	6	11	GA$^+$	2823.9	7863.5	6
	I-ICP	894.3		5		I-ICP	951.3		5
	m-SS	42.7	1.7	2		m-SS	59.7	4.9	2
	SS*	63.2	2.9	3		SS*	82.2	45.1	3
	ASGD	1233.3	166.8	7		ASGD	885.8	356.7	7
	Dyn-GA	302.0	121.4	5		Dyn-GA	414.8	258.2	5
	r-GA$^+$	32.7	0.1	1		r-GA$^+$	47.7	0.2	2
4	GA$^+$	32.7	0.2	2	12	GA$^+$	47.3	0.4	1
	I-ICP	631.7		6		I-ICP	416.6		6
	m-SS	32.9	0.1	3		m-SS	56.0	7.1	3
	SS*	53.9	2.6	4		SS*	153.9	86.1	4
	ASGD	61063.5	309.1	7		ASGD	56932.0	568.6	7
	Dyn-GA	299.3	144.1	5		Dyn-GA	179.8	59.5	4
	r-GA$^+$	51.1	0.5	1		r-GA$^+$	35.5	0.3	2
5	GA$^+$	51.4	0.2	2	13	GA$^+$	35.0	0.2	1
	I-ICP	517.7		6		I-ICP	237.6		6
	m-SS	53.7	1.0	3		m-SS	40.4	2.3	3
	SS*	112.2	12.4	4		SS*	193.1	62.0	5
	ASGD	34796.2	223.8	7		ASGD	31521.1	6.6	7
	Dyn-GA	154.0	114.2	5		Dyn-GA	105.7	50.8	5
	r-GA$^+$	43.7	0.3	1		r-GA$^+$	30.5	0.2	1
6	GA$^+$	43.8	0.2	2	14	GA$^+$	30.7	0.3	2
	I-ICP	330.3		6		I-ICP	341.3		6
	m-SS	44.1	0.5	3		m-SS	32.2	1.9	3
	SS*	56.7	4.5	4		SS*	74.9	41.1	4
	ASGD	110131.2	1022.7	7		ASGD	112134.4	1027.4	7
	Dyn-GA	326.5	174.0	4		Dyn-GA	192.2	115.8	4
	r-GA$^+$	56.6	0.7	1		r-GA$^+$	40.7	1.2	1
7	GA$^+$	1091.8	4965.8	6	15	GA$^+$	1104.8	5128.7	6
	I-ICP	437.8		5		I-ICP	608.8		5
	m-SS	56.7	4.1	2		m-SS	45.0	5.9	2
	SS*	63.8	46.2	3		SS*	103.8	66.6	3
	ASGD	1017.4	252.7	7		ASGD	512.8	233.2	6
	Dyn-GA	354.3	146.9	5		Dyn-GA	298.1	144.8	5
	r-GA$^+$	45.2	0.1	2		r-GA$^+$	29.7	0.1	2
8	GA$^+$	44.5	0.3	1	16	GA$^+$	29.5	0.3	1
	I-ICP	478.0		6		I-ICP	1587.8		7
	m-SS	48.1	0.4	3		m-SS	37.2	5.1	3
	SS*	122.7	8.2	4		SS*	150.2	78.3	4

Table 2.5: First experiment: result of Nemenyi's test comparing r-GA$^+$ with the remaining algorithms. The table reports the average rankings of the algorithms and the adjusted p-value for each comparison.

Algorithm	Mean Rank	p-value
r-GA$^+$	1.38	
m-SS	2.50	0.0339
GA$^+$	3.00	0.0029
SS*	3.75	0.0001
Dyn-GA	4.62	-
I-ICP	5.81	-
ASGD	6.94	-

Table 2.6: First experiment: the number of scenarios in which the algorithm on the row has a better mean MSE value than that on the column.

	ASGD	Dyn-GA	r-GA$^+$	GA$^+$	I-ICP	m-SS	SS*
ASGD	-	0	0	0	1	0	0
Dyn-GA	16	-	0	5	16	0	1
r-GA$^+$	16	16	-	11	16	15	16
GA$^+$	16	11	5	-	12	9	11
I-ICP	15	0	0	4	-	0	0
m-SS	16	16	1	7	16	-	16
SS*	16	15	0	5	16	0	-

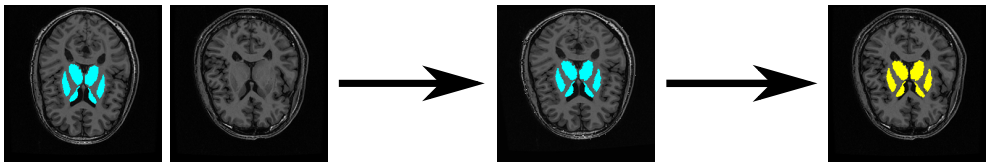


Figure 2.6: The atlas-based segmentation of deep brain structures in brain MRI. First, the atlas is registered to the input image and the resulting transformation is applied to the labelled region of the atlas (in blue). The resulting region is overlapped on the input image (in yellow) to determine the output of the process.

preliminary step in a more complex segmentation approach.

2.3.2.1 Setup

Images Thirteen T_1-weighted brain MRI were retrieved from the NMR database [112]. The deep nuclei structures in each image have been manually delineated by an expert in order to create the ground-truth data used to evaluate the registration. Figure 2.7 shows one of the images along with the corresponding deep nuclei.

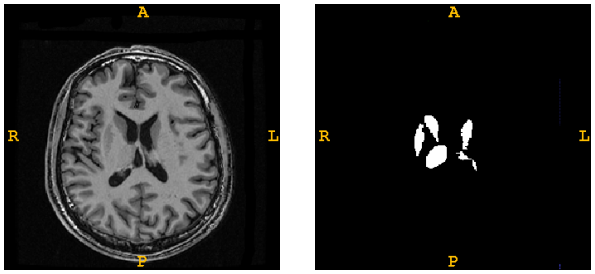


Figure 2.7: A slice of a 3D MRI brain image used in the second experiment (left) and the corresponding deep brain structure (right).

Registration scenarios Nine registration scenarios were created by selecting a pair of different images at random. No transformation was applied on the images; however, the location of the brain in each image is different due to the variability in the pose of the patient during the acquisition of the images.

Algorithms and parameters settings Given the nature of this experiment, we compare only intensity-based algorithms, i.e. r-GA$^+$, GA$^+$ and ASGD. The transformation model is affine transform, which involves rotation, translation, scaling and shearing, and it can be represented using 12 real parameters. Affine transform is a popular choice in registration of medical images [113]. It is flexible enough to present a wide range of transformations and it does not produce anatomically unrealistic results, as it could happen with deformable models. An affine transform (Section 1.1.2) is specified by a linear transformation (a 3×3 matrix $A = \{a_{i,j}\}$) and a translation vector t. In SS$^+$, both elements are combined in a vector $(a_{1,1}, a_{1,2}, \ldots, a_{3,2}, a_{3,3}, t_x, t_y, t_z)$. The valid range of each matrix element $a_{i,j}$ is computed form the limits to imposed over rotation, scaling and shearing.

In this experiment we have ground-truth data to evaluate the registration, but we do not know the concrete parameters values of the optimal registration transformations like in the first experimental study. Therefore, we estimated parameters values intervals considering a big enough range to include all registration solutions for this application. We allowed rotations between -90 and 90 degrees, scaling in the range $[0.9, 1.1]$, shearing in the interval $[-0.1, 0.1]$ and translations between -15 and 15 centimeters.

For r-GA$^+$, we kept the same configuration used in the first experiment. In a preliminary experimentation, we tested several configurations for ASGD varying the number of resolutions (2, 3 and 4) and iterations (500, 1000 and 2000). In what follows we report only the results obtained with the best configuration, which uses 4 resolutions and 1000 iterations.

The stopping criterion is a time limit of 10 minutes. Note that no synthetic transformation is applied on the images; the registration aims to compensate for the variability in the pose of the patient during the acquisition of the images. Therefore, with respect to the first experiment, the ''magnitude'' of the transformations to be found is much smaller. This results in a smaller search space, which in turn yields faster convergence

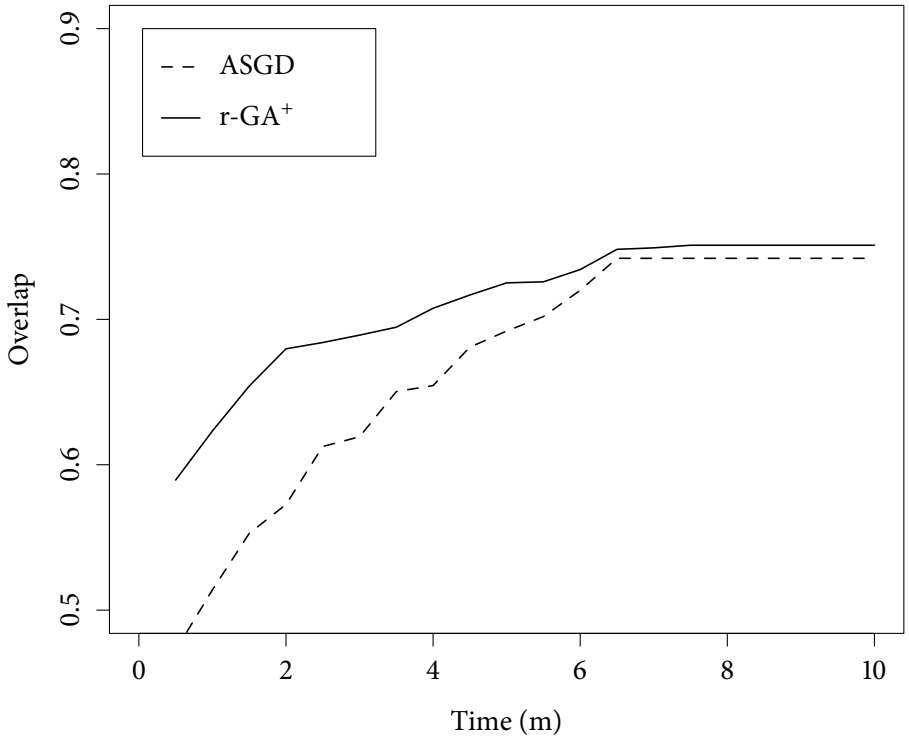


Figure 2.8: The average results of ASGD and r-GA$^+$ against running time in the first scenario of the second study.

for the algorithms, justifying the smaller time limit of this second experiment. Figure 2.8 shows an example of this phenomenon on the first scenario.

Validation procedure The quality of atlas-based segmentation depends closely on the accuracy of the registration step, although the anatomical variability of the target region can limit its effectiveness. In this experiment we validate the results of the registration algorithms by carrying out atlas-based segmentation of deep nuclei. For each scenario we performed 32 independent runs of each algorithm. The model image is used as atlas, while the scene is employed as input image. The segmented region obtained from the registration V_R is then compared with the ground-truth V_{GT}. The overlapping of the two regions is commonly measured using the Dice's coefficient [114], given by

$$\mathrm{Dice}(V_R, V_{\mathrm{GT}}) = \frac{2|V_R \cap V_{\mathrm{GT}}|}{|V_R| + |V_{\mathrm{GT}}|}$$

where $|\cdot|$ is the number of voxels. A value of 1 means perfect overlapping, while 0 means the two regions do not overlap at all.

2.3.2.2 Analysis of results

The results of the second experiment are reported in Table 2.7. We computed the mean and standard deviation of the overlap for each scenario. Table 2.9 shows the count of wins for the algorithms in the comparison.

The overlap values can differ considerably across the scenarios, reflecting the fact that the effectiveness of this kind of segmentation can vary depending on the concrete anatomy of the patients. ASGD and GA$^+$ had a similar performance. They have almost identical mean rank values (2.33 and 2.44) and a similar number of wins against each other (5 and 4). Again, GA$^+$ occasionally has quite large standard deviation values compared to the others, e.g. scenario number 2. r-GA$^+$ ranked first in 8 out of 9 scenarios and came second in the last one, delivering the best performance both in terms of ranking and number of wins. The results Nemenyi's test (Table 2.8) show the advantage of r-GA$^+$ over the other algorithms is statistically significant. The adjusted p-values of the tests are both 0.019.

2.4 Conclusions

In this chapter, 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.

The merit of this new approach is proved experimentally in two separate studies involving synthetic and real-world medical images. Each study included a comparison with other state-of-the-art IR methods using a wide range of approaches to the problem.

We proved that our genetic approach is competitive with the best IR techniques. Through a combination of MH and IR components, we addressed the initial convergence issues of r-GA$^+$ and designed a method able to outperform all other algorithms on the majority of the scenarios.

Table 2.7: Detailed results of the second experiment. For each scenario, the table reports the average overlap, standard deviation and ranking of the algorithms in the comparison.

	Algorithm	Overlap		Rank
		mean	sd	
1	ASGD	.742	.001	3
	r-GA$^+$	.755	.012	1
	GA$^+$	.751	.010	2
2	ASGD	.616	.005	2
	r-GA$^+$	.618	.007	1
	GA$^+$	.615	.033	3
3	ASGD	.677	.003	2
	r-GA$^+$	.679	.008	1
	GA$^+$	.676	.012	3
4	ASGD	.691	.001	3
	r-GA$^+$	.706	.016	1
	GA$^+$	.698	.011	2
5	ASGD	.756	.010	2
	r-GA$^+$	.760	.009	1
	GA$^+$	.755	.009	3
6	ASGD	.738	.003	2
	r-GA$^+$	.739	.005	1
	GA$^+$	.734	.011	3
7	ASGD	.686	.009	3
	r-GA$^+$	.729	.004	1
	GA$^+$	.717	.015	2
8	ASGD	.741	.001	3
	r-GA$^+$	.750	.007	2
	GA$^+$	.751	.020	1
9	ASGD	.754	.004	1
	r-GA$^+$	.749	.014	2
	GA$^+$	.745	.017	3

Table 2.8: Second experiment: result of Nemenyi's post-hoc procedure when comparing r-GA$^+$ with the remaining algorithms. The table reports the average rankings of the algorithms and the adjusted p-value for each comparison.

Algorithm	Mean Rank	p-value
r-GA$^+$	1.22	
ASGD	2.33	0.0190
GA$^+$	2.44	0.0190

Table 2.9: Second experiment: the number of scenarios in which the algorithm on the row has a better mean overlap value than that on the column.

	ASGD	r-GA$^+$	GA$^+$
ASGD	-	1	5
r-GA$^+$	8	-	8
GA$^+$	4	1	-

Chapter 3

Scatter Search for intensity-based medical image registration

In this Chapter, we present a novel IR algorithm based on SS. By implementing a new optimization procedure, we aimed to improve over our GA-based proposal and to create a faster, more accurate and robust IR algorithm. The choice of SS has a number of reasons. First, the results of our previous study, in which the two main competitors of our algorithm were both based on SS. Also, the excellent performance obtained by SS-based algorithms in experimental comparisons of feature-based IR algorithms in different contexts [47, 56, 10]. Furthermore, the fact that those algorithms used different approaches to IR, namely matching- and parameter-based, which indicates SS is so flexible that it can be tailored to deal with both combinatorial and continuous optimization.

The new algorithm is validated in an extended version of the two experimental comparisons presented in the previous chapter. Besides including additional algorithms and registration scenarios, the study presents a visual comparison of the results and a more comprehensive statistical analysis.

The Chapter is organized as follows. In Section 3.1 we review the SS optimization procedure and describe its adaptation to IR. Section 3.2 presents the experimental comparison used to validate our new proposal and the analysis of the results. Finally, Section 3.3 provides conclusions.

3.1 Scatter search

3.1.1 The Scatter Search optimization procedure

Scatter search was originally proposed by Glover in [34] in the context of integer programming. The main idea behind SS is to recombine systematically a *reference set* of high-quality solutions, possibly obtained using different techniques. Most of SS implementation to date have been based on the SS *templates* presented in [43] and [42];

our exposition follows the well-known five-methods template introduced in the latter. These methods are:

- a *diversification generation method*, that generates a set of diverse trial solutions;
- a *subset generation method*, by which sets of solutions from the reference set are created;
- a *solution combination method*, that combines a set of solutions into one or more new solutions.
- an *improvement method*, used to enhance a solution, usually applying an heuristic method;
- a *reference set update method* to build and maintain a reference set of solutions selected for their quality or diversity;

An outline of SS is shown in Figure 3.1, while Algorithm 5 shows a complete SS procedure [42]. At the beginning, a number of solutions are created, improved and stored in a temporary container *P*. The reference set is then updated by selecting the best solutions between *P* and the reference set itself. Next, the process enters a loop (line 8) that iterates the core SS procedure until a stopping condition is met, controlling the duration of the optimization. Then, an inner loop begins. First, subsets of solutions from the reference set are created. Each subset is combined into a new solution, which is then improved and stored in another container, *Pool*. Next, the reference set is updated with the best solutions among those in *Pool* and the current reference sets. If no new solution has entered the reference set, the inner loop ends. Finally, a new set *P* is created as in the beginning of the algorithm.

A popular variant of the canonical SS design is the *2-tier* design [42], in which the reference set is divided in two tiers. One tier, called the *quality* reference set, stores the b_1 most high quality solutions, while the other, the *diversity* reference set, contains the b_2 solutions having high diversity with respect to those in the first tier. Each tier is ordered according to either quality or diversity, and during the reference set update procedure the new solutions having the highest values of quality or diversity are considered for inclusion into the reference set. The advantage of this design is that it adds diversity to the search instead of focusing only around the best solutions found, avoiding premature convergence.

3.1.2 SS design for intensity-based IR

In what follows we present the design of the proposed IR method, called SS$^+$. The effectiveness of any SS implementation depends on a proper design of the previous five methods, that must be specific to the optimization problem at hand. However, just like r-GA$^+$, SS$^+$ has been designed to support multiple transformation models and similarity metrics.

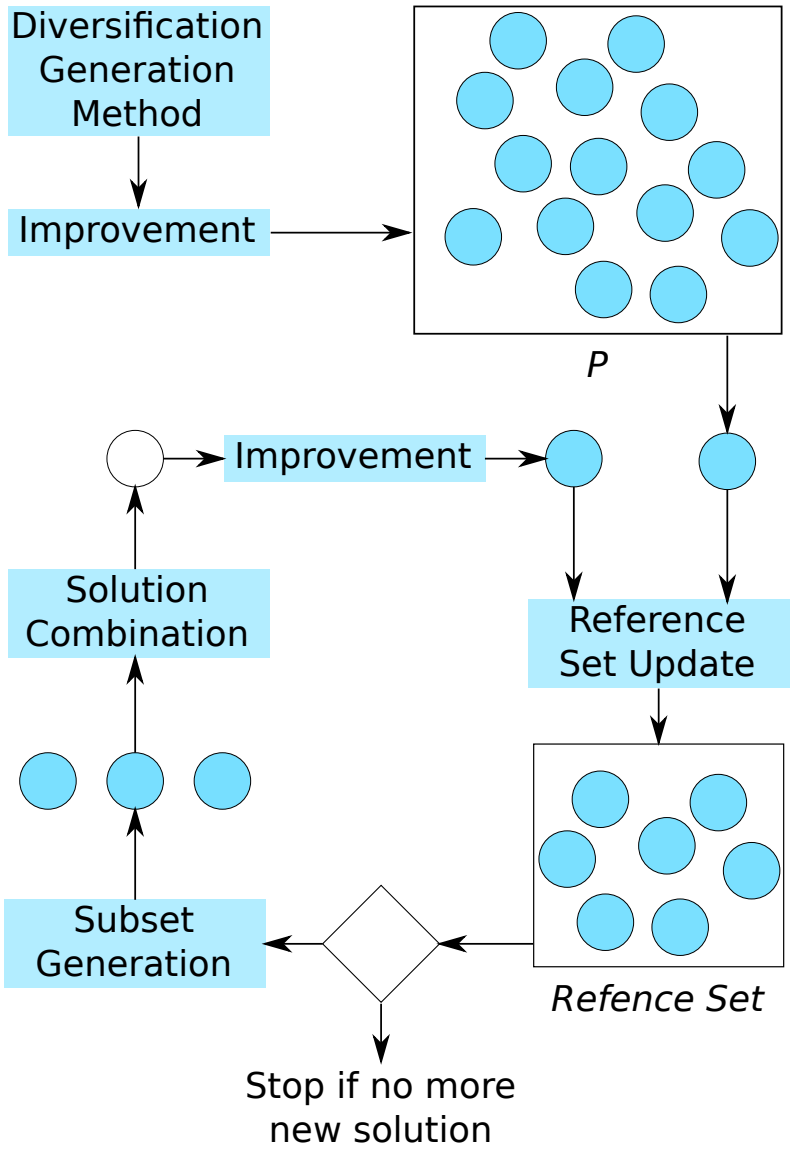


Figure 3.1: A graphical representation of the SS process.

Algorithm 5: A complete optimization procedure based on Scatter Search.

```
ReferenceSet ← ∅
P ← ∅
while |P| < PSize do
    x ← DiversificationGeneration()
    x′ ← ImprovementMethod(x)
    if x′ ∉ P then add x′ to P
end
while ¬ stop condition do
    Update the ReferenceSet by selecting the best b solutions in ReferenceSet ∪ P
    NewElements ← TRUE
    Pool ← ∅
    while NewElements do
        Subsets ← SubsetGeneration()
        NewElements ← FALSE
        while Subsets ≠ ∅ do
            S ← POP(Subsets)
            x ← SolutionCombination(S)
            x′ ← ImprovementMethod(x)
            if x′ ∉ Pool then Add x′ to Pool
        end
        Update the ReferenceSet by selecting the best b solutions in ReferenceSet ∪ Pool
        if ReferenceSet has new elements then
            NewElements ← TRUE
        end
        Build a new set P using the diversification generation and improvement
        methods
    end
end
```

Our efforts have been devoted to the design of the optimizer component. Our proposal is based on a SS algorithm with 2-tier design. First, we present the five SS methods especially designed to deal with IR as a continuous optimization problem.

Diversification generation method We adopt an approach based on frequency memory [115] to ensure the search space is explored in a uniform manner. The range of each transformation parameter is divided in four sub-ranges of equal size and a frequency counter is associated to each of them. A solutions is constructed in two steps. First, for each transformation parameter, a sub-range is chosen at random with a probability inversely proportional to its frequency count. Then, the actual values of the parameters are selected at random within the selected sub-range with uniform probability. Finally, the corresponding frequency counter is increased.

Subset generation method Our method generates subsets having two elements. First, the solutions from the quality reference set are considered, yielding $b_1(b_1-1)/2$

subsets. Then, the diversity reference set is used, so that other $b_2(b_2-1)/2$ subsets are generated. Finally, we combine solutions from the quality and diversity reference set, adding $b_1 b_2/2$ sets to the result.

Solution combination method To combine solutions we use the BLX-α crossover operator [100], a common choice in real-coded Evolutionary Algorithms. Two solutions x and y are provided as input. For each position i of their encoding, the operator computes the value $d_i = |x_i - y_i|$ and then it randomly generates a value z in the interval

$$[\quad \min(x_i, y_i) - \alpha d, \quad \max(x_i, y_i) + \alpha d \quad]$$

with uniform probability. The value z is assigned to the i-th positions of the resulting solution. The parameter α controls the width of the ranges in which the new values are drawn.

Improvement method Our improvement methods is a local search based on a crossover operator. To improve a solution x, we perform crossover with another randomly chosen solution from either the reference set or P to produce a number of new solutions; the best solution is then considered and the whole process is repeated for a number of iterations. If all the solutions generated by the crossover are worse than the original solution, the improvement method returns its original input. We use the ''parent-centric'' BLX-α crossover called PMX-α [116]. In contrast with the BLX-α, the ranges for the new solution parameters are

$$[\quad x_i - \alpha d, \quad x_i + \alpha d \quad]$$
$$[\quad y_i - \alpha d, \quad y_i + \alpha d \quad]$$

which indeed results in solutions that are closer to their parents. This operator has already been used in IR and it yielded the best results in the comparison of memetic approaches to range IR carried out in [49].

Reference set update method The update policy aims to maintain the highest quality solutions in the quality reference set and the most diverse solutions in the diversity one. The update is performed in three steps. First, a new quality reference set is created using the b_1 having the highest quality values in both the current quality reference set and the pool. Then, the algorithm computes a *diversity* measure over each solution in the pool and in the diversity reference set. Finally, the b_2 solutions having the highest distance value in both the pool and the diversity reference set are placed in the new diversity reference set.

The diversity value of a solution x measures the distance between a solution and the quality reference set S. It is defined as the minimum of the square difference between

x and each solution y in the quality reference set, as in

$$\text{diversity}(x, S) = \min \{ \text{ distance}(x, y) \mid y \in S \} \tag{3.1}$$

$$\text{distance}(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 \tag{3.2}$$

where n is the number of components of a solution.

Our SS$^+$ proposal also includes a number of more specific components that exploit the features of the intensity-based IR problem and are essential to improve the performance of SS. Some components are shared with our GA-based proposal from Chapter 2.

Duplication control A common issue in real-coded evolutionary algorithm is that extremely similar solutions can be generated. Despite being different in their representation, such solutions can be so close to each other that for all purposes they encode the very same object. This can cause the quality reference set to contain only almost identical copies of the same solution, which makes pointless the use of multiple solutions. We utilize a duplication control mechanism. We consider two solutions to be different only if their distance (Eq. 3.2) is above a given threshold. This condition is checked before the inclusion of a solution in a set (lines 6 and 19 in Algorithm 5).

Multi-resolution strategy Here we follow the same approach used with r-GA$^+$. Before the registration, the input images are processed applying both down-sampling and gaussian smoothing in order to create two image pyramids. To handle the change from one resolution to the next we just need to compute the fitness of the solutions. The solutions in the reference sets do not need any change as we encoded transformations using physical units and therefore the solutions are not affected by the change of the image's size. After the recomputation of the fitness, the process continues as if an iteration of the outermost loop just ended (line 8 in Algorithm 5).

Restart mechanism and dynamic boundary The motivation of these two components is described in Chapter 2. At the end of the first resolution, the algorithm might have found a very low-quality transformation and refining such transformation is unlikely to produce a good final solution. Therefore, the algorithm perform the first resolution a predefined number of times, independently of its outcome, and at the end of this process, the best solutions found are considered for the second resolution. When switching to the finer resolutions, a dynamic boundary mechanism improves the intensification of the search by limiting the range of the transformation parameters inside this area where the best solutions lie. We use the following formulas for the range of the transformation parameters l, u:

$$l_i' = \max\{x_i - (u_i - l_i)/\gamma,\ l_i\}$$
$$u_i' = \min\{x_i + (u_i - l_i)/\gamma,\ u_i\}$$

where $\gamma > 1$ is the shrinking factor and x is the best current solution.

3.2 Experimental study

The experimental study in this chapter is an extension of that carried out in the previous chapter. Besides adding SS$^+$ to the comparison, we included two additional tools for the analysis of the results. The first is a multiple sign-test, which compares the algorithms in a pairwise fashion using the number of wins and losses. Sign-test is essentially a modified binomial test, and it provides an additional way to detect significant differences in the performance of the algorithms. Second, we included a visual comparison of the average registration results.

Before going into the actual study, we present some preliminary results that aim to highlight the crucial role played by each of the four specific IR components introduced in Section 3.1.2 on the performance of SS$^+$. To this end, we selected a simple and a complex scenario, as evaluated according to the quality of the images and the magnitude of the transformation involved. Then, we tested SS$^+$ enabling different combinations of components, starting with all the components being disabled and incrementally adding them one at the time. The results are shown in Figure 3.2. In both scenarios, each component brings a noticeable improvement on the results, which can span across a few orders of magnitude. In particular, restart has a major effect on the variability of the results, so it is beneficial in terms of robustness, while dynamic boundary affects mostly the accuracy. An analogous pattern was observed throughout the study.

3.2.1 Registration of simulated brain MRIs

For this experiment we used exactly the same setup of Section 2.3.1 for what concerns images, registration scenarios and parameter settings for all algorithms in the previous comparison. As for SS$^+$, the registration is performed in two resolutions; at the first resolution the images are smoothed (gaussian smoothing, $\sigma = 4$) and downsampled by a factor of 4 in each dimension. The first resolution is repeated twice (i.e. one restart) independently of the results. The algorithm used the same configuration in both resolutions: five elements in the quality and in the diversity reference set, P's size set to 32, blend factor $\alpha = 0.3$ for the solution combination method and 50 iterations of PMX with $\alpha = 0.5$ for the improvement method. Mutual information was used as similarity metric.

For the stopping criterion we kept the same time limit: 20 seconds for feature-based approaches and 20 minutes for intensity-based ones. We checked that this time limit allows also SS$^+$ to reach convergence, as shown in Figure 3.3.

3.2.1.1 Results

Table 3.2 reports the results of the first experiment. For each scenario, we reported mean and standard deviation of the MSE values obtained by the algorithms along with their ranks. The average ranks (Table 3.3) and the count of wins (Table 3.4) provide a basic view of the results of the comparison. We also include a visual comparison of the average results of four scenarios (Table 3.1).

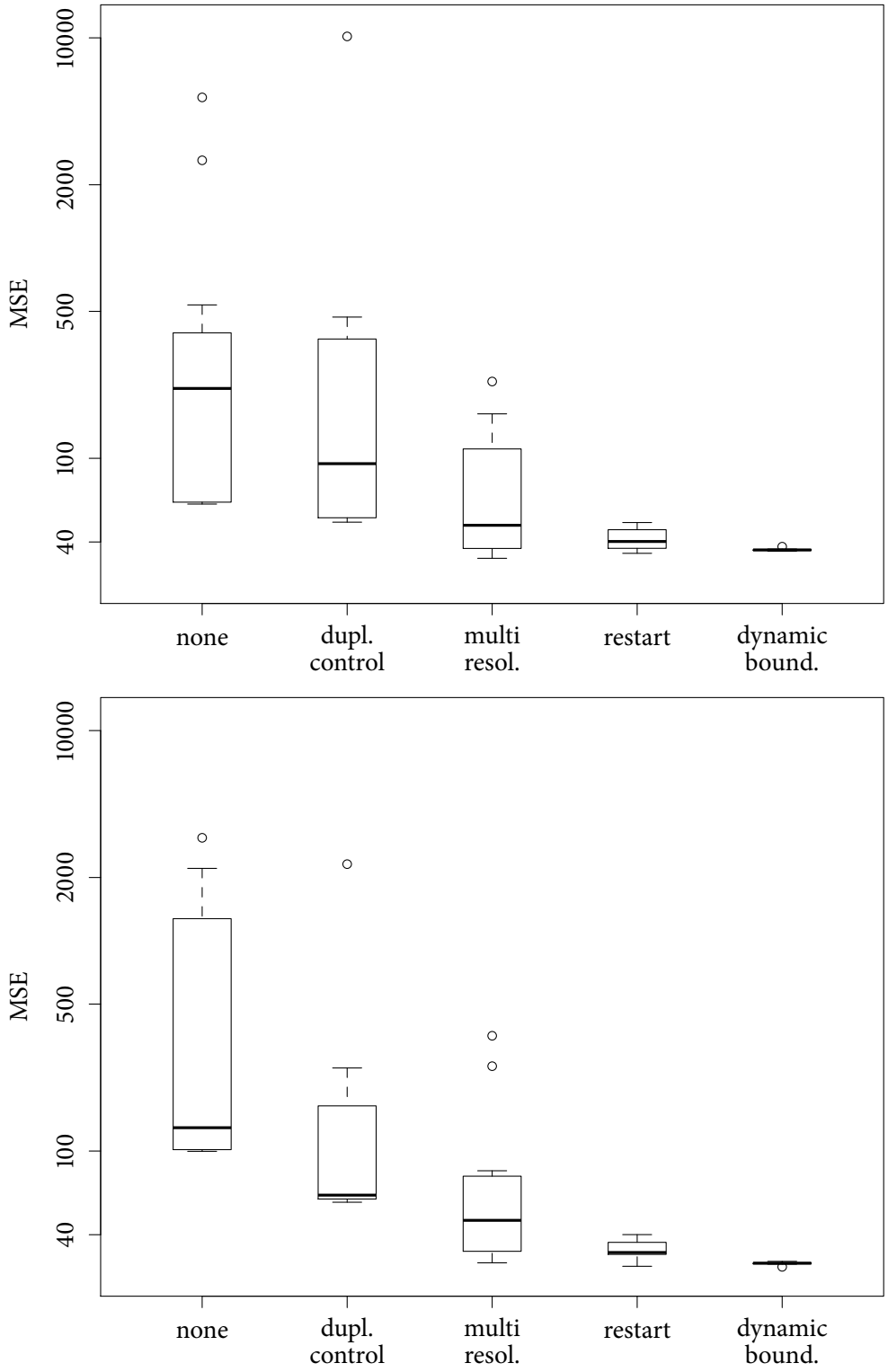


Figure 3.2: The effect of the specific IR components of SS$^+$ on its performance. The results refer to the first experimental study, scenarios I_1 vs $T_1(I_2)$ (top) and I_2 vs $T_4(I_4)$ (bottom). In each plot, the components are enabled in an incremental fashion from left to right, e.g. the boxplot labeled ''restart'' shows the results of SS$^+$ using duplication control, multiple resolutions and restart. Logarithmic scale is used to show the improvement brought by each component despite the differences in their order of magnitude.

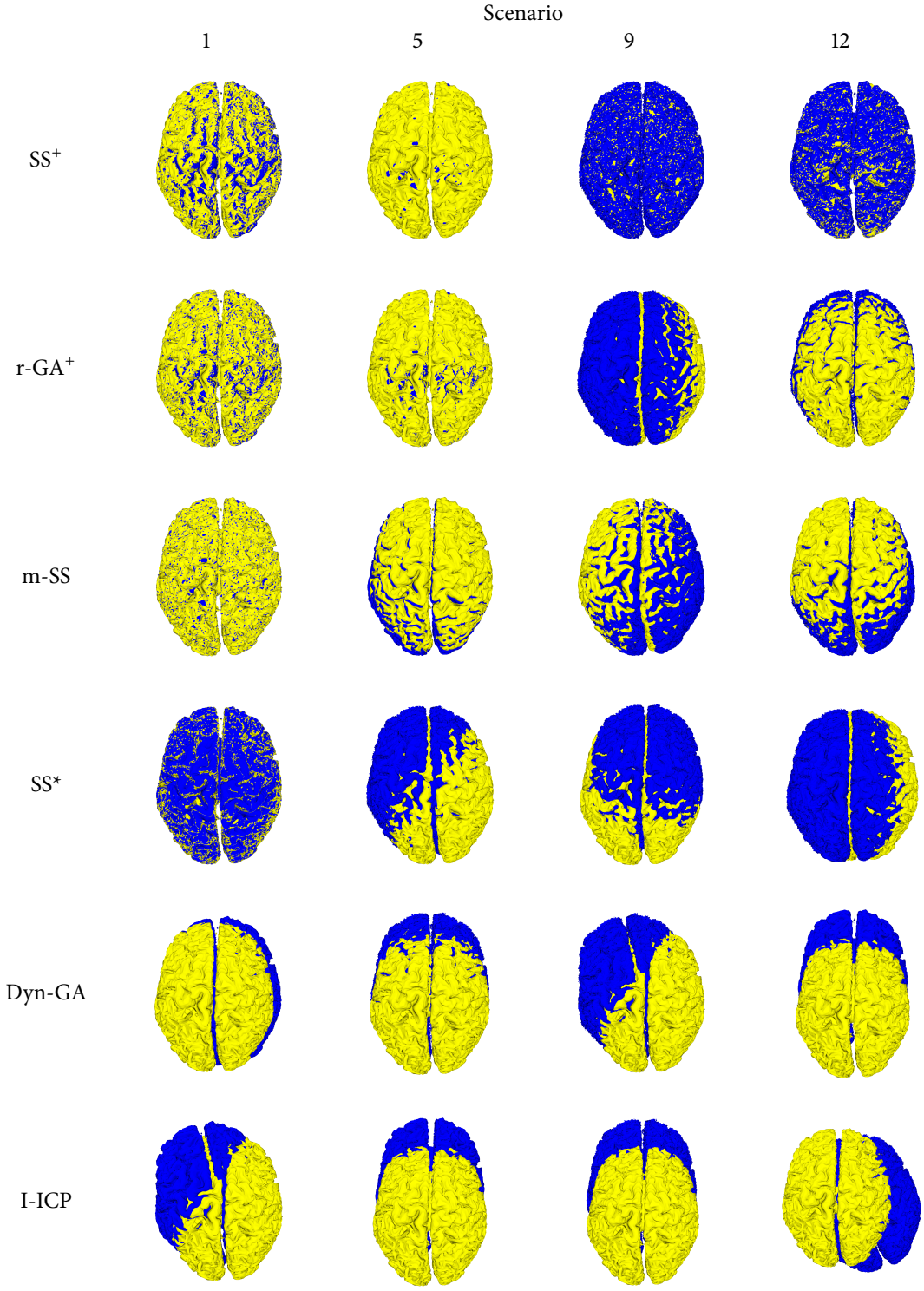


Table 3.1: A visual comparison of the results of the first experiment. Each entry shows the solution having the MSE value closest to the median MSE scored by the corresponding algorithm. The scene is coloured in blue, while the model is in yellow.

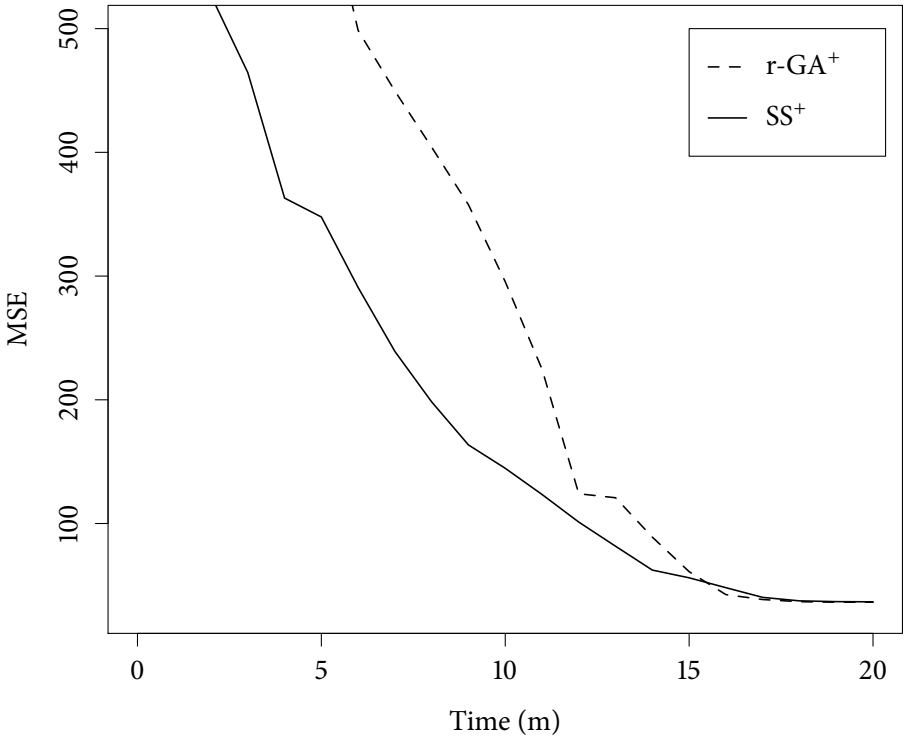


Figure 3.3: The results of ASGD, r-GA$^+$ and SS$^+$ against running time on the first scenario of the first experimental study. Results are averaged over 15; the score of ASGD is above 50,000 and it has been omitted.

From the highest to the lowest error ranking is ASGD, I-ICP, Dyn-GA, SS*, m-SS, r-GA$^+$ and SS$^+$. ASGD scored the largest MSE values in all but one of the scenarios. Its performance varies greatly depending on the scenario, but in general the mean MSE is at least one order of magnitude away from the best solutions. I-ICP delivered a better, more steady performance, but still with very large MSE values. Comparing the mean values, Dyn-GA scored better than I-ICP in all scenarios, with less variability between different scenarios, but the gap with the best results is large nevertheless. Also, the standard deviation is often comparable with the mean, meaning that the quality of the solutions varies a lot even in the same scenario. SS* delivered a good performance: it scored constantly quite close to the best results and it ranking fourth in all but one scenario.

Overall, m-SS scored the third average rank, 2.88, after r-GA$^+$ and SS$^+$. It ranked third in all but one scenario, showing a stable behavior. Its solution are generally very close to the best in the comparison, but still it was outperformed by both SS$^+$ and r-GA$^+$ in 15 scenarios. r-GA$^+$ and SS$^+$ scored similar results in terms of MSE, but SS$^+$ was consistently better than the other. Indeed, r-GA$^+$ was outperformed in all but one scenarios, ranking 2.0, while SS$^+$ has a better ranking, 1.12, resulting from delivering

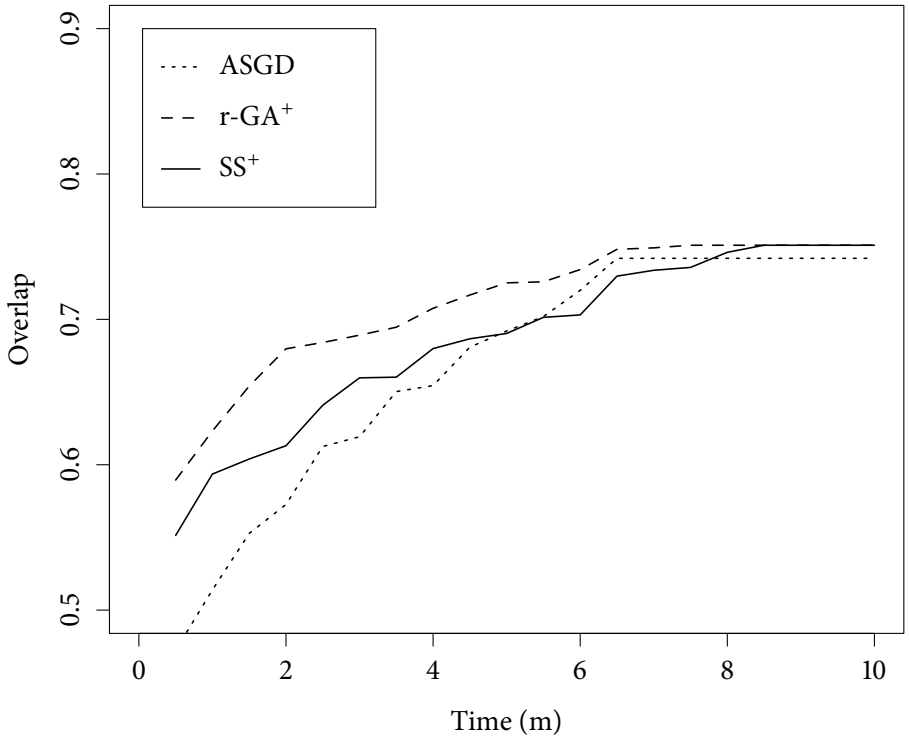


Figure 3.4: The average results of ASGD, r-GA$^+$ and SS$^+$ against running time on the first scenario of the second experimental study.

the best average solutions in 15 over 16 cases. It is also the algorithm with the lowest variability, which is constantly below 0.1.

Table 3.3 reports the p-value of Nemenyi's test comparing SS$^+$ against r-GA$^+$, m-SS and SS*. We included only the best ranking algorithms to avoid lowering the power of the test. In all three cases the test confirms the performance of SS$^+$ is significantly better than the competitors, with the highest p-value being that of r-GA$^+$, 0.0133. The sign-test comparing the number of wins (Table 3.5) has similar results. All algorithms have a p-value of 0.001 or less, confirming the superiority of SS$^+$'s performance at an even higher degree of confidence.

3.2.2 Atlas-based segmentation of real-world MRIs

This experiment follows the same setup of Section 2.3.2, but we created additional registration scenarios for a total of 18. The configuration of SS$^+$ is that same one used in the first experiment. Likewise, the time limit of 10 minutes allows SS$^+$ to reach convergence, as shown in Figure 3.4.

	Algorithm	MSE mean	MSE sd	Rank		Algorithm	MSE mean	MSE sd	Rank
	ASGD	61816.4	795.7	7		ASGD	58146.8	661.0	7
	Dyn-GA	194.9	50.5	5		Dyn-GA	255.4	228.2	5
	r-GA$^+$	36.3	0.5	3		r-GA$^+$	53.1	0.2	2
1	I-ICP	344.4		6	9	I-ICP	704.3		6
	m-SS	36.2	0.6	1		m-SS	59.9	1.9	3
	SS*	37.0	1.5	4		SS*	183.6	33.0	4
	SS$^+$	36.3	0.0	2		SS$^+$	52.3	0.1	1
	ASGD	34773.9	238.0	7		ASGD	35695.3	2465.3	7
	Dyn-GA	107.5	52.1	5		Dyn-GA	163.1	57.5	5
	r-GA$^+$	36.5	0.5	2		r-GA$^+$	46.5	0.2	2
2	I-ICP	130.7		6	10	I-ICP	1493.2		6
	m-SS	36.6	0.6	3		m-SS	47.4	1.1	3
	SS*	43.4	3.6	4		SS*	89.2	40.8	4
	SS$^+$	36.5	0.0	1		SS$^+$	46.2	0.0	1
	ASGD	111870.0	154.6	7		ASGD	111384.4	574.2	7
	Dyn-GA	211.0	137.3	5		Dyn-GA	224.9	87.3	5
	r-GA$^+$	41.7	1.6	2		r-GA$^+$	58.6	2.4	2
3	I-ICP	894.3		6	11	I-ICP	951.3		6
	m-SS	42.7	1.7	3		m-SS	59.7	4.9	3
	SS*	63.2	2.9	4		SS*	82.2	45.1	4
	SS$^+$	40.7	0.0	1		SS$^+$	57.1	0.0	1
	ASGD	1233.3	166.8	7		ASGD	885.8	356.7	7
	Dyn-GA	302.0	121.4	5		Dyn-GA	414.8	258.2	5
	r-GA$^+$	32.7	0.1	2		r-GA$^+$	47.7	0.2	2
4	I-ICP	631.7		6	12	I-ICP	416.6		6
	m-SS	32.9	0.1	3		m-SS	56.0	7.1	3
	SS*	53.9	2.6	4		SS*	153.9	86.1	4
	SS$^+$	32.3	0.0	1		SS$^+$	46.6	0.0	1
	ASGD	61063.5	309.1	7		ASGD	56932.0	568.6	7
	Dyn-GA	299.3	144.1	5		Dyn-GA	179.8	59.5	4
	r-GA$^+$	51.1	0.5	2		r-GA$^+$	35.5	0.3	2
5	I-ICP	517.7		6	13	I-ICP	237.6		6
	m-SS	53.7	1.0	3		m-SS	40.4	2.3	3
	SS*	112.2	12.4	4		SS*	193.1	62.0	5
	SS$^+$	50.8	0.0	1		SS$^+$	34.6	0.1	1
	ASGD	34796.2	223.8	7		ASGD	31521.1	6.6	7
	Dyn-GA	154.0	114.2	5		Dyn-GA	105.7	50.8	5
	r-GA$^+$	43.7	0.3	2		r-GA$^+$	30.5	0.2	1
6	I-ICP	330.3		6	14	I-ICP	341.3		6
	m-SS	44.1	0.5	3		m-SS	32.2	1.9	3
	SS*	56.7	4.5	4		SS*	74.9	41.1	4
	SS$^+$	43.5	0.1	1		SS$^+$	30.7	0.1	2
	ASGD	110131.2	1022.7	7		ASGD	112134.4	1027.4	7
	Dyn-GA	326.5	174.0	5		Dyn-GA	192.2	115.8	5
	r-GA$^+$	56.6	0.7	2		r-GA$^+$	40.7	1.2	2
7	I-ICP	437.8		6	15	I-ICP	608.8		6
	m-SS	56.7	4.1	3		m-SS	45.0	5.9	3
	SS*	63.8	46.2	4		SS*	103.8	66.6	4
	SS$^+$	55.6	0.0	1		SS$^+$	40.0	0.0	1
	ASGD	1017.4	252.7	7		ASGD	512.8	233.2	6
	Dyn-GA	354.3	146.9	5		Dyn-GA	298.1	144.8	5
	r-GA$^+$	45.2	0.1	2		r-GA$^+$	29.7	0.1	2
8	I-ICP	478.0		6	16	I-ICP	1587.8		7
	m-SS	48.1	0.4	3		m-SS	37.2	5.1	3
	SS*	122.7	8.2	4		SS*	150.2	78.3	4
	SS$^+$	43.8	0.0	1		SS$^+$	28.9	0.0	1

Table 3.2: Detailed results of the first experiment. For each scenario, the table reports the average MSE, standard deviation and ranking of the algorithms in the comparison.

Algorithm	Mean Rank	p-value
SS$^+$	1.12	
r-GA$^+$	2.00	0.0133
m-SS	2.88	0.0000
SS*	4.06	0.0000
Dyn-GA	4.94	-
I-ICP	6.06	-
ASGD	6.94	-

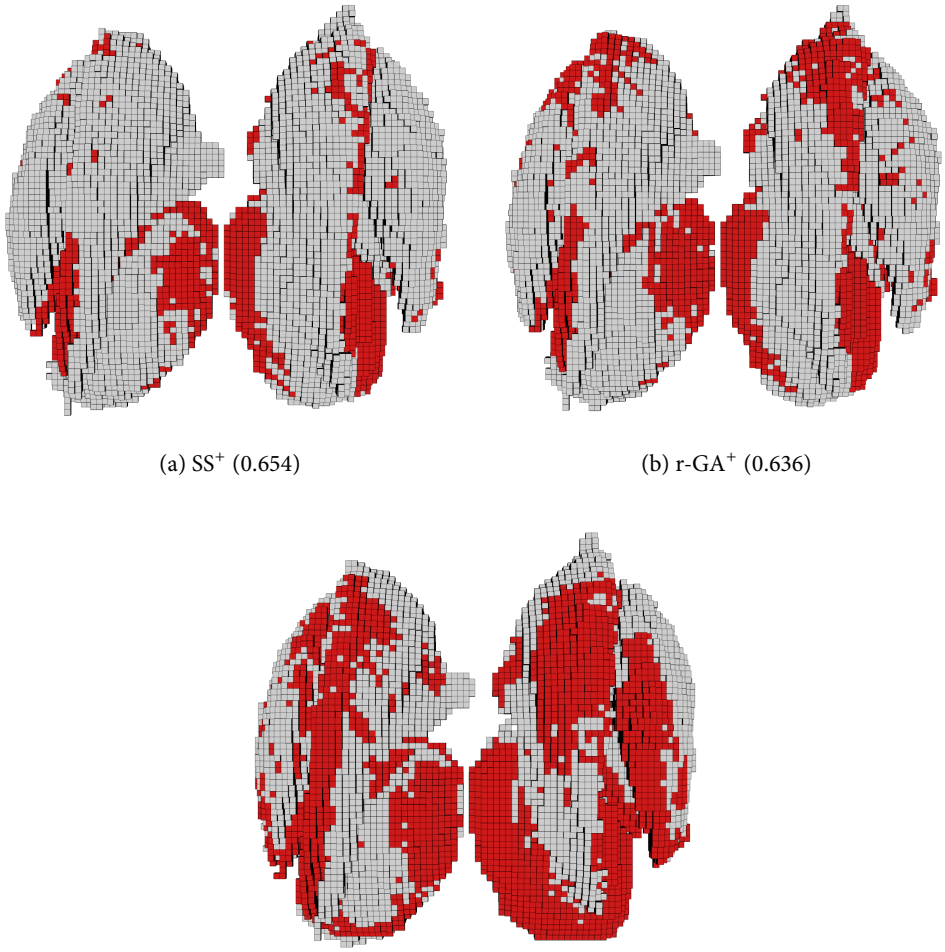
Table 3.3: First experiment: result of Nemenyi's test comparing SS$^+$ with the remaining algorithms. The table reports the average rankings of the algorithms and the adjusted p-value for each comparison.

	ASGD	Dyn-GA	r-GA$^+$	I-ICP	m-SS	SS*	SS$^+$
ASGD	-	0	0	1	0	0	0
Dyn-GA	16	-	0	16	0	1	0
r-GA$^+$	16	16	-	16	15	16	1
I-ICP	15	0	0	-	0	0	0
m-SS	16	16	1	16	-	16	1
SS*	16	15	0	16	0	-	0
SS$^+$	16	16	10	16	14	16	-

Table 3.4: First experiment: the number of scenarios in which the algorithm on the row has a better mean MSE value than that on the column.

Algorithm	Losses	p-value
ASGD	16	0.0002
Dyn-GA	16	0.0002
r-GA$^+$	15	0.0010
I-ICP	16	0.0002
m-SS	15	0.0010
SS*	16	0.0002

Table 3.5: First experiment: result of sign test comparing SS$^+$ with the other algorithms. The table lists the algorithms along with their number of scenarios in which they have been outperformed by SS$^+$ (Table 3.4, bottom row) and the associated adjusted p-value.



(a) SS$^+$ (0.654)

(b) r-GA$^+$ (0.636)

(c) ASGD (0.619)

Figure 3.5: Second experiment. A visualization of the overlapping between the automatically segmented volume (red) and the ground truth (white). The figures refers to the seventh scenario. The solutions used to create the figures are those having the closest overlap value to the mean result of the corresponding algorithm. In parenthesis is the overlap value of each solution.

3.2.2.1 Results

The results of the second experiment are reported in Table 3.6. We computed the mean and standard deviation of the overlap and ranked the algorithm accordingly for each scenario. Tables 3.7 and 3.8 show the mean ranks and the count of wins for the three algorithms in the comparison. Finally, Figure 3.5 shows a comparison of the average results in the seventh scenario.

The overlap values can differ considerably across the scenarios, reflecting the fact

that the effectiveness of this kind of segmentation can vary depending on the concrete anatomy of the patients. In general, r-GA$^+$ and ASGD have similar results, whereas SS$^+$ ranked constantly better than the others. In contrast with the previous experiment, even though it ranked last on average, ASGD delivered an acceptable performance. This is due to the lower ''magnitude'' of the transformations involved, more suitable for a local search method. However, note that the low variability of its overlap values means almost all solutions have a lower quality than the average solution found by the other two algorithms. r-GA$^+$ had a similar performance: it has almost the same mean rank and it ranked better than ASGD in slightly more than half of the scenarios (11 cases).

The performance of SS$^+$ is the best one of the comparison. Our proposal outperformed the other algorithms both in terms of mean rank (1.39 against 2.22 and 2.39) and number of wins (14 and 15 against r-GA$^+$ and ASGD, respectively). The results Nemenyi's test, reported in Table 3.7, shows the difference with both algorithms is statistically significant to a high degree (0.0124 for r-GA$^+$ and 0.0054 for ASGD). The sign-test (Table 3.9) confirms the thesis, although with slightly higher p-values (0.0151 for r-GA$^+$ and 0.0309 for ASGD).

3.3 Conclusions

We have described the design and implementation of SS$^+$, a novel technique to solve medical IR using SS. We used an advanced, 2-tier SS design, and developed the five components of the SS template specifically to tackle IR. In addition, we integrated a multi-resolution strategy with two optimization components. First, the restart mechanism, that allows the algorithm to deliver a more robust performance at the cost of a low extra computational effort. Second, the dynamic boundary, that focuses the search on the appropriate region of the search space as the optimization progresses.

The validation of the new method extends the two experimental comparisons already introduced in the previous Chapter. In both studies, SS$^+$ delivered the best mean performance, with a statistically significant improvement over all the competitors.

	Algorithm	Overlap mean	Overlap sd	Rank		Algorithm	Overlap mean	Overlap sd	Rank
	ASGD	0.742	0.001	3		ASGD	0.738	0.003	2
1	r-GA$^+$	0.751	0.010	2	10	r-GA$^+$	0.734	0.011	3
	SS$^+$	0.751	0.008	1		SS$^+$	0.740	0.011	1
	ASGD	0.679	0.003	1		ASGD	0.717	0.011	3
2	r-GA$^+$	0.669	0.019	3	11	r-GA$^+$	0.736	0.012	2
	SS$^+$	0.673	0.026	2		SS$^+$	0.750	0.013	1
	ASGD	0.616	0.005	1		ASGD	0.684	0.005	3
3	r-GA$^+$	0.615	0.033	2	12	r-GA$^+$	0.704	0.024	2
	SS$^+$	0.596	0.026	3		SS$^+$	0.716	0.022	1
	ASGD	0.677	0.003	1		ASGD	0.686	0.009	3
4	r-GA$^+$	0.676	0.012	3	13	r-GA$^+$	0.717	0.015	2
	SS$^+$	0.677	0.008	2		SS$^+$	0.718	0.014	1
	ASGD	0.682	0.000	1		ASGD	0.680	0.005	3
5	r-GA$^+$	0.671	0.006	2	14	r-GA$^+$	0.713	0.011	1
	SS$^+$	0.670	0.005	3		SS$^+$	0.693	0.019	2
	ASGD	0.691	0.001	3		ASGD	0.741	0.001	3
6	r-GA$^+$	0.698	0.011	2	15	r-GA$^+$	0.751	0.020	2
	SS$^+$	0.722	0.018	1		SS$^+$	0.769	0.023	1
	ASGD	0.621	0.007	3		ASGD	0.751	0.012	3
7	r-GA$^+$	0.635	0.015	2	16	r-GA$^+$	0.769	0.017	2
	SS$^+$	0.652	0.027	1		SS$^+$	0.779	0.011	1
	ASGD	0.756	0.010	2		ASGD	0.754	0.004	2
8	r-GA$^+$	0.755	0.009	3	17	r-GA$^+$	0.745	0.017	3
	SS$^+$	0.773	0.012	1		SS$^+$	0.756	0.020	1
	ASGD	0.634	0.006	3		ASGD	0.624	0.004	3
9	r-GA$^+$	0.656	0.014	2	18	r-GA$^+$	0.672	0.044	2
	SS$^+$	0.670	0.022	1		SS$^+$	0.689	0.030	1

Table 3.6: Detailed results of the second experiment. For each scenario, the table reports the average overlap, standard deviation and ranking of the algorithms in the comparison.

Algorithm	Mean rank	p-value
SS$^+$	1.39	
r-GA$^+$	2.22	0.0124
ASGD	2.39	0.0054

Table 3.7: Second experiment: result of Nemenyi's post-hoc procedure when comparing SS$^+$ with the remaining algorithms. The table reports the average rankings of the algorithms and the adjusted p-value for each comparison.

	ASGD	r-GA$^+$	SS
ASGD	-	7	4
r-GA$^+$	11	-	3
SS$^+$	14	15	-

Table 3.8: Second experiment: the number of scenarios in which the algorithm on the row has a better mean overlap value than that on the column.

Algorithm	Losses	p-value
ASGD	14	0.0309
r-GA$^+$	15	0.0151

Table 3.9: Second experiment: result of sign test comparing SS$^+$ with ASGD and r-GA$^+$. The table lists the algorithms along with their number of scenarios in which they have been outperformed by SS$^+$ (Table 3.8, bottom row) and the associated adjusted p-value.

Chapter 4

Automatic offline parameter tuning for image registration

In general, the effectiveness of optimization algorithms can be heavily dependent on the setting of their parameter values. This poses a series of challenges for both the design and the comparison of algorithms. Finding appropriate settings is a complex task, and for many years, this was done manually, using a trial-and-error approach with preliminary experiments. As a consequence, algorithm development requires a lot of expertise and it is very time consuming for algorithm designers. Additionally, when comparing algorithms, the effort invested in finding appropriate parameter settings for each algorithm may be very different, and the potential uneven tuning may compromise the ability of experimental studies to assess the intrinsic quality of an algorithm.

In recent years, several methods have been proposed for automatically configuring algorithms parameters, a process also called parameter *tuning* [117, 118, 119, 120]. These methods are increasingly acknowledged in the research community for being important to find effective parameters. However, these automatic algorithm configuration methods, or tuners, are still very rarely used for the comparison of algorithms. In this chapter, we use such methods to assess the performance of various algorithms for IR. The procedure we use for this task is the following. First, we automatically configure the different algorithms we compare, and from this process we obtain parameter configurations that have been found to be the best by the tuner. Then, in the actual comparison, we use these parameter settings instead of the default parameter settings that have been proposed in the literature. In this way, we expect to compare the different algorithms at stake without introduction of human bias.

Our goal is tow-fold. First, we want to assess the applicability of our IR methods through a comparison with other established techniques in the medical imaging field. Second, we want to demonstrate the potential of automatic configuration techniques for comparing algorithms. More generally, to the best of our knowledge, this work is also the first to apply automatic configuration techniques to the processing of medical images.

This chapter is structured as follows. In Section 4.1, we first review the automatic

configuration of algorithms and the tool we employed. In Section 4.2 we report the setup of both the tuning and the comparison of tuned algorithms, followed by the experimental results and their analysis. Finally, Section 4.3 concludes the chapter.

4.1 Automatic Configuration of the Algorithms

The automatic configuration of algorithms, also called automatic *tuning*, has received a strong attention in recent research. In particular, in recent years a number of new algorithmic tools for the automatic configuration of algorithms have been developed. These include methods such as ParamILS [117], SMAC [118], SPO [119], SPO$^+$ [121], or iterated race [122, 120], which is available as an R [123] package and which we used in this study. In this section, we explain the goal of the automatic configuration process, and we briefly overview the tool that we used to perform this process.

4.1.1 Offline Automatic Configuration

Automatic parameter configuration can be done *online* to set the values of the parameters while the algorithm is running. This online adaptation of parameters is sometimes referred to as parameter adaptation. It is typically applied only to a small subset of key parameters, since it implies a significant overhead for the algorithm to ''learn'' good values for the parameters, in addition to the exploration of the search space of the instance to be tackled.

In this study, we use *offline* automatic configuration. In this case, the purpose is to automatically configure optimization algorithms before they are deployed, that is, before they are applied to instances that are not yet known. Two clearly delimited phases are involved in this process. In a primary tuning phase, an algorithm configuration is selected, given a set of training instances. In a secondary production (or testing) phase, the selected algorithm configuration is used to solve unknown instances of the same problem. The goal is to find, during the tuning phase, an algorithm configuration that optimizes some cost measure over the set of instances that will be seen during the production phase. In other words, the ultimate purpose is that the high-quality configuration of the algorithm found during the tuning phase generalizes to similar but unknown instances.

4.1.2 The IRACE software package

Birattari et al. [124, 125, 126] proposed an automatic configuration approach, F-Race, based on *racing* [127], with the use of Friedman's non-parametric two-way analysis of variance by ranks, to test for significantly inferior candidate configurations. This proposal was later improved by repeating the race process, refining iteratively the sampling distribution. The resulting automatic configuration approach was called Iterated F-race, and formally described in (I/F-Race) [122, 128]. However, no implementation of it has been made publicly available at that time. Later, the `irace` package has been proposed, that implements a general *iterated racing* procedure, which includes I/F-Race as a

special case. It also implements several extensions, some described in [125, 126], such as the use of the paired t test instead of Friedman's test. Several original contributions were also implemented to improve further the effectiveness of the tuning procedure. For more details, the reader can refer to [120].

The `irace` package has already been extensively tested in several successful research projects. For instance, Dubois-Lacoste et al. [129, 130, 131] used `irace` for tuning new state-of-the-art algorithms for the permutation flow-shop scheduling problem. However, to the best of our knowledge, this study is the first to apply an automatic configuration tool to image registration, and to the field of medical images, in general.

The advantage of the `irace` tool is that it handles several parameter types: continuous, integer, categorical, and ordered. Continuous and integer parameters take values within a range specified by the user. Categorical parameters can take any value among a set of possible ones explicitly given by the user. An ordered parameter is a categorical parameter with a pre-defined strict order of its possible values. We also relied on `irace`'s capability to parallelize the configuration process in order to reduce considerably the amount of time required for it.

4.2 Experimental Comparison

In this study we consider our proposal SS$^+$ together with r-GA$^+$ and four gradient-based medical IR methods introduced in Section 1.3.2.1: gradient descent (GD), quasi-Newton (QN), nonlinear conjugate gradient (NCG) and adaptive stochastic gradient descent (ASGD). The target IR application is the atlas-based segmentation of deep brain structures described in Section 2.3.2. We matched the setup used in the previous studies except that a larger number of instances is considered. The experimentation consist of two phases. At the beginning, the IR algorithms are tuned. The best configuration of each algorithm is then used in the actual comparison. In addition, the comparison includes a version of each algorithm using its default settings.

4.2.1 Tuning Setup

The study is performed using 10-fold cross validation. We used 70 instances, that were randomly partitioned in 10 subsets. We performed 10 independent comparisons; for each comparison, all algorithms were tuned using 9 subsets of instances (that is, 63 instances) for the tuning phase, and the remaining subset (7 instances) was used for the comparison of the configurations obtained from the tuning. That is, each instance is used for testing in exactly one independent comparison, and used for training in all the others. In this way, we perform 10 repetitions of the whole tuning process for each algorithm, always keeping a clear separation between the training and the testing instances, so the algorithm automatically tuned are compared on instances that were never seen during the tuning.

We allowed a tuning budget of 1000 evaluations, i.e. `irace` can call the algorithm to be tuned a maximum of 1000 times to find the best possible configuration.

In our comparison, we compare both deterministic and stochastic algorithms. Unlike deterministic algorithms, which are evaluated during the tuning process through a single run, stochastic ones (r-GA$^+$, SS$^+$ and ASGD) are run 15 times independently and the average cost is returned as the result of the evaluation to the tuner. Table 4.1 presents the list of parameters that were automatically configured, with their types. In case of real or integer parameters, the table shows the ranges considered as candidate values, and for categorical variable all possible values are given explicitly.

The comparison also includes a version of each algorithm using its default parameters settings. The suffix ''-def'' is used after the name of the algorithm to label those untuned versions, leading to the algorithms SS$^+$-def, r-GA$^+$-def, GD-def, QN-def, NCG-def and ASGD-def. The default settings, recommended in the corresponding literature, are presented in Table 4.2. In addition to r-GA$^+$ and SS$^+$, all classic comparison algorithms are written in C++ and integrated in Elastix [105].

4.2.2 Comparison setup

Following the 10-fold cross validation setup, the algorithms are tested on the testing data using the configurations obtained from the tuning. Non-deterministic algorithms are run 30 times on each instance, and their average overlap value is used in the comparison with deterministic algorithms. Our analysis examines several aspects of the results using the same methodology of the previous studies. The per-instance performance is evaluated by ranking the algorithms according to their overlap value. The overall performance is assessed by computing the average ranking over all the instances. In addition, we count the number of wins, to allow a pairwise comparison. In the final part of the analysis, statistical tests are used to determine which results are significantly different.

4.2.3 Results

The detailed results of the experiments for 6 of the 70 instances are reported in Table 4.3. The mean overlap and standard deviation values for each instance are reported. The average ranks (Table 4.4) and the count of wins (Table 4.5) provides an overall view of the results of the comparison.

4.2.3.1 Overall effect of the tuning

First, we discuss the impact of the tuning on the performance of the algorithms, with respect to their default settings (Table 4.2). As shown in Table 4.4, the tuning leads to a clear improvement in terms of average ranking for all algorithms. Also, when each tuned configuration is compared to the default one in terms of number of instances (see Table 4.5), we can observe that all tuned algorithms performed better than their default version in more than half of the cases. SS$^+$ and r-GA$^+$ benefited the most of the tuning, improving their performances in 57 out of 70 cases. ASGD improved in 53 cases, which is remarkable given that the algorithm uses an online adaptation mechanism for its parameters and, thus, was not a priori supposed to be subject to improvements from

Table 4.1: List of the parameters that are automatically configured. Given is the name of the parameter, the type and the domain of the parameters.

Parameter	Type	Domain
	Scatter Search	
PSize	integer	$[20, 40]$
b1	integer	$[4, 8]$
b2	integer	$[4, 8]$
BLX-alpha	real	$[0, 1]$
MaximumNumberOfIterations	integer	$[5, 15]$
Restarts	integer	$[2, 8]$
NumberOfResolutions	integer	$[2, 4]$
	Genetic Algorithm	
CrossoverProbability	real	$[0.5, 0.9]$
MutationProbability	real	$[0.05, 0.2]$
TournamentSize	integer	$[2, 6]$
MaximumNumberOfIterations	integer	$[25, 100]$
Restarts	integer	$[2, 8]$
NumberOfResolutions	integer	$[2, 4]$
	Gradient Descent	
NumberOfResolutions	integer	$[2, 5]$
MaximumNumberOfIterations	integer	$[500, 2000]$
a	integer	$[400, 1600]$
A	integer	$[50, 200]$
α	real	$[0.5, 0.7]$
	Quasi-Newton	
NumberOfResolutions	integer	$[2, 5]$
MaximumNumberOfIterations	integer	$[500, 2000]$
LBFGSUpdateAccuracy	integer	$[20, 50]$
	Nonlinear Conjugate Gradient	
NumberOfResolutions	integer	$[2, 5]$
MaximumNumberOfIterations	integer	$[500, 2000]$
ConjugateGradientType	categorical	{SteepestDescent, FletcherReeves PolakRibiere, DaiYuan, HestenesStiefel, DaiYuanHestenesStiefel}
	Adaptive Stochastic Gradient Descent	
NumberOfResolutions	integer	$[2, 5]$
MaximumNumberOfIterations	integer	$[500, 2000]$

Table 4.2: Parameter settings of the default version of the algorithms.

Parameter	Value
Scatter Search (SS$^+$-def)	
`PSize`	32
`b1`	5
`b2`	5
`BLX-alpha`	0.3
`MaximumNumberOfIterations`	10
`Restarts`	2
`NumberOfResolutions`	2
Genetic Algorithm (r-GA$^+$-def)	
`PopulationSize`	50
`CrossoverProbability`	0.7
`MutationProbability`	0.1
`TournamentSize`	3
`MaximumNumberOfIterations`	50
`Restarts`	5
`NumberOfResolutions`	2
Gradient Descent (GD-def)	
`NumberOfResolutions`	3
`MaximumNumberOfIterations`	1000
`a`	400
`A`	50
α	0.602
Quasi-Newton (QN-def)	
`NumberOfResolutions`	2
`MaximumNumberOfIterations`	1000
`LBFGSUpdateAccuracy`	5
Nonlinear Conjugate Gradient (NCG-def)	
`NumberOfResolutions`	3
`MaximumNumberOfIterations`	1000
`ConjugateGradientType`	DaiYuanHestenesStiefel
Adaptive Stochastic Gradient Descent (ASGD-def)	
`NumberOfResolutions`	3
`MaximumNumberOfIterations`	1000

Table 4.3: Detailed results of the experiment for the first six instances. For each instance, the table reports the average overlap, the standard deviation and the ranking of the algorithms in the comparison.

	Algorithm	Overlap mean	Overlap sd	Rank
1	ASGD	0.762	0.001	4
	ASGD-def	0.755	0.002	9
	NCG	0.756		8
	NCG-def	0.757		7
	r-GA$^+$	0.771	0.008	2
	r-GA$^+$-def	0.759	0.015	5
	GD	0.758		6
	GD-def	0.740		11
	QN	0.742		10
	QN-def	0.708		12
	SS$^+$	0.774	0.007	1
	SS$^+$-def	0.763	0.010	3
2	ASGD	0.301	0.005	12
	ASGD-def	0.680	0.003	6
	NCG	0.702		1
	NCG-def	0.598		11
	r-GA$^+$	0.686	0.016	5
	r-GA$^+$-def	0.676	0.014	8
	GD	0.701		2
	GD-def	0.669		9
	QN	0.686		4
	QN-def	0.628		10
	SS$^+$	0.690	0.016	3
	SS$^+$-def	0.679	0.011	7
3	ASGD	0.272	0.043	12
	ASGD-def	0.595	0.005	7
	NCG	0.616		4
	NCG-def	0.503		10
	r-GA$^+$	0.629	0.026	2
	r-GA$^+$-def	0.611	0.023	5
	GD	0.589		8
	GD-def	0.575		9
	QN	0.488		11
	QN-def	0.614		6
	SS$^+$	0.637	0.018	1
	SS$^+$-def	0.617	0.016	3
4	ASGD	0.734	0.001	5
	ASGD-def	0.723	0.002	8
	NCG	0.707		9
	NCG-def	0.635		12
	r-GA$^+$	0.744	0.012	2
	r-GA$^+$-def	0.731	0.018	4
	GD	0.727		7
	GD-def	0.729		6
	QN	0.689		11
	QN-def	0.702		10
	SS$^+$	0.747	0.010	1
	SS$^+$-def	0.735	0.013	3
5	ASGD	0.705	0.002	5
	ASGD-def	0.694	0.002	10
	NCG	0.711		4
	NCG-def	0.676		11
	r-GA$^+$	0.713	0.012	3
	r-GA$^+$-def	0.697	0.011	8
	GD	0.724		1
	GD-def	0.699		9
	QN	0.702		6
	QN-def	0.637		12
	SS$^+$	0.717	0.007	2
	SS$^+$-def	0.700	0.008	7
6	ASGD	0.676	0.005	6
	ASGD-def	0.658	0.006	10
	NCG	0.382		12
	NCG-def	0.672		7
	r-GA$^+$	0.693	0.016	4
	r-GA$^+$-def	0.696	0.024	2
	GD	0.665		9
	GD-def	0.678		5
	QN	0.670		8
	QN-def	0.637		11
	SS$^+$	0.697	0.011	3
	SS$^+$-def	0.703	0.011	1

Table 4.4: Result of Nemenyi's post-hoc procedure when comparing SS$^+$ with the other algorithms. The table reports the average rankings of the algorithms and the adjusted p-value for each comparison.

Algorithm	Mean Rank	p-value
SS$^+$	3.19	
r-GA$^+$	4.44	0.0140
GD	5.10	0.0004
SS$^+$-def	5.89	0.0000
ASGD	6.54	0.0000
GD-def	6.66	0.0000
QN	6.91	0.0000
r-GA$^+$-def	7.09	0.0000
QN-def	7.64	0.0000
NCG	8.30	0.0000
NCG-def	8.37	
ASGD-def	7.87	

the tuning. GD, QN and NCG improved in 49, 38 and 37 cases, respectively. Overall, the use of automatic tuning lead to improvement in 57 + 57 + 53 + 49 + 38 + 37 = 291 cases over 420. We tested this result using a binomial test to assess its significance, which confirms the benefits of the tuning, with a p-value of 1.776×10^{-15}.

4.2.3.2 Comparison of the tuned algorithms

The overlap values can differ considerably across the instances, reflecting the fact that the effectiveness of this kind of segmentation can vary depending on the concrete anatomy of the patients. From the highest (worst) to the lowest (best) average rank we have the order NCG, QN, ASGD, GD, r-GA$^+$ and SS$^+$. NCG delivered the worst performance of the group: it was outperformed by all other tuned algorithms in 43 over 70 scenarios and scored the worst average ranking (8.3).

QN and ASGD have similar average ranking (6.91 and 6.54, respectively), but rather different behaviors. While the results of QN are quite consistent throughout the instances, ASGD occasionally delivered a mean overlap value below half of those of the other algorithms, as can be seen in the case of instances 2 and 3 in Table 4.3. It is interesting to point out that this inconsistent behavior of ASGD was not observed at all in previous comparisons without tuning (Chapters 2 and 3). GD scored consistently close to the best results, scoring the third best mean ranking (5.10) after r-GA$^+$ and SS$^+$, and beating the remaining algorithms in 43 instances. We remark that under tuning, this ''simplest'' algorithm among those compared is able to deliver a rather good performance.

r-GA$^+$ placed second with a ranking of 4.44. Although in no single scenario r-GA$^+$ delivered the best result, its solutions are usually pretty close to the highest value. Finally,

Table 4.5: The number of instances in which the algorithm on the row has a better mean overlap value than that on the column.

	ASGD	ASGD-def	NCG	NCG-def	r-GA$^+$	r-GA$^+$-def	GD	GD-def	QN	QN-def	SS$^+$	SS$^+$-def
ASGD	-	53	46	42	15	32	26	41	41	41	13	28
ASGD-def	17	-	44	40	13	29	14	27	31	37	10	26
NCG	24	26	-	37	14	28	16	22	27	32	11	25
NCG-def	28	30	33	-	11	29	21	19	33	29	11	20
r-GA$^+$	55	57	56	59	-	57	42	51	45	50	0	57
r-GA$^+$-def	38	41	42	41	13	-	26	31	36	41	9	0
GD	44	56	54	49	28	44	-	49	43	52	23	41
GD-def	29	43	48	51	19	39	21	-	38	44	16	33
QN	29	39	43	37	25	34	27	32	-	38	22	31
QN-def	29	33	38	41	20	29	18	26	32	-	18	26
SS$^+$	57	60	59	59	70	61	47	54	48	52	-	57
SS$^+$-def	42	44	45	50	13	70	29	37	39	44	13	-

Table 4.6: Result of sign test. The table lists the algorithms along with their number of instances in which they have been outperformed by SS^+ (Table 4.5) and the associated adjusted p-value.

Algorithm	Losses	p-value
ASGD	56	0.0000
ASGD-def	57	0.0000
NCG	58	0.0000
NCG-def	61	0.0000
r-GA^+	70	0.0000
r-GA^+-def	60	0.0000
GD	46	0.0115
GD-def	52	0.0002
QN	47	0.0112
QN-def	53	0.0001
SS^+-def	57	0.0000

SS^+ performed better than all other algorithms in 47 of 70 instances and it reached the lowest average rank (3.19), delivering, thus, the best overall performance. Remarkably, SS^+ outperformed its best competitor r-GA^+ in all 70 scenarios. The significance of this result is confirmed by two statistical tests. Table 4.4 reports the p-value of Nemenyi's test comparing SS^+ against the best ranking algorithms. In all cases, the test confirms that the performance of SS^+ is significantly better than those of the competitors, with the highest p-value being that of r-GA^+, 0.014. The sign-test with respect to the number of wins (Table 4.6) confirms that the difference between SS^+ and the others algorithms is significant with even higher confidence.

4.3 Conclusion

In this chapter, we compared our proposed algorithms against four other well-established medical IR algorithms. Unlike many comparisons of algorithms proposed in the literature, we relied on automatic configuration techniques to set appropriately the parameter values of all the algorithms at hand, a crucial aspect of their effectiveness. Our experimental setup consisted of two phases. In a first training phase, we applied an automatic configuration tool [120] to all algorithms involved in the comparison. Our experimental results showed a consistent improvement through the automatic configuration procedure with respect to the default parameter settings. In the second phase, the testing phase, we compared all algorithms with the settings that were obtained in the first phase. This comparison is made on unseen instances, that is, instances that are different from the training instances used in the first phase. Our experimental comparison, based on statistical tests to assess the significance of the observed differences, shows that SS^+ performs significantly better than the other algorithms in terms of the quality of the results.

Chapter 5

Biomedical image segmentation using deformable models and metaheuristics

In this Chapter, IR is used as a component in a medical image segmentation approach based on level set. This new geometric deformable model combines region- and edge-based information with the prior shape knowledge introduced using deformable registration. The system adapts the weight of each piece of information through a machine learning step based on GAs.

5.1 Introduction

Image segmentation is commonly defined as the partitioning of an image into non-overlapping regions that are homogeneous with respect to some visual feature, such as color or texture [132]. In many medical imaging applications, segmentation algorithms play a crucial role by automatically identifying anatomical structures and other regions of interest. Such algorithms are nowadays in the core of multiple tasks, like quantification and measurement of tissue volumes, localization of pathologies or computer-integrated surgery. It is important to highlight that manual segmentation is not only tedious and time consuming but, sometimes, also inaccurate, hence the importance of developing automatic and accurate segmentation methods.

In particular, medical image segmentation is usually challenging due to poor image contrast, noise, diffuse organ/tissue boundaries, and artifacts. These problems can cause considerable difficulties when applying traditional segmentation techniques, such as edge detection or thresholding. Consequently, an intelligent way of proceeding is to incorporate as much prior knowledge as possible about the particular object and image modality to segment. To address these difficulties, deformable models have been extensively studied and widely used in medical image segmentation with interesting results [133, 134].

A single source of prior knowledge is usually not enough to satisfactorily tackle medical image segmentation problems. Therefore, the development of hybrid approaches combining different sources of information has been a major focus in the field of image segmentation [135, 136, 137]. In this Chapter, the search/learning abilities of MHs and the capability of geometric deformable models to handle topological changes are combined. Three sources of information (a region term, a shape prior, and an edge term) are used to accurately segment the organs of interest in different medical image modalities: microscopy, X-ray CT, and MRI. In our proposal, MHs have capital importance in two stages. First, during the training process of the new model, the tuning of the parameters is carried out by a GA. Second, in the proper segmentation stage, the shape prior is obtained by deformable registration using our proposal SS^{+}.

Every image modality has its own peculiarities, thus the training phase allows our model to learn the most suitable parameters for a specific modality/anatomical area using few images as paradigmatic examples. In turn, the segmentation phase uses only one manually segmented reference image to generate the prior shape knowledge that will guide, together with the region- and edge-based terms, the evolution of the moving contour.

To assess the quality of the new approach, we developed an experimental comparison including seven state-of-the-art segmentation methods. The study was carried out on four different datasets, for a total of 22 microscopy, 11 MR, and 5 CT images.

This Chapter is structured as follows: in Section 5.2 we provide the theoretical foundations necessary to understand our work. In Section 5.3, a general overview of the method is presented, providing details about the different terms used in our deformable model. Finally, Section 5.4 presents the results and the statistical analysis of the experimental comparison, followed, in Section 5.5, by some final remarks and a discussion about possible future developments.

5.2 Theoretical Background

In this section, an overview of the main techniques applied in our approach (geometric deformable models, IR and MHs) and previous related work are presented.

5.2.1 Deformable Models

The term ‘‘deformable models’’ (DMs) was first used in the late eighties [138] with reference to curves or surfaces, defined within the image domain, that are deformed under the influence of ‘‘internal’’ forces, related with the curve features, and ‘‘external’’ forces, related with the features of the image regions surrounding the curve. Internal forces enforce regularity constraints and keep the model smooth during deformation, while external forces are defined to attract the model toward features of the object of interest.

DMs are segmentation techniques that use prior information about the shape of the object to be located or segmented. They start with some initial boundary shape

represented in the form of a curve, and iteratively modify it by applying various shrink/-expansion operations according to some energy function. DMs can either be region-based or edge-based approaches, depending on the feature they rely on to segment the object of interest. Region-based methods usually proceed by partitioning the image into connected regions by grouping neighboring pixels with similar features. Edge-based methods, instead, are focused on contour detection, relying on discontinuities in image values between distinct regions.

There are basically two types of DM depending on the kind of shape representation used: parametric/explicit and geometric/implicit.

- Parametric Deformable Models. This type of DM represents curves and surfaces explicitly in their parametric forms during deformation, allowing direct interaction with the model and leading to a compact representation for fast real-time implementation. As examples of parametric DMs we could cite ''snakes'' or Active Contour Models (ACMs) [139], Active Shape Models (ASMs) [140], Active Appearance Models [141, 142], and Topological Active Nets (TANs) [143].

- Geometric Deformable Models. Geometric DMs are based on curve evolution theory [144, 145, 146, 147, 148]. Curves and surfaces are adapted using only geometric measures, resulting in deformations that are independent of the parameterization but, as in parametric DMs, also rely on image data to delineate object boundaries. Since the adaptation is independent of parameterization, the evolving curves and surfaces can be represented implicitly as a level set of a higher-dimensional function and topological changes can be handled automatically.

Among geometric models, the Level Set (LS) method [147, 148] relies on an evolving closed surface defined by a moving interface, the front, which expands into the image. The interface $\Gamma(t)$ can be characterized as a Lipschitz continuous function:

$$\begin{cases} \phi(t,\mathbf{x}) > 0 & \text{for } \mathbf{x} \text{ inside } \Gamma(t) \\ \phi(t,\mathbf{x}) < 0 & \text{for } \mathbf{x} \text{ outside } \Gamma(t) \\ \phi(t,\mathbf{x}) = 0 & \text{for } \mathbf{x} \text{ on } \Gamma(t) \end{cases}$$

The front or evolving boundary, denoted by Γ, is represented by the zero level $\Gamma(t) = \{\mathbf{x}|\phi(t,\mathbf{x}) = 0\}$ of a LS function $\phi(t,\mathbf{x})$. The dynamics of ϕ can be described by the following general form:

$$\frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0$$

known as the LS equation, where F is called the speed function and ∇ is the spatial gradient operator. F can depend on position, time, the geometry of the interface (e.g., its normal or its mean curvature), or the different image features.

In any case, the definition of the LS function ϕ is essential. One common choice is the signed distance function $d(\mathbf{x})$, which gives the distance of a point to the surface

and the sign: generally $d > 0$ if the point $\mathbf{x}$ is outside and $d < 0$ if it is inside the surface (assuming it is a closed surface). This definition is especially interesting to avoid numerical instabilities and inaccuracies during computations. But even with this definition, ϕ will not remain a signed distance function all the time and a reinitialization procedure to keep the LS intact will be needed [149].

5.2.2 Image Segmentation using Deformable Models and Metaheuristics

It is fundamental to understand that the combination of internal and external forces in a DM determines a target function to minimize, whose minimum is theoretically located at the boundary of the object to segment. This target function can be very complex (noisy, highly-multimodal) and the classic algorithms fail at minimizing it [150]. Hence, the global search capabilities of MHs can be very beneficial to optimize this function. Furthermore, the automatic learning of DM parameters is also possible with these intelligent techniques [151]. In fact, this automatic parameter configuration is even desirable since it is known that *manual* parameter tuning is time consuming and may introduce a bias in comparing an algorithm with a reference, due to better knowledge of the algorithm under consideration with respect to the reference, or to possible different time spent tuning each of them.

In the literature several examples can be found which hybridize parametric DMs and MHs. In [152] and [153], snakes are combined with an optimization procedure based on GAs. In [154] a GA evolves a population of medial-based shapes extracted from a training set, using prior shape knowledge to produce feasible deformations while also controlling the scale and localization of these deformations. In [155] an ACM is applied to the automatic segmentation of PET images of liver; a GA is used to find optimal parameters values for the edge detection step. Also, in [150], different MHs are compared (PSO, SS, GA, SA, DE) to optimize an ASM and localize the hippocampus in microscopy images. With respect to standard and extended TANs, in [156], [157] and [158], the minimization of TANs energy to segment CT images is carried out by means of GA and memetic algorithms, Differential Evolution and SS, respectively.

In relation to geometric DMs, much fewer proposals of hybridization have been presented. In [159] a GA is used to perform LS curve evolution using texture and shape information to automatically segment the prostate in CT and MRI pelvic images. In [151], a GA is used to find an optimal set of parameters that characterize the LS method in CT and MRI segmentation. In [160], the initial segmentation based on the LS method is refined using swarms of intelligent agents. Finally, in [161], a comparative study on the segmentation of histological images is carried out where different geometric approaches are initialized using MHs and parametric DMs.

5.3 The HybridLS method

In this section, we present a novel segmentation approach based on the LS method, called HybridLS, that combines edge, region and prior shape knowledge of the target

object to guide the LS evolution. Moreover, we take advantage of the beneficial characteristics of MHs to automatically learn the inherent parameters of a specific type of object using training data (a set of already segmented images).

In its first stage, using an atlas of the target object, HybridLS performs atlas-based segmentation of the image under consideration. This requires the availability of a single image of a similar target object, along with its segmentation. From such data, the initial registration-based step provides a prior segmentation that will allow the LS to start its evolution near the area to be segmented. This benefits both the speed and the accuracy of the segmentation. In fact, with a default initialization over the whole image, features located far from the target area are more likely to negatively influence the evolution of the LS.

The LS moves under the influence of three force terms, each providing information about a different characteristic of the current contour. There are a region, an edge and a prior term. The region term minimizes the inhomogeneity of the intensity values inside and outside the surface enclosed by the evolving contour, while the edge term attracts the curve towards natural boundaries and other edges of the image. Finally, the prior term moves the LS towards the prior segmentation obtained by the registration, incorporating the information gathered at the first stage of the method in the later segmentation process. Note that this is rather different than just using the prior as initial contour for the LS. Indeed, the prior term, rather than its initial location, influences the evolution of the contour, and can balance the other forces when they are small or inconsistent, leading to a more ''conservative'' segmentation with respect to using the initial contour.

Figure 5.1 provides an overview of HybridLS. The total force acting on the LS is a linear combination of the force terms

$$F_{\text{tot}} = w_{\text{r}}F_{\text{r}}(C) + w_{\text{e}}F_{\text{e}}(C) + w_{\text{p}}F_{\text{p}}(C, P) \tag{5.1}$$

where C is the current contour and P is the prior segmentation. Along with the specific parameters for each term, the use of weights provides flexibility to our approach, allowing it to be adapted to the features and particularities of the objects to be segmented. In HybridLS, a GA is in charge of tuning the weights and the parameters of each term based on training data.

In what follows, we describe the components of HybridLS, starting with the computation of the registration-based prior. Then, we define the three force terms and show how to compute them. Finally, we provide details about the GA and the parameter learning phase.

5.3.1 Registration-based prior

The registration algorithm we used is SS^+ (Chapter 3). In this work, the applications of SS^+ are extended to histological and CT images. The registration is performed in two steps, beginning with affine registration. Affine transform can remove large misalignments between the images. Then, a deformable B-Spline-based registration takes care of adjusting the overlap locally, to match the finer details.

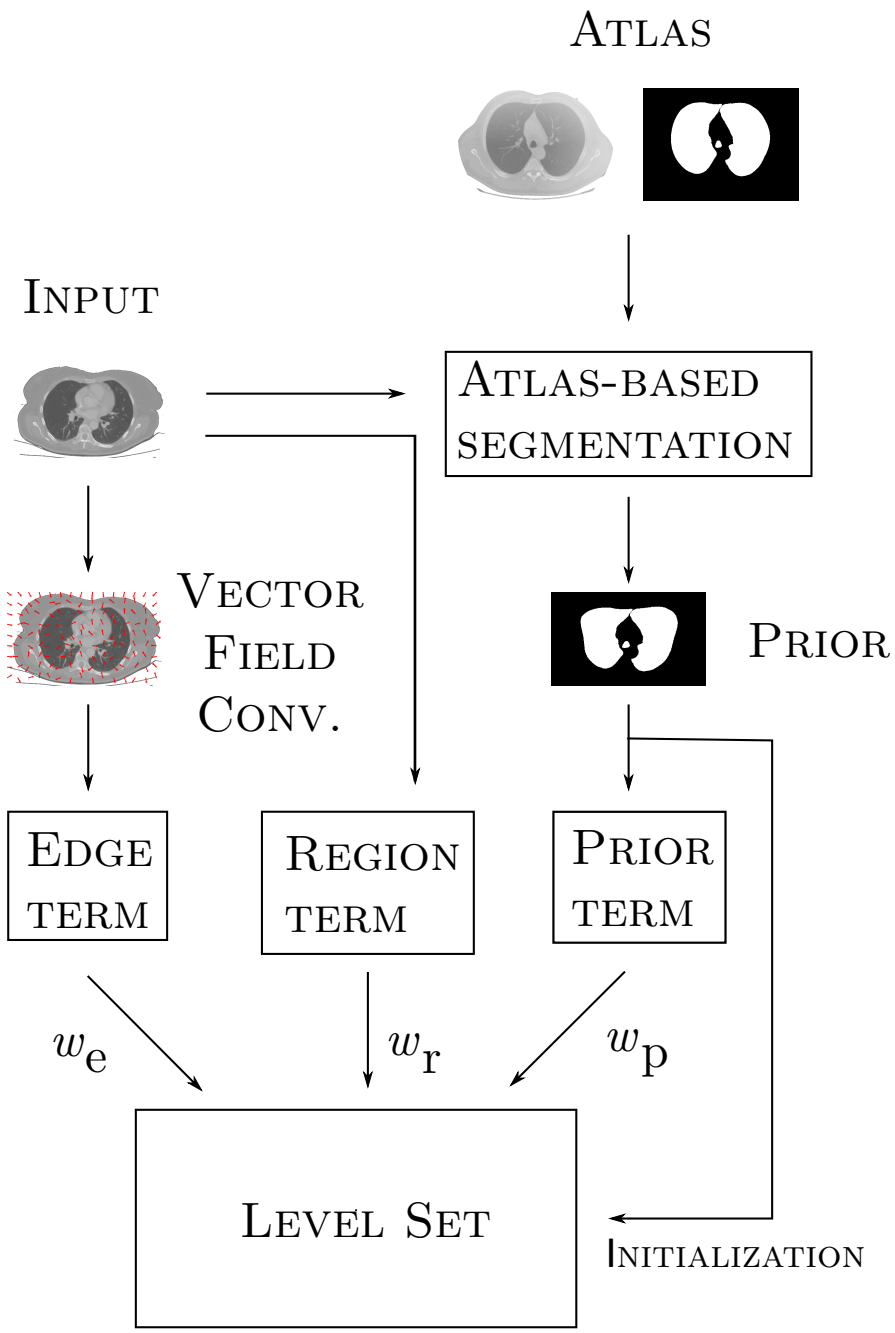


Figure 5.1: The interaction among the components of HybridLS.

To compute a prior, one of the training images plays the role of the atlas. In cases where the target object has a large anatomical variability, a single atlas cannot express the whole distribution of possible shapes the object can assume, leading to poor registration results. This can be improved by using multiple atlases and selecting the most similar atlas for the registration to the target.

In HybridLS, the prior is obtained considering multiple images to be used as atlas. To select the actual atlas, all candidate images are registered to the target image using affine registration. Then, the candidate atlas having the highest similarity metric value is selected for the further B-Spline registration step.

5.3.2 Force terms

5.3.2.1 Region term

Our region term is borrowed from the classic ''Active Contours Without Edges'' (CV) [162] method by Chan and Vese. This algorithm was designed to detect objects whose boundaries are not necessarily defined by gray level gradients; indeed, it ignores edges completely, converting CV in a region-based method. The idea is to separate the image into two regions having homogeneous intensity values. More formally, the process minimizes the energy functional shown in Equation 5.2. The functional is used to evolve a LS representing the contour C, using the conventional variational calculus approach.

$$\begin{aligned} E(C) = {} & \mu \cdot \text{Length}(C) + \nu \cdot \text{Area}(C) \\ & + \lambda_1 \int_C |I(x,y) - \bar{I}_C|^2 \, dx \, dy + \\ & + \lambda_2 \int_{\Omega \setminus C} |I(x,y) - \bar{I}_{\Omega \setminus C}|^2 \, dx \, dy \end{aligned} \tag{5.2}$$

In the equation, I is the intensity value of the image to be segmented and $\bar{I}$ is its average value. Along with the length of C and its area, there are a third and fourth term representing the variance of the intensity level (i.e., the homogeneity) inside and outside C. Each term has a weight that determines its influence on the total energy, so that, for instance, the smaller μ, the more the length of the curve can increase without penalizing the minimization.

In HybridLS, we are interested in a pure region-based term without area or length restrictions, therefore we just use the two homogeneity terms. Therefore, in terms of force acting on the LS, we get

$$F_r(x,y,C) = \begin{cases} \lambda_1 \, |I(x,y) - \bar{I}_C|^2 & (x,y) \in C \\ \lambda_2 \, |I(x,y) - \bar{I}_{\Omega \setminus C}|^2 & (x,y) \notin C \end{cases} \tag{5.3}$$

5.3.2.2 Edge term

The edge term incorporates the information about the boundaries in the image. Basically, the edge term pulls each point of the contour towards the closest edge. Our edge term is based on Vector Field Convolution (VFC) [163]. Compared to other edge-based forces such as Gradient Vector Flow [164], VFC has a lower computational cost and shows better robustness to noise and initialization. In addition, it showed good performance as external force for DMs [158, 163].

The VFC is static, in the sense that it does not depend on the current LS but only on the target image, therefore the field is calculated only once. The computation of the force occurs in two independent steps. First, an edge map of the target image is obtained applying Gaussian smoothing followed by the Sobel edge detector [3]. Then, the edge map is convolved with a vector field kernel K in which each vector point to the origin. The magnitude of the vectors decreases with the distance d, in such a way that distant edges produce a lower force than close edges (the actual value is $1/d^{\gamma+1}$). For a point c of contour C, the edge term is simply the normal component of the VFC with respect to C.

5.3.2.3 Prior term

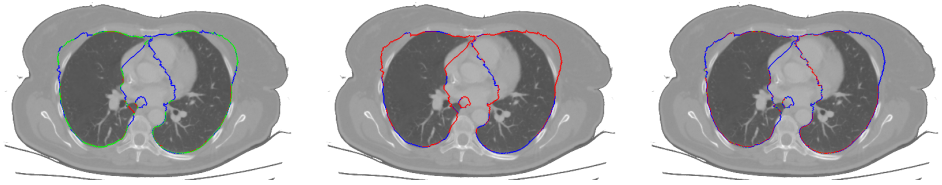
The aim of the prior term is to move the LS towards the prior segmentation. Also, we want the module of every force vector to be proportional to the overlap between the current evolving curve and the prior segmentation. The idea about how to compute the actual force comes from the region term. If one considers the prior segmentation as a binary image, having an intensity value inside the object and another one outside, this image has two regions that are perfectly homogeneous. This is exactly the kind of result our region-based term was designed to deliver. Therefore, to compute the prior term we simply calculate the region term on the *prior* image, rather than on the target image.

$$F_p(x, y, C) = \begin{cases} |P(x, y) - \overline{P}_C|^2 & (x, y) \in C \\ |P(x, y) - \overline{P}_{\Omega \setminus C}|^2 & (x, y) \notin C \end{cases} \tag{5.4}$$

Notice that the weights λ_1, λ_2 serve no purpose in this case, so they have been removed.

Figure 5.2 shows a visual example of the effect of three force terms in segmenting a lung CT. The top row shows the input image (a), the prior segmentation (b) and the vector field convolution (c). In the bottom row, three copies of the input image are overlapped with the current contour, which is colored according to the corresponding force. Green means the force is close to zero, while blue and red colors mark inward and outward forces, respectively. In this example, the region term (d) is correctly moving the contour towards the lungs boundaries, as they define two very homogeneous areas. The prior term (e) is just pulling towards the prior segmentation (b). Finally, the edge term (f) is moving the LS towards the closest edges, regardless whether these belongs to the lungs boundaries or not.

(a) input image (b) atlas-based prior (c) vector field convolution



(d) region term (e) prior term (f) edge term

Figure 5.2: A visualization of the different force terms and the elements from which they are computed.

In the implementation, we used the Shi and Karl's Fast-Two-Cycle (FTC) algorithm [165], which avoids the need of solving partial differential equations (PDE). It is a narrow band technique that restricts the calculations of the LS to a much smaller region than the whole grid, and significantly speed up the curve evolution process. This method also separates the evolution process into two different cycles: one cycle for the data dependent term and a second cycle for the smoothness regularization.

5.3.3 Parameter learning using metaheuristics

HybridLS has the ability to learn optimal parameter settings for every specific dataset. Provided a training set of already segmented images of the same class, the parameters are learned using a classic machine learning approach: configurations of parameters are tested on the training data, and the results are compared with the ground truth to assess their quality. In the most basic approach, all combinations of parameters need to be tested, but this exhaustive search is very time consuming, if not even impossible when a large number of parameters are involved. Fortunately, we can overcome this problem by using MHs, since a properly designed MH has the ability of learning optimal parameter values faster than an exhaustive search.

In this work, we developed a GA to learn the weights of the force terms (w_r, w_e, w_p) and their corresponding parameters (λ_1, λ_2 for the region term and γ for the edge term). A solution of the problem, or an individual in GA terms, is a string of real values encoding the parameters values. The quality of a solution s is defined as the average quality of the segmentations obtained using the parameters values in s. In this case, we measured the average Dice coefficient obtained segmenting the images in the training set. The GA uses a population with fixed size and random uniform initialization. The other components are elitism, tournament selection, blend crossover (BLX-α) and

random mutation.

It is important to notice that, when testing combinations of parameter values, not all segmentation steps need all parameters. For instance, the VFC of an image depends only on γ. Having this in mind, and in order to speed up the learning process, we saved in a cache all the information that are shared between different configurations. This is especially important for the prior, which is the most computationally demanding step in the segmentation process by far. The prior do not use any of the parameters in the learning process, therefore only one per image is needed but, since the registration algorithm is non-deterministic, we represented its variability by creating a pool of 30 priors for each image. The priors used in tuning the segmentation parameters were drawn at random from the pool. This approach led to an impressive speedup of the training process. Once the priors and the VFC of each training image has been computed, a single parameter configuration could be tested in less than a second.

5.4 Experimental Setup

One of the main aims of this research is to develop a method that, combining the advantages of geometric DMs, MHs and prior shape knowledge, can achieve remarkable results with different medical image modalities and anatomical structures of interest. To accomplish this purpose, three image modalities with completely different characteristics and various structures have been tested. In this section, these datasets will be described, as well as the anatomical structures to be segmented. Then, we will present the different methods included in the comparison, and devote two separate sections to the atlas registration and the tuning of the parameters (given their critical importance in our pipeline). Finally, the overall results of segmentation will be presented and analyzed.

5.4.1 Datasets

Three kinds of biomedical image modalities were used to verify the global performance of the different methods over different datasets. We focused our interest on microscopy histological images derived using In Situ Hybridization, X-Ray CT, and MRI.

- In Situ Hybridization-derived images (ISH). 26 microscopy histological images were downloaded from the Allen Brain Atlas (ABA) [166]. The anatomical structure to segment was the hippocampus, and the ground truth was created manually by an expert in molecular biology: every image was manually segmented 5 times and, for each group of 5 manual segmentations, the consensus image was calculated and used as ground truth. The typical resolution of ABA images is about 15,000 × 7,000 pixels, and the ROIs about 2,500 × 2,000 pixels.

- Magnetic Resonance Imaging (MRI). A set of 17 T_1-weighted brain MRI were retrieved from a NMR database with their associated manual segmentations [112]. The deep brain structures to segment were caudate, putamen, globus pallidus,

and thalamus. All MR images used in training and test have a resolution of 256 × 256 pixels.

- X-Ray Computed Tomography (CT). A set of 10 CT images were used in the experiments [158]. Four of them correspond to a human knee and the other six to human lungs. The gray value of all pixels have been inverted so the bone and the lungs are the darker objects in the image. Knee images have an average size of 410 × 435 pixels, while Lung images have 510 × 350 pixels.

All four datasets, considering lungs and knee as different image sets, were divided in training and test data. The training images were used by HybridLS to learn the parameters, while the test images were the ones used in the final experiments to check the performance of the methods.

In ISH, 22 images were used for testing and 4 as a training set. As atlas for the registration, the actual references in the ABA were employed to obtain the shape prior. With respect to MRI, 3 images were used as training set, another 3 were used as atlas, and the remaining 11 as test set. Finally, in relation to CT, one image of every organ was used as training and atlas for the registration, leaving 3 lung and 2 knee images for testing the system.

5.4.2 Methods included in the comparison

In our comparisons we have included both deterministic and non-deterministic methods, as well as classic and very recent proposals. The stress has been focused on DMs, and their hybridization with MHs, but other kinds of approaches have also been taken into account.

- Active Shape Models (and Iterative Otsu Thresholding Method) refined using Random Forests (ASM + RF) [150, 167]. This method, published in 2012, uses a medial-based shape representation in polar coordinates, with the objective of creating simple models that can be managed in an easy and fast manner. Such a parametric model is moved and deformed by a DE algorithm according to an intensity-based similarity function between the model and the object itself. After that, Otsu's thresholding method [168] is iteratively applied on every region identified by the located control points. Finally, Random Forests [169] is applied to expand the segmented area to the regions that were not properly localized. This segmentation algorithm has shown very good performance in histological images, but needs a training set of shapes to manually create the parametric template and its possible deformations, as well as a training set of textural patterns for the expansion phase. Due to these restrictions it was only applied to ISH images.

- Soft Thresholding (ST) [170]. This deterministic method, presented in 2010, is based on relating each pixel in the image to the different regions via a membership function, rather than through hard decisions, and such a membership function

is derived from the image histogram. In a first stage, the normalized histogram of the image is calculated and a sum of weighted known distributions is fit to it. Each probability distribution represents the probability for a pixel with a certain value to belong to the corresponding region. This segmentation technique is totally automatic, and the spatial operations performed make the thresholding more robust to noise and artifacts. Having been successfully applied to CT, MRI and ultrasound, it seemed interesting to apply it also to microscopy histological images and compare its performance with other state-of-the-art methods.

- Atlas-based deformable segmentation (DS). This method refers to the atlas-based segmentation procedure used in HybridLS to compute the prior. This is actually a stand-alone segmentation method, therefore it is included in the experimental study as a representative of registration-based segmentation algorithms. Moreover, comparing DS's and HybridLS's results will assess the influence of the prior term on the performance of the second method. During the whole study, the setup and the atlas selection mechanism of DS (Section 5.3.1) are always the same whether the method is used stand-alone or embedded in another segmentation technique.

- Geodesic Active Contours (GAC) [171]. This technique, introduced in 1997, connects classical 'snakes' based on energy minimization and geometric active contours based on the theory of curve evolution. It is based on active contours that evolve in time according to intrinsic geometric measures of the image: the evolving contours naturally split and merge, allowing the simultaneous detection of several objects and both interior and exterior boundaries.

  The Partial Differential Equation of the GAC is the following:

$$u_t = \alpha \cdot div(g\nabla u/|\nabla u|)|\nabla u| + \beta \cdot g|\nabla u| \tag{5.5}$$

  where g is a positive and strictly decreasing function, ∇ is the gradient operator computed on image I, div is the divergence operator (that measures the magnitude of a vector field's source or sink at a given point), and α and β are the contour (internal force) and expansion (external force) weights, respectively. This method is equivalent to the minimization of the length of curve C according to a Riemannian metric, and such a metric depends here on the local gradient of the image I.

  In this study, two implementations of GAC have been tested. The first one uses as initial contour the whole image, while the second one, called DSGAC, employs the segmentation obtained using DS to create the initial contour of the geometric DM.

- Chan&Vese Level Set Model (CV) [162]. This implicit DM was also included in the comparison to check its performance in comparison with the other approaches (see section 5.3.2.1). Also in this case, like in GAC, two implementations

have been tested. The first one uses the whole image as initial contour, and the second one employs the segmentation result obtained by DS as the LS initial contour.

5.4.3 Parameter settings

As HybridLS has an automatic parameter learning phase, it would be unfair to compare it against other methods without some kind of parameter tuning. A manual tuning is time consuming and error-prone, while using the GA to tune all methods could introduce a bias, as the behavior of the GA could vary with each method. In general, we want the competitors to deliver their best performance, regardless of their parameter sensitivity or their ability to be tuned. Therefore, we decided to tune the competitors with an exhaustive search using the *test data*, rather than the training one. This means the results reported for all methods but HybridLS are actually the best average results they can obtain on these datasets. This gives them a clear advantage over HybridLS, as for the latter the parameters are learned using the training data only.

For CV, GAC, DSGAC and DSCV, all the possible combinations of the values in Table 5.1 were tested. Also, a pre-processing and a post-processing stages were included to improve the results obtained. The post-processing stage makes a refinement of the results removing the connected components smaller than a certain amount of pixels, while the pre-processing is a median filter to remove salt-and-pepper-like noise present in some of the images. Moreover, for DSGAC and DSCV, 10 different initial masks were created using DS and the best one was used in the tuning. The number of iterations for GAC and CV was set to 500 to ensure the process reached convergence. In a few cases, on the ISH dataset, CV failed to converge within the limit due to poor parameters values. This occurred only while producing very low quality, degenerate segmentations, therefore the early stopping did not affect the tuning process.

After tuning these methods, the minimum allowed size in pixels of connected components was set to 75, 200 and 25000 for MRI, CT and ISH, respectively. For ASM+RF, the parameters used (Table 5.3) were those suggested in the bibliography for the ISH dataset, which has been deeply tested by the authors.

For HybridLS, the parameters settings were learned by the GA using the training data. The size of the population was set to 50 individuals, and the evolution lasted 50 generations. The probability of crossover and mutation was set to 0.7 and 0.1, respectively, and the size of the tournament was 3. The range of λ_1, λ_2 was restricted to $\{1, 2, 5\}$ to match the settings used with the other methods.

The final parameters configurations are reported in Table 5.2. It is interesting to remark how the GA detected a different level of importance for each term across the datasets. For instance, in MRI the edge term is not used ($w_e = 0$) since our machine learning system determines that, for a good segmentation, the region term and prior shape knowledge are enough. When segmenting CT lungs the only term used is the region-based one. In this case, λ_1 and λ_2 were set to 5 and 2, respectively. This means that our final segmentation will have a more uniform foreground region (since the

Table 5.1: Combination of parameters tested for CV, GAC, DSCV and DSGAC.

Parameter	Values
α contour_weight	{1, 2, 3}
β expansion_weight	{-1, -0.5}
μ weightLengthTerm	{0.01, 0.1, 0.25, 0.5, 0.75}
λ_1	{ 1, 2, 5}
λ_2	{ 1, 2, 5}
size median filter	{1, 3, 5, 10}
minimum size allowed	{1, 50, 75, 100, 200, 5000, 25000}

energy contributed by the ''variance'' in the foreground region has a larger weight), at the expense of allowing more variation in the background.

5.4.4 Experimental results

To evaluate the performance of the segmentation methods, we employed three standard segmentation metrics: the Dice coefficient (DSC), the Jaccard similarity index (JI) and the Hausdorff distance (HD). Both the Dice coefficient and the Jaccard index measure set agreement: a value of 0 indicates no overlap with the ground truth, and a value of 1 indicates perfect agreement. In turn, the Hausdorff distance measures the mutual proximity of two images, calculated as the maximum distance from a point in the ground truth to the closest point in the segmentation.

It is important to remark that ASM+RF, DS, DSCV, DSGAC and HybridLS are non-deterministic, since stochastic methods, like Differential Evolution or Scatter Search, are embedded in these algorithms. It is essential to execute such algorithms several times to estimate and compare their performances. In this case, 20 repetitions per image were run and the mean, median and standard deviation values were calculated over the whole set of results (see Table 5.5). For instance, in ISH the mean Dice value of DS is 0.876, and represents the average of 440 experiments performed (20 repetitions per image and 22 images).

We also performed a statistical analysis of the results. When comparing two methods, we used Wilcoxon rank-sum test [172], a non-parametric statistical test that checks whether one of two independent samples tends to have larger values than the other. When multiple comparisons were performed, Holm's correction [104] was applied to the p-values to control the family-wise error rate. Note that, in the Lungs and Knee datasets, the number of images is not large enough to allow the comparison of the deterministic methods (ST, CV and GAC), therefore they have been excluded from the test.

The experiments were run on a computer with an Intel Core i5-2410M processor and 4.00 GB of RAM. In Table 5.4, some concise information about the running time of each algorithm is provided with an illustrative purpose. The fastest method is ST with MRI, taking only 1 second per image, while the slowest are the different applications

Table 5.2: Parameters obtained after tuning ST, GAC, CV, DS+GAC, DS+CV, and training HybridLS.

CV	GAC	CV+DS	GAC+DS	HybridLS
		Magnetic Resonance Imaging		
500 iterations	500 iterations	500 iterations	500 iterations	$\lambda_1 = 5$
$\nu = 0$	β = -1	$\nu = 0$	β = -0.5	$\lambda_2 = 1$
$\mu = 0.01$	$\alpha = 3$	$\mu = 0.01$	$\alpha = 3$	$w_r = 5.1$
$\lambda_1 = \lambda_2 = 1$	medFiltSize = 3	$\lambda_1 = 1$	medFiltSize = 1	$w_p = 1.1$
medFiltSize = 1		λ_2 =1		$w_e = 0$
		medFiltSize = 5		$\gamma = 1.5$
		Computerized Tomography - Knee		
500 iterations	500 iterations	500 iterations	500 iterations	$\lambda_1 = 2$
$\nu = 0$	$\alpha = 1$	$\nu = 0$	$\alpha = 3$	$\lambda_2 = 5$
$\mu = 0.01$	β = -0.5	$\mu = 0.01$	β = -0.5	$w_r = 4.8$
$\lambda_1 = 5$	medFiltSize = 1	$\lambda_1 = 1$	medFiltSize = 1	$w_p = 0.9$
$\lambda_2 = 2$		$\lambda_2 = 1$		$w_e = 2$
medFiltSize = 3		medFiltSize = 1		$\gamma = 1.5$
		Computerized Tomography - Lungs		
500 iterations	500 iterations	500 iterations	500 iterations	$\lambda_1 = 5$
$\nu = 0$	β = -1	$\nu = 0$	β = -1	$\lambda_2 = 2$
$\mu = 0.01$	$\alpha = 2$	$\mu = 0.01$	$\alpha = 3$	$w_r = 1.5$
$\lambda_1 = 5$	medFiltSize = 3	λ_1 =1	medFiltSize = 3	$w_p = 0$
$\lambda_2 = 2$		$\lambda_2 = 5$		$w_e = 0$
medFiltSize = 3		medFiltSize = 3		$\gamma = 1.5$
		In Situ Hybridization-derived images		
500 iterations	500 iterations	500 iterations	500 iterations	$\lambda_1 = 1$
$\nu = 0$	β = -1	$\nu = 0$	β = -1	$\lambda_2 = 1$
$\mu = 0.01$	$\alpha = 3$	$\mu = 0.01$	$\alpha = 3$	$w_r = 1.9$
$\lambda_1 = \lambda_2 = 1$	medFiltSize = 10	$\lambda_1 = 1$	medFiltSize = 10	$w_p = 2.2$
medFiltSize = 5		$\lambda_2 = 1$		$w_e = 1$
		medFiltSize = 5		$\gamma = 2$

Table 5.3: Parameters used in ST, DS and ASM+RF. All parameters were taken from the original proposals. Note that the parameters of ST are the same for all image modalities.

ST	ASM+RF	DS
L = 2 regions	Cr = 0.9	Metric = AdvancedNormalizedCorrelation
Relative max	F = 0.7	Optimizer = ScatterSearch
normalization	Uniform Crossover	SS_b = 12
	DE/target-to-best/1	PSize = 32
	Population Size = 80	BLX-α = 0.3
	Iterations = 250	LS-iterations = 25
	Median Filter [25×25]	NumberOfIterations = 15
	RF with 500 trees	NumberOfResolutions = 3
		NumberOfSpatialSamples = 2000 5000 10000
		Restarts = 3

Table 5.4: Average execution time (in seconds) per method and kind of image. The programming environment is also reported.

	Prog. Env.	ISH	Lungs	Knee	MRI
ASM+RF	MATLAB, C++	35	-	-	-
ST	MATLAB	39	1.7	2.5	1
CV	MATLAB	87	36	67	11
GAC	MATLAB, C++	32	15	16	1.5
DSCV	MATLAB, C++	582	384	310	429
DSGAC	MATLAB, C++	493	342	265	407
DS	C++	471	326	245	404
HybridLS	C++	545	331	252	405

of DS to ISH, employing up to 10 minutes to process an image. Nevertheless, several factors affect the accuracy of a comparison in terms of execution time. First, some of the methods have been developed in MATLAB and others in C++. Moreover, the size of the images differs from one image modality to another, as well as some of the pre- and post-processing stages we used. Finally, the nature of the algorithms is completely different, making them hardly comparable. For instance, ST is based on the image histogram only, while others have a registration step that imposes a notable computational overhead.

5.4.4.1 Analysis

The experimental results are reported in Table 5.5. Visual examples of two segmentations obtained by the methods on each dataset are provided in Figure 5.4. For simplicity, our discussion focuses on the results in terms of mean DSC, however this choice does not affect the outcome of the comparison, as there is an almost perfect agreement with the

other validation measures we considered.

The performance of the segmentation methods varies greatly across the four datasets. The easiest problem to be solved has been the segmentation of Lungs in CT images, with all methods but GAC and DS scoring higher than 0.95. The most complex task has shown to be the segmentation of deep anatomical structures in brain MRI, where four of the compared methods have obtained an average DSC of 0.2 or less).

The per-dataset results are shown in Figure 5.3 using boxplots and in Table 5.6 through the average rankings. Obviously, the performance of every method depends on the nature of the image to be segmented. For instance, techniques based on grey intensity level (such as CV and ST) yielded worse results in image modalities with less contrast and small differences in terms of pixel intensity like MRI.

HybridLS has obtained the best results in all biomedical image datasets. It achieved the best values for the mean DSC and JI metrics, and it was ranked as the best method in every image modality. The Wilcoxon test (Table 5.6) showed, with really high confidence, that the difference between HybridLS and the other methods is statistically significant in all but one case (DS on MRI). This behavior is also robust, as shown by the low standard deviation values. We can then conclude that our proposal is the best segmentation method in the comparison.

The DS method has been one of the best-performing algorithms, ranking second or third over three datasets. More in general, all methods using the registration-based initialization scored better than those using a standard one. This applies also to CV and GAC: in all but one case, both DSCV and DSGAC ranked better than their counterpart, with a statistically significant difference (Table 5.7).

Overall, DSGAC delivered an acceptable performance, ranking above average in three datasets out of four. This is remarkable, as the regular GAC ranked constantly in the last three positions, and it can be explained by the high sensitivity of GAC to its initialization.

DSCV ranked around average in all datasets, performing slightly worse, although more consistently, than DSGAC. The plain CV method achieved a bad performance, ranking last or second to last in three datasets. Only on the Lungs dataset, where the grey value is enough to segment the target quite accurately, CV delivered good results.

ST results showed a similar pattern to CV. It performed better than CV, but being ST based on the histogram it showed limited ability to cope with complex scenarios. On the other hand, ST is the fastest method on the group and it has virtually no parameters to be set.

ASM+RF obtained some of the best results with ISH images, being also one of the fastest techniques, but it is fair to underline its ad-hoc nature. It needs a training set of shapes to create the template and its possible deformations, and it also needs a training set of textural patterns for the expansion phase. Also, it is not able to manage topological changes in a natural way, as geometric DMs can do.

Finally, from the values of HD, it is interesting to notice how methods without shape deformation restrictions, like ST, CV and GAC, have a higher (worse) HD with respect to others introducing prior shape knowledge, like ASM+RF, DS and HybridLS.

Table 5.5: Segmentation Results using 3 different metrics: Dice Similarity Coefficient (DSC), Jaccard Index (JI), and Hausdorff Distance (HD).

Dataset	Method	Dice Coefficient			Jaccard Index			Hausdorff Distance		
		mean	median	stdev	mean	median	stdev	mean	median	stdev
ISH	ASM+RF	0.885	0.896	0.040	0.797	0.812	0.061	114.9	94.73	52.78
	CV	0.589	0.723	0.257	0.460	0.567	0.242	839.8	844.77	331.8
	DS	0.876	0.907	0.078	0.787	0.829	0.108	101.43	75.16	79.77
	DSCV	0.673	0.764	0.203	0.538	0.618	0.203	263.34	176.7	234.19
	DSGAC	0.791	0.830	0.143	0.674	0.709	0.172	215.23	150.72	196.90
	GAC	0.670	0.722	0.181	0.528	0.564	0.192	693.25	707.56	265.58
	HybridLS	0.888	0.918	0.079	0.806	0.849	0.109	103.61	70.109	100.11
	ST	0.728	0.775	0.175	0.597	0.632	0.192	578.87	665.3	303.31
Knee	CV	0.230	0.230	0.072	0.131	0.131	0.046	179.68	179.68	30.49
	DS	0.687	0.685	0.227	0.569	0.563	0.271	45.01	45.3	25.3
	DSCV	0.528	0.527	0.079	0.363	0.361	0.073	69.64	76.26	16.46
	DSGAC	0.810	0.811	0.142	0.705	0.705	0.204	84.61	84.61	4.04
	GAC	0.486	0.486	0.310	0.349	0.349	0.276	187.08	187.08	29.52
	HybridLS	0.868	0.872	0.087	0.777	0.782	0.136	13.69	13.61	1.73
	ST	0.398	0.398	0.088	0.250	0.250	0.069	144.86	144.86	36.79
Lungs	CV	0.973	0.992	0.034	0.949	0.983	0.063	52.9	60.2	38.59
	DS	0.896	0.897	0.062	0.818	0.814	0.102	39.37	51.47	18.02
	DSCV	0.966	0.985	0.034	0.935	0.971	0.062	26.51	25.49	9.32
	DSGAC	0.950	0.952	0.027	0.906	0.908	0.049	37.82	29.15	27.19
	GAC	0.670	0.627	0.251	0.542	0.457	0.309	143.93	137.56	34.43
	HybridLS	0.996	0.997	0.001	0.992	0.994	0.003	2.94	3.6	1.41
	ST	0.979	0.990	0.023	0.960	0.981	0.044	51.55	60.2	42.88
MRI	CV	0.155	0.171	0.042	0.084	0.093	0.024	93.72	93.4	4.67
	DS	0.752	0.783	0.056	0.606	0.643	0.071	10.83	10.0	3.19
	DSCV	0.204	0.213	0.054	0.115	0.119	0.033	76.79	76.32	5.95
	DSGAC	0.585	0.613	0.087	0.418	0.442	0.084	25.11	25.49	2.41
	GAC	0.124	0.139	0.035	0.066	0.074	0.020	90.85	90.13	5.45
	HybridLS	0.758	0.780	0.048	0.612	0.639	0.062	7.81	6.7	2.87
	ST	0.175	0.181	0.053	0.097	0.100	0.032	84.93	85.58	6.28

Table 5.6: Average rank achieved by every method per image modality and adjusted p-value of Wilcoxon test comparing each algorithm against HybridLS.

Dataset	Method	Mean Rank	p-value
ISH	HybridLS	1.82	
	DS	2.50	0.000
	ASM+RF	2.64	0.000
	DSGAC	4.14	0.000
	ST	5.68	0.000
	DSCV	6.23	0.000
	GAC	6.36	0.000
	CV	6.64	0.000
Knee	HybridLS	1.50	
	DSGAC	2.00	0.000
	DS	3.00	0.000
	DSCV	4.50	0.000
	GAC	5.00	-
	ST	5.50	-
	CV	6.50	-
Lungs	HybridLS	1.00	
	ST	2.33	-
	CV	3.00	-
	DSCV	3.67	0.000
	DSGAC	5.33	0.000
	DS	5.67	0.000
	GAC	7.00	-
MRI	HybridLS	1.27	
	DS	1.73	0.46
	DSGAC	3.00	0.000
	DSCV	4.18	0.000
	ST	5.09	0.000
	CV	5.73	0.000
	GAC	7.00	0.000

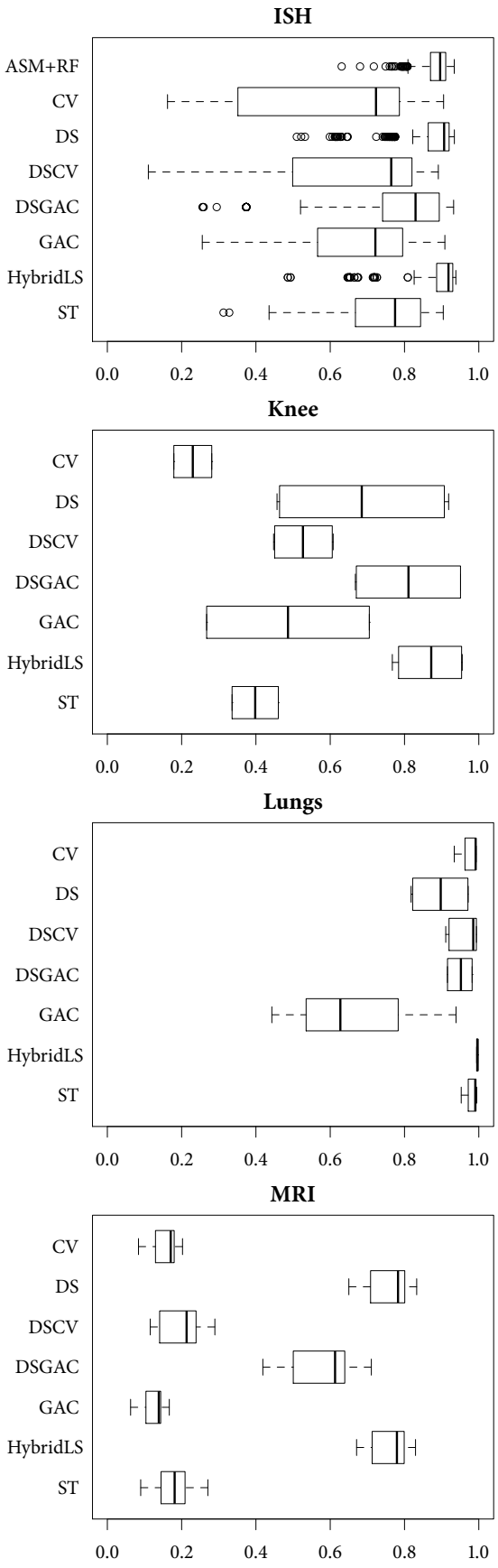


Figure 5.3: Box-plot representing the DSC for all methods.

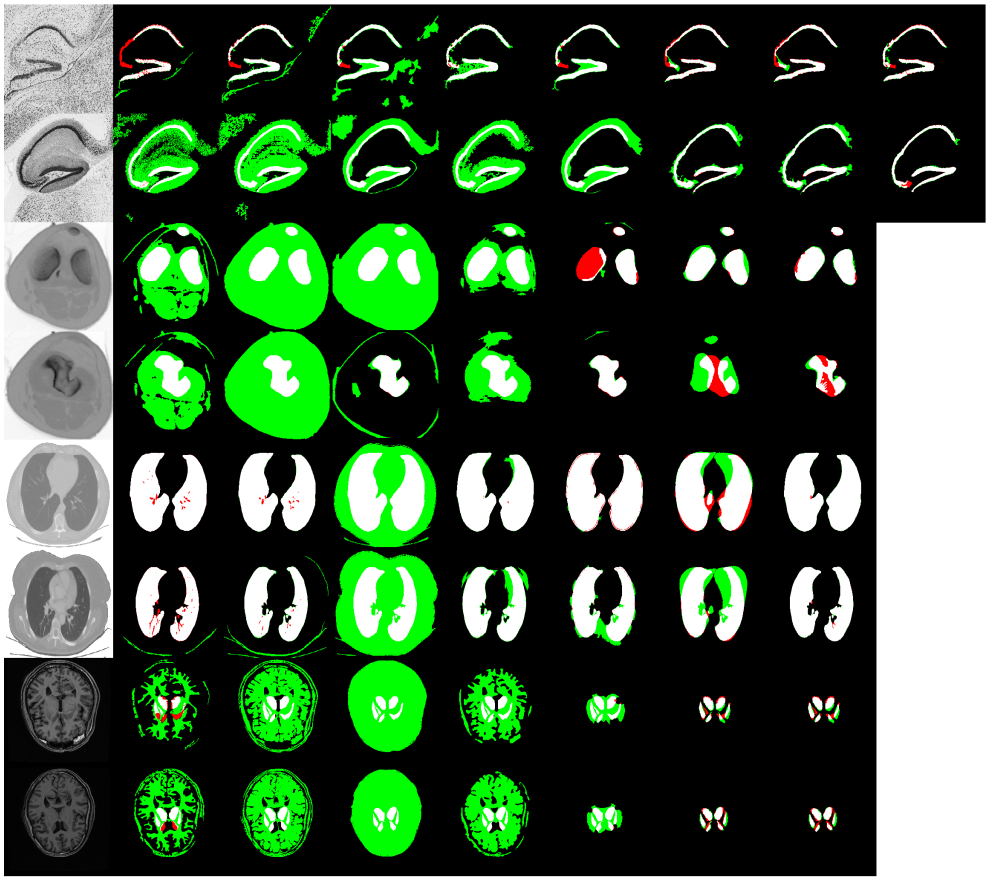


Figure 5.4: Some visual examples of the results obtained. Two images per image modality and structure to segment have been selected: the first two rows correspond to ISH, the next two rows to CT-Knee, and the last four to CT-Lungs and MRI. White represents true positives, red false negatives, and green false positives.

Table 5.7: Results of pairwise Wilcoxon test comparing all the methods but HybridLS.

ISH

	ASM+RF	CV	DS	DSCV	DSGAC	GAC
CV	0.00					
DS	0.01	0.00				
DSCV	0.00	0.38	0.00			
DSGAC	0.00	0.00	0.00	0.00		
GAC	0.00	0.92	0.00	0.92	0.00	
ST	0.00	0.32	0.00	0.61	0.26	0.61

Knee

	DS	DSCV
DSCV	0.00	
DSGAC	0.00	0.00

Lungs

	DS	DSCV
DSCV	0.00	
DSGAC	0.00	0.00

MRI

	CV	DS	DSCV	DSGAC	GAC
DS	0.00				
DSCV	0.02	0.00			
DSGAC	0.00	0.00	0.00		
GAC	0.10	0.00	0.00	0.00	
ST	0.79	0.00	0.26	0.00	0.08

5.5 Discussion

It is crucial to highlight the main features of HybridLS:

- it is an accurate and also general segmentation method (it obtains very good results with all the medical image modalities tested, even overcoming well-consolidated techniques);
- its overall standard deviation is the lowest among the different methods compared, therefore we can affirm that the developed approach is consistent and stable in terms of performance;
- it does not need a training set of textures or shapes to segment the object of interest (it needs only one reference image to obtain the shape prior);

- it performs self-adaptation of its own parameters depending on the medical image modality to segment;
- and uses MHs to generate the shape prior and to perform the previously mentioned learning of parameters.

Thanks to the automatic learning of the model parameters, our hybrid proposal is able to perform an effective segmentation with very different medical image modalities, adapting the importance of every term to each image modality and anatomical structure.

Chapter 6

Final considerations

6.1 Conclusions

In this dissertation we have proposed different intensity-based medical IR methods based on metaheuristics and evolutionary computation. We have proven that, with an appropriate design, these techniques are competitive with gradient-based approaches traditionally used in medical imaging as well as with other outstanding feature-based algorithms. In particular, our proposal based on Scatter Search was able to significantly outperform all competitors in two experimental studies, despite being compared with a large and different group of state of the art techniques.

In the following items, we analyze the results obtained in the pursuit of the objectives set at the beginning of the work:

- *Study the state of the art in medical image registration using metaheuristics and evolutionary computation.*

  After a deep study of the field, we have identified common patterns in the design of this kind of algorithms and turned the results of the analysis into recommendations for the development of novel techniques. A key element is the ability to switch from a broad, global search at the earliest stage of the optimization, to a more focused search once a coarse approximation of the solution has been found.

  The study has been published in the form of a literature review in

  – A. Valsecchi, S. Damas, J. Santamaría. Evolutionary intensity-based medical image registration: a review. *To appear on* Current Medical Imaging Reviews.

- *Design new IR methods taking advantage of robust optimization algorithms based on EC or other MHs.*

  The expertise acquired through the literature review has been put into practice in the design of two novel techniques. An initial proposal has been based on

genetic algorithms; advanced MH components, namely restart and dynamic boundary, were introduced and integrated with the multiple resolution strategy. This problem-specific approach is the ultimate responsible for the excellent performance of the method. In the second proposal, further improvement has been obtained by using Scatter Search for the optimization. Its flexible, modular design and the reduced population have proven to be easier to customize and more effective in facing registration problems. Besides the optimization component, both our approaches have been designed in a modular fashion, to be able to support different kind of images, transformation models, similarity metrics and other minor components. Finally, they have been integrated in a popular, publicly-available medical IR toolkit, Elastix, with the aim of facilitating the adoption from practitioners.

- *Validate the performance of our proposals through a thorough experimental comparison and analysis.*

  We have designed two experimental studies very different from one another. The first study employed simulated brain MRI; we created a series of scenarios of different complexity by selecting images having noise and anatomical lesions, as well as by applying know transformations to the images. The comparison included a variety of outstanding competitors, including classic IR techniques and feature-based algorithms. The second study focused on a common medical application of IR, atlas-based segmentation. The experiment employed a dataset of real-word brain MRI, retrieved through a collaboration with the Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (ICM) at the Université Paris 6, CNRS, Paris (France). At the end of both studies, we obtained quantitative results that were analyzed using proper non-parametric statistical tests.

  The two proposed algorithms and the experimental comparisons have resulted in a number of papers:

  - A. Valsecchi, S. Damas, J. Santamaría, L. Marrakchi-Kacem. Intensity-based image registration using scatter search. *Submitted to* Artificial Intelligence in Medicine.
  - A. Valsecchi, S. Damas, J. Santamaría, L. Marrakchi-Kacem. Genetic Algorithms for Voxel-based Medical Image Registration. Proceedings of the IEEE Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI), 2013, Singapore, April 2013, pages 22–29.
  - A. Valsecchi, S. Damas, J. Santamaría, L. Marrakchi-Kacem. Voxel-based Medical Image Registration using Genetic Algorithms. Proceedings of the International Conference On Medical Imaging Using Bio-Inspired and Soft Computing (MIBISOC13), Brussels, Belgium, May 2013.

  - A. Valsecchi, S. Damas, J. Santamaría. An Image Registration Approach using Genetic Algorithms. Proceedings of the IEEE Congress on Evolutionary Computation (CEC) 2012, Brisbane, Australia, pages 416–423.
  - A. Valsecchi, S. Damas, J. Santamaría. Una Aproximación al Registrado de Imágenes Médicas con Algoritmos Genéticos. VIII Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados 2012, pages 667–674.

- *Apply approaches for algorithm configuration to image registration.*

  We studied the application of offline parameter tuning techniques to IR algorithms. Through a stochastic process known as iterated racing, algorithms are automatically tuned for a specific application and can deliver the best possible results without the burden of a manual or exhaustive tuning. In addition to demonstrating the applicability of this kind of techniques to IR, we developed a comparison of *tuned* IR algorithms. Through tuning, we removed any bias that could have affected the comparison due to of an improper setting of the algorithms' parameters. Two major conclusions were reached. On one hand, all algorithms in the comparison benefited from the tuning; on the other, our two proposals delivered the best performances in the comparison, proving the effectiveness of MH-based IR also in this context.

  This study was carried out in collaboration with Artificial Intelligence research laboratory of the Université Libre de Bruxelles (Belgium). Our findings have been published in the following paper:

  - A. Valsecchi, J. Dubois-Lacoste, T. Stützle, S. Damas, J. Santamaría, L. Marrakchi-Kacem. Evolutionary Medical Image Registration using Automatic Parameter Tuning. Proceedings of the IEEE Congress on Evolutionary Computation (CEC) 2013, Cancun, June 2013, pages 1326–1333.

- *Integrate our algorithms into larger medical applications based on image registration.*

  We developed a novel segmentation algorithm based on level set. Our SS-based registration algorithm is used to perform atlas-based segmentation, resulting in a prior segmentation of the target object. The prior is used both as initial contour and as a force driving the evolution of the level set, together with other force terms that exploit information about edges and regional boundaries. We carried out a large comparison of segmentation methods on multiple medical datasets, including histological, CT and MRI images, and our algorithm outperformed all seven competitors with a significant margin.

  The study, performed in collaboration with the Intelligent Bio-Inspired System Laboratory at the University of Parma (Italy), resulted in the paper:

  - P. Mesejo, A. Valsecchi, L. Marrakchi-Kacem, S. Cagnoni, S. Damas. Biomedical Image Segmentation using Geometric Deformable Models and Metaheuristics. *To appear on* Computerized Medical Imaging and Graphics.

6.2 Future work

Next, we will discuss some open research lines concerning the issues tackled in this dissertation. Besides, we consider some extensions of our proposals that will be developed as future work.

- *Apply our proposals to other specific medical applications.*

  IR is used in countless applications in medical image analysis, and obviously we have considered just a limited number of them. Among others, we mention functional imaging and ultra high field MRI as examples of areas that can provide novel and challenging IR scenarios.

- *Consider other advanced metaheuristics for the optimizer component.*

  Although the performance of Scatter Search has been excellent, we may improve our results by considering other metaheuristics or approaches to the problem. An ideal candidate is to be found in memetic algorithms. These are not just state of the art techniques for continuous optimization problems, but also fit perfectly the two-stage/two-component pattern we observed during the literature review.

- *Develop GPU implementations to tackle real-time or interoperative IR.*

  Metaheuristics and evolutionary computation techniques are generally computationally expensive and difficult to apply in real-time applications. Our algorithms are not exceptions; however, implementing these algorithms on GPUs can lead to extremely large speedups. This may allow the use of our proposals in the context of interoperative image registration, which is a commonly used in image-guided surgery systems, closing one of the main gaps with gradient-based techniques.

6.3 Recommendations

Finally, having analyzed the literature on intensity-based IR methods using MHs and designed two novel methods ourselves, we make the following recommendations for the design of new, generic IR method.

- Follow the parameter-based strategy and use a metaheuristic with a real-coded design and appropriate operators for a continuous optimization problems. Matching-based algorithms are complex to design even for a 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 dividing 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 increase 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 applications, and it can be easily turned into a similarity or rigid transformation model by setting simple constraints on the transformation parameters. For deformable registration, B-splines is the current method of choice.

Conclusiones y trabajo futuro

En esta tesis doctoral hemos propuesto distintos métodos de registrado de imagen médica mediante el uso de técnicas de computación evolutiva y otras metaheurísticas. Hemos demostrado que, con un diseño apropiado, estas técnicas son competitivas con los enfoques basados en el gradiente utilizados tradicionalmente en imagen médica, así como con otros algoritmos basados en características. En particular, nuestra adaptación de Scatter Search al problema ha sido capaz de superar de manera significativa a todos los competidores en dos estudios experimentales, a pesar de ser comparado con un grupo grande y diverso de técnicas del estado del arte.

En los puntos siguientes se analizan los resultados obtenidos en la búsqueda de los objetivos fijados al comienzo del trabajo:

- *Estudio del estado del arte en registrado de imágenes médicas mediante metaheurísticas y computación evolutiva* .

  Después de un profundo estudio del campo, hemos identificado patrones comunes en el diseño de este tipo de algoritmos. Un elemento clave es la capacidad de cambiar de una búsqueda global en la etapa más temprana de la optimización, a una búsqueda más centrada una vez que se ha encontrado una solución aproximada.

  El estudio ha sido publicado en una revista JCR como una revisión del estado del arte

  - A. Valsecchi, S. Damas, J. Santamaría. Evolutionary intensity-based medical image registration: a review. *A aparecer en* Current Medical Imaging Reviews.

- *Diseño de nuevos métodos de registrado mediante algoritmos de optimización basados en computación evolutiva y otras metaheurísticas.*

  La experiencia adquirida a través de la revisión de la literatura se ha puesto en práctica en el diseño de dos nuevas técnicas. La propuesta inicial se ha basado en algoritmos genéticos; se han introducido y integrado con la estrategia de resolución múltiple componentes metaheurísticos avanzados, es decir, reinicio y frontera dinámica. Este enfoque específico al problema es el responsable último del excelente rendimiento del método. En la segunda propuesta, se ha obtenido una mejora adicional mediante el uso de la búsqueda dispersa para el problema

optimización correspondiente. Su diseño modular y flexible, junto al uso de una población reducida, han demostrado ser más fácil de personalizar y más eficaces para hacer frente a problemas de registrado.

Además del componente de optimización, ambos propuestas han sido diseñadas de forma modular, con el objetivo de soportar diferentes tipos de imágenes, modelos de transformación, métricas de similitud y otros componentes menores. Por último, los algoritmos se han integrado en una herramienta de registrado médico, Elastix, con el objetivo de facilitar la adopción por parte de los clínicos.

- *Validación del rendimiento de nuestras propuestas a través de un comparativa experimental y un análisis estadístico.*

  Hemos diseñado dos estudios experimentales muy diferentes entre sí. En el primer estudio se han empleado resonancias magnéticas simuladas del cerebro; se crearon una serie de escenarios de diferente complejidad mediante la selección de imágenes con diferentes niveles de ruido, lesiones anatómicas, y mediante la aplicación de transformaciones adicionales a las imágenes . La comparativa confronta nuestra propuesta contra amplia variedad de competidores, incluyendo tanto técnicas clásicas como algoritmos basados en características. El segundo estudio se centró en una aplicación médica común del registrado: la segmentación basada en atlas. Para dicho experimento se ha empleado un conjunto de resonancias magnéticas reales de cerebros, obtenidos a través de una colaboración con el Centre de Recherche del' Institut du Cerveau et de la Moelle épinière (ICM) en la Université Paris 6, París (Francia). Al final de ambos estudios, se obtuvieron resultados cuantitativos que fueron analizados mediante el uso de técnicas estadísticas adecuadas.

  Los dos algoritmos propuestos y las comparaciones experimentales han dado lugar a una serie de publicaciones:

  - A. Valsecchi, S. Damas, J. Santamaría, L. Marrakchi-Kacem. Intensity-based image registration using scatter search. *Enviado a* Artificial Intelligence in Medicine.
  - A. Valsecchi, S. Damas, J. Santamaría, L. Marrakchi-Kacem. Genetic Algorithms for Voxel-based Medical Image Registration. Proceedings of the IEEE Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI), 2013, Singapore, April 2013, pages 22–29.
  - A. Valsecchi, S. Damas, J. Santamaría, L. Marrakchi-Kacem. Voxel-based Medical Image Registration using Genetic Algorithms. Proceedings of the International Conference On Medical Imaging Using Bio-Inspired and Soft Computing (MIBISOC13), Brussels, Belgium, May 2013.
  - A. Valsecchi, S. Damas, J. Santamaría. An Image Registration Approach using Genetic Algorithms. Proceedings of the IEEE Congress on Evolutionary Computation (CEC) 2012, Brisbane, Australia, pages 416–423.

  - A. Valsecchi, S. Damas, J. Santamaría. Una Aproximación al Registrado de Imágenes Médicas con Algoritmos Genéticos. VIII Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados 2012, pages 667–674.

- *Aplicación de técnicas de ajuste automático a algoritmos de registrado de imagen.*

  Se ha estudiado la aplicación de técnicas de optimización de parámetros a los algoritmos de registrado. A través de un proceso estocástico conocido como carreras iteradas, los algoritmos se ajustan automáticamente para la aplicación especifica podiendo ofrecer los mejores resultados posibles sin la carga de un ajuste manual o exhaustiva. Además de demostrar la aplicabilidad de este tipo de técnicas al registrado, hemos desarrollado una comparativa más objetiva de los algoritmos de registrado a consecuencia del ajuste automático de sus parámetros. Se han alcanzado dos conclusiones. Por un lado, todos los algoritmos en la comparativa se han beneficiado del ajuste. Por otro lado, nuestras dos propuestas han proporcionado los mejores resultados, demostrando la eficacia de las metaheurísticas también en este contexto.

  Este estudio ha sido llevado a cabo en colaboración con laboratorio de Inteligencia Artificial de la Universidad Libre de Bruselas (Bélgica) . Nuestros resultados han sido presentados en las actas del congreso siguiente:

  - A. Valsecchi, J. Dubois-Lacoste, T. Stützle, S. Damas, J. Santamaría, L. Marrakchi-Kacem. Evolutionary Medical Image Registration using Automatic Parameter Tuning. Proceedings of the IEEE Congress on Evolutionary Computation (CEC) 2013, Cancun, June 2013, pages 1326–1333.

- *Integración de nuestros algoritmos en aplicaciones médicas mas amplias basadas en registrado de imágenes.*

  Hemos desarrollado un nuevo algoritmo de segmentación basado en level sets. Nuestro algoritmo de registrado basado en búsqueda dispersa se ha utilizado para llevar a cabo una segmentación a partir de un atlas, lo que resulta en una segmentación preliminar del objeto. Esta se utiliza como contorno inicial y como motor de la evolución del level set, unto con otros términos de fuerza que aprovechan la información de bordes y regiones. Hemos llevado a cabo una gran comparativa de métodos de segmentación en múltiples conjuntos de datos médicos, incluyendo imágenes histológicas, tomografías computarizadas e imágenes de resonancia magnética, en donde nuestro algoritmo supera a siete competidores de manera significativa.

  El estudio se ha realizado en colaboración con el Laboratorio de Sistemas Bio-Inspirados de la Universidad de Parma (Italia). Este ha sido aceptado para su publicación en una revista JCR:

  - P. Mesejo, A. Valsecchi, L. Marrakchi-Kacem, S. Cagnoni, S. Damas. Biomedical Image Segmentation using Geometric Deformable Models and Metaheuristics. *en prensa* Computerized Medical Imaging and Graphics.

A continuación, vamos a hablar de algunas de las líneas de investigación abiertas en relación con los temas tratados en este tesis. Además, presentamos algunas extensiones de nuestras propuestas que se desarrollarán como parte de los trabajos futuros.

- *Aplicación de nuestras propuestas a otros problemas de imagen médica.*

  El registrado se utiliza en innumerables aplicaciones en análisis de imágenes médicas, y obviamente, hemos considerado sólo un número limitado de ellas. Entre otros, podemos mencionar la imagen funcional y la resonancia magnética de alto campo como ejemplos de áreas que pueden proporcionar escenarios nuevos y desafiantes.

- *Consideración de otras metaheurísticas avanzadas para el optimizador.*

  Aunque el rendimiento de la búsqueda dispersa ha sido excelente, se podrían mejorar los resultados considerando otras metaheurísticas o aborando el problema desde otros enfoques. Un candidato ideal son los algoritmos meméticos. Estos pueden considerarse el estado del arte en problemas de optimización continua, y también encajan a la perfección en el patrón de dos componentes observado en la revisión de la literatura.

- *Desarrollo de implementaciones en GPU para hacer frente a aplicaciones en tiempo real.*

  Metaheurísticas, y técnicas de computación evolutiva en general, son computacionalmente caros y difíciles de aplicar en tiempo real. Nuestros algoritmos no son excepciones, sin embargo, su implementación en GPUs puede conducir a reduciones drásticas en el tiempo de cómputo necesario. Esto haría posible el uso de estas técnicas en el contexto del registrado de imagen intraoperatorio, comúnmente utilizado en sistemas de cirugía guiada por imagen.

Bibliography

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

[2] X. Y. Wang, S. Eberl, M. Fulham, S. Som, and D. D. Feng, ''Data registration and fusion,'' in *Biomedical information technology* (D. D. Feng, ed.), pp. 187–210, Burlington, USA: Academic Press, 2008.

[3] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Addison-Wesley, 2nd ed., 2001.

[4] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, 2007.

[5] S. Klein, M. Staring, and J. P. W. Pluim, ''Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines,'' *IEEE Trans. on Image Processing*, vol. 16, no. 12, pp. 2879–2890, 2007.

[6] P. J. Besl and N. D. McKay, ''A method for registration of 3D shapes,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, pp. 239–256, 1992.

[7] F. Glover and G. A. Kochenberger, eds., *Handbook of Metaheuristics*. Boston, USA: Kluwer Academic Publishers, 2003.

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

[9] M. Zhang, M. Koeppen, and S. Damas, ''Special issue on computational intelligence in computer vision and image processing,'' *Computational Intelligence Magazine, IEEE*, vol. 8, no. 1, 2013.

[10] S. Damas, O. Cordón, and J. Santamaría, ''Medical image registration using evolutionary computation: An experimental survey,'' *IEEE Computational Intelligence Magazine*, vol. 6, pp. 26 –42, nov. 2011.

[11] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, ''Nonrigid registration using free-form deformations: Application to breast mr images,'' *IEEE Transactions on Medical Imaging*, vol. 18, pp. 712–721, 1999.

[12] M. Svedlow, C. D. Mc-Gillem, and P. E. Anuta, ''Experimental examination of similarity measures and preprocessing methods used for image registration,'' in *Symposium on Machine Processing of Remotely Sensed Data* (P. Swain, D. Morrison, and D. Parks, eds.), vol. 4(A), (Indiana, USA), pp. 9–17, 1976.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] P. Chalermwat, T. El-Ghazawi, and J. LeMoigne, ''2-phase GA-based image registration on parallel clusters,'' *Future Gener. Comp. Sy.*, vol. 17, pp. 467–476, 2001.

[18] S. Zambanini, R. Sablatnig, H. Maier, and G. d. Langs, ''Automatic image-based assessment of lesion development during hemangioma follow-up examinations,'' *Artificial Intelligence in Medicine*, vol. 50, no. 2, pp. 83–94, 2010.

[19] Z. Zhang, ''Iterative point matching for registration of free-form curves and surfaces,'' *Int. J. Comput. Vision*, vol. 13, no. 2, pp. 119–152, 1994.

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

[21] S. Rusinkiewicz and M. Levoy, ''Efficient variants of the ICP algorithm,'' in *Third International Conference on 3D Digital Imaging and Modeling (3DIM'01)*, (Quebec, Canada), pp. 145–152, 2001.

[22] R. Woods, S. Grafton, J. Watson, N. Sicotte, and J. Mazziotta, ''Automated image registration: II. intersubject validation of linear and nonlinear models,'' *Journal of computer assisted tomography*, vol. 1, no. 22, pp. 153–65, 1998.

[23] K. De Jong, *Evolutionary Computation*. The MIT Press, 2002.

[24] A. Eiben and J. Smith, *Introduction to Evolutionary Computation*. Springer, Berlin, 2003.

[25] D. Fogel, *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. Wiley-IEEE Press, 2005.

[26] G. Dantzig, *Linear Programming and Extensions*. Princeton University Press, 1963.

[27] A. H. Land and A. G. Doig, ''An automatic method of solving discrete programming problems,'' *Econometrica*, vol. 28, no. 3, pp. 497–520, 1960.

[28] C. Blum and A. Roli, ''Metaheuristics in combinatorial optimization: Overview and conceptual comparison,'' *ACM Comput. Surv.*, vol. 35, pp. 268–308, Sept. 2003.

[29] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, ''Optimization by simulated annealing,'' *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[30] F. Glover and C. McMillan, ''The general employee scheduling problem. an integration of {MS} and {AI},'' *Computers & Operations Research*, vol. 13, no. 5, pp. 563 – 573, 1986.

[31] H. R. Lourenço, O. C. Martin, and T. Stützle, ''Iterated local search,'' in *Handbook of Metaheuristics* (F. Glover and G. Kochenberger, eds.), pp. 321–353, Kluwer Academic Publishers, 2003.

[32] T. A. Feo and M. G. Resende, ''A probabilistic heuristic for a computationally difficult set covering problem,'' *Operations Research Letters*, vol. 8, no. 2, pp. 67 – 71, 1989.

[33] P. Moscato, ''On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms,'' Tech. Rep. C3P Report 826, California Institute of Technology, 1989.

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

[35] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor: The University of Michigan Press, 1975.

[36] H.-G. Beyer and H.-P. Schwefel, ''Evolution strategies – a comprehensive introduction,'' *Natural Computing*, vol. 1, no. 1, pp. 3–52, 2002.

[37] J. R. Koza, *Genetic programming: on the programming of computers by means of natural selection*. Cambridge, MA, USA: MIT Press, 1992.

[38] R. Storn and K. Price, ''Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces,'' *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[39] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*. Oxford, 1999.

[40] M. Dorigo, *Optimization, Learning and Natural Algorithms*. PhD thesis, Politecnico di Milano, Italy, 1992.

[41] J. Kennedy and R. Eberhart, ''Particle swarm optimization,'' in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, pp. 1942–1948, 1995.

[42] M. Laguna and R. Martí, *Scatter search: methodology and implementations in C*. Boston, USA: Kluwer Academic Publishers, 2003.

[43] F. Glover, ''A template for scatter search and path relinking,'' in *Artificial Evolution* (J.-K. Hao, E. Lutton, E. Ronald, M. Schoenauer, and D. Snyers, eds.), vol. 1363 of *Lecture Notes in Computer Science*, pp. 1–51, Springer Berlin Heidelberg, 1998.

[44] G. Pascale and L. Troiano, ''A niche based genetic algorithm for image registration,'' in *Ninth International Conference on Enterprise Information Systems (ICEIS 2007)*, pp. 342–347, 2007.

[45] K. Brunnström and A. Stoddart, ''Genetic algorithms for free-form surface matching,'' in *International Conference of Pattern Recognition*, (Vienna, Germany), pp. 689–693, 1996.

[46] O. Cordón and S. Damas, ''Image Registration with Iterated Local Search,'' *Journal of Heuristics*, vol. 12, pp. 73–94, 2006.

[47] O. Cordón, S. Damas, J. Santamaría, and R. Martí, ''Scatter Search for the 3D Point Matching Problem in Image Registration,'' *INFORMS Journal on Computing*, vol. 20, pp. 55–68, 2008.

[48] M. Jenkinson and S. Smith, ''A global optimisation method for robust affine registration of brain images,'' *Medical Image Analysis*, vol. 5, no. 2, pp. 143–156, 2001.

[49] J. Santamaría, O. Cordón, S. Damas, J. García-Torres, and A. Quirin, ''Performance evaluation of memetic approaches in 3D reconstruction of forensic objects,'' *Soft Comput.*, vol. 13, no. 8-9, pp. 883–904, 2009.

[50] Y. S. Ong, M. Lim, N. Zhu, and K. Wong, ''Classification of adaptive memetic algorithms: a comparative study,'' *IEEE Transaction on Systems, Man, and Cybernetics*, vol. 36, no. 1, pp. 141–152, 2006.

[51] Y. S. Ong, M. H. Lim, and X. Chen, ''Memetic computation — Past, present & future,'' *IEEE Computational Intelligence Magazine*, vol. 5, no. 2, pp. 24–31, 2010.

[52] A. Valsecchi, J. Dubois-Lacoste, T. Stützle, S. Damas, J. Santamaría, and L. Marrakchi-Kacem, ''Evolutionary medical image registration using automatic parameter tuning,'' in *IEEE Congress on Evolutionary Computation*, 2013.

[53] C. Robertson and R. B. Fisher, ''Parallel Evolutionary Registration of Range Data,'' *Computer Vision and Image Understanding*, vol. 87, pp. 39–50, 2002.

[54] K. Fok, T. Wong, and M. Wong, ''Evolutionary computing on consumer graphics hardware,'' *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 69–78, 2007.

[55] L. Silva, O. R. P. Bellon, and K. L. Boyer, *Robust range image registration using genetic algorithms and the surface interpetenetration measure*. Singapore: World Scientific, 2005.

[56] J. Santamaría, O. Cordón, and S. Damas, ''A comparative study of state-of-the-art evolutionary image registration methods for 3D modeling,'' *Comput. Vis. Image Underst.*, vol. 115, no. 9, pp. 1340–1354, 2011.

[57] 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.

[58] 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.

[59] 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.

[60] C. K. Chow, H. T. Tsui, and T. Lee, ''Surface registration using a dynamic genetic algorithm,'' *Pattern Recogn.*, vol. 37, pp. 105–117, 2004.

[61] M. Wachowiak and A. Elmaghraby, ''The continuous tabu search as an optimizer for 2d-to-3d biomedical image registration,'' in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2001* (W. Niessen and M. Viergever, eds.), vol. 2208 of *Lecture Notes in Computer Science*, pp. 1273–1274, Springer Berlin Heidelberg, 2001.

[62] F. Glover and M. Laguna, *Tabu search*. Norwell, MA, EE. UU.: Kluwer Academic, 1997.

[63] 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.

[64] R. Battiti and G. Tecchiolli, ''The continuous reactive tabu search: Blending combinatorial optimization and stochastic search for global optimization,'' *Annals of Operations Research*, vol. 63, no. 2, pp. 151–188, 1996.

[65] J. A. Nelder and R. Mead, ''A Simplex Method for Function Minimization,'' *The Computer Journal*, vol. 7, pp. 308–313, Jan. 1965.

[66] O. Cordón, S. Damas, and J. Santamaría, ''A Fast and Accurate Approach for 3D Image Registration using the Scatter Search Evolutionary Algorithm,'' *Pattern Recogn. Lett.*, vol. 27, no. 11, pp. 1191–1200, 2006.

[67] O. Cordón, S. Damas, and J. Santamaría, ''Feature-based image registration by means of the chc evolutionary algorithm,'' *Image Vision Comput.*, vol. 22, pp. 525–533, 2006.

[68] I. D. Falco, A. D. Cioppa, D. Maisto, and E. Tarantino, ''Differential evolution as a viable tool for satellite image registration,'' *Applied Soft Computing*, vol. 8, no. 4, pp. 1453 – 1462, 2008.

[69] H.-G. Beyer and K. Deb, ''On self-adaptive features in real-parameter evolutionary algorithms,'' *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 3, pp. 250–270, 2001.

[70] F. J. Solis and R. J.-B. Wets, ''Minimization by random search techniques,'' *Mathematics of operations research*, vol. 6, no. 1, pp. 19–30, 1981.

[71] M. J. D. Powell, ''An efficient method for finding the minimum of a function of several variables without calculating derivatives,'' *The Computer Journal*, vol. 7, pp. 155–162, Jan. 1964.

[72] 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.

[73] M. Wachowiak, R. Smolíková, Y. Zheng, J. Zurada, and A. Elmaghraby, ''An approach to multimodal biomedical image registration utilizing particle swarm optimization,'' *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 289–301, 2004. cited By (since 1996) 130.

[74] M. Clerc and J. Kennedy, ''The particle swarm - explosion, stability, and convergence in a multidimensional complex space,'' *Evolutionary Computation, IEEE Transactions on*, vol. 6, no. 1, pp. 58–73, 2002.

[75] G. Kagadis, K. Delibasis, G. Matsopoulos, N. Mouravliansky, P. Asvestas, and G. Nikiforidis, ''A comparative study of surface- and volume-based techniques for the automatic registration between ct and spect brain images,'' *Medical Physics*, vol. 29, no. 2, pp. 201–213, 2002. cited By (since 1996) 24.

[76] Q. Wang and X. Li, ''Application of improved genetic algorithm in practical medical image registration,'' *International Journal of Digital Content Technology and its Applications*, vol. 5, no. 10, pp. 60–67, 2011. cited By (since 1996) 9.

[77] X. Xu and R. Dony, ''Differential evolution with powell's direction set method in medical image registration,'' vol. 1, (Arlington, VA), pp. 732–735, 2004. cited By (since 1996) 8; Conference of 2004 2nd IEEE International Symposium on

Biomedical Imaging: Macro to Nano; Conference Date: 15 April 2004 through 18 April 2004; Conference Code: 64593.

[78] J. P. W. Pluim, J. Maintz, and M. Viergever, ''Image registration by maximization of combined mutual information and gradient information,'' *Medical Imaging, IEEE Transactions on*, vol. 19, no. 8, pp. 809–814, 2000.

[79] H. Talbi and M. Batouche, ''Hybrid particle swarm with differential evolution for multimodal image registration,'' vol. 3, (Hammamet), pp. 1567–1572, 2004. cited By (since 1996) 19; Conference of 2004 IEEE International Conference on Industrial Technology, ICIT; Conference Date: 8 December 2004 through 10 December 2004; Conference Code: 66063.

[80] S. Winter, B. Brendel, I. Pechlivanis, K. Schmieder, and C. Igel, ''Registration of ct and intraoperative 3-d ultrasound images of the spine using evolutionary and gradient-based methods,'' *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 3, pp. 284–296, 2008. cited By (since 1996) 25.

[81] N. Hansen and S. Kern, ''Evaluating the cma evolution strategy on multimodal test functions,'' in *Parallel Problem Solving from Nature - PPSN VIII* (X. Yao, E. Burke, J. Lozano, J. Smith, J. Merelo-Guervós, J. Bullinaria, J. Rowe, P. Tiňo, A. Kabán, and H.-P. Schwefel, eds.), vol. 3242 of *Lecture Notes in Computer Science*, pp. 282–291, Springer Berlin Heidelberg, 2004.

[82] R. He and P. Narayana, ''Global optimization of mutual information: Application to three-dimensional retrospective registration of magnetic resonance images,'' *Computerized Medical Imaging and Graphics*, vol. 26, no. 4, pp. 277–292, 2002. cited By (since 1996) 44.

[83] D. Jones, C. Perttunen, and B. Stuckman, ''Lipschitzian optimization without the lipschitz constant,'' *Journal of Optimization Theory and Applications*, vol. 79, no. 1, pp. 157–181, 1993.

[84] R. P. Woods, S. T. Grafton, J. D. G. Watson, N. L. Sicotte, and J. C. Mazziotta, ''Automated image registration: Ii. intersubject validation of linear and nonlinear models,'' *Journal of Computer Assisted Tomography*, vol. 22, no. 1, pp. 153–165, 1998.

[85] N. Castellanos, P. Angel, and V. Medina, ''Nonrigid medical image registration technique as a composition of local warpings,'' *Pattern Recognition*, vol. 37, no. 11, pp. 2141–2154, 2004. cited By (since 1996) 7.

[86] X. Li, P. Zhang, R. Brisman, and G. Kutcher, ''Use of simulated annealing for optimization of alignment parameters in limited mri acquisition volumes of the brain,'' *Medical Physics*, vol. 32, no. 7, pp. 2363–2370, 2005. cited By (since 1996) 9.

[87] Y.-W. Chen, C.-L. Lin, and A. Mimori, ''Multimodal medical image registration using particle swarm optimization,'' vol. 3, (Kaohsiung), pp. 127–131, 2008. cited By (since 1996) 5; Conference of 8th International Conference on Intelligent Systems Design and Applications, ISDA 2008; Conference Date: 26 November 2008 through 28 November 2008; Conference Code: 76231.

[88] Q. Li and I. Sato, ''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)*, vol. 4682 LNAI, pp. 1120–1130, 2007. cited By (since 1996) 6; Conference of 3rd International Conference on Intelligent Computing, ICIC 2007; Conference Date: 21 August 2007 through 24 August 2007; Conference Code: 71052.

[89] D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens, ''Temporal subtraction of thorax cr images using a statistical deformation model,'' *IEEE Transactions on Medical Imaging*, vol. 22, no. 11, pp. 1490–1504, 2003. cited By (since 1996) 20.

[90] J. Du, S. Tang, T. Jiang, and Z. Lu, ''Intensity-based robust similarity for multimodal image registration,'' *International Journal of Computer Mathematics*, vol. 83, no. 1, pp. 49–57, 2006. cited By (since 1996) 6.

[91] P. Meer, D. Mintz, A. Rosenfeld, and D. Kim, ''Robust regression methods for computer vision: A review,'' *International Journal of Computer Vision*, vol. 6, no. 1, pp. 59–70, 1991.

[92] D. Zhou, J. Sun, C.-H. Lai, W. Xu, and X. Lee, ''An improved quantum-behaved particle swarm optimization and its application to medical image registration,'' *International Journal of Computer Mathematics*, vol. 88, no. 6, pp. 1208–1223, 2011. cited By (since 1996) 4.

[93] J. Sun, B. Feng, and W. Xu, ''Particle swarm optimization with particles having quantum behavior,'' in *Evolutionary Computation, 2004. CEC2004. Congress on*, vol. 1, pp. 325–331 Vol.1, 2004.

[94] T. Inaba, L. He, K. Suzuki, K. Murakami, and Y. Chao, ''A genetic-algorithm-based temporal subtraction for chest radiographs,'' *JACIII*, vol. 13, no. 3, pp. 289–296, 2009.

[95] R. Poli, ''Analysis of the publications on the applications of particle swarm optimisation,'' *J. Artif. Evol. App.*, vol. 2008, pp. 4:1–4:10, Jan. 2008.

[96] 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 Transactions on Medical Imaging*, vol. 17, pp. 463–468, 1998.

[97] J. Fitzpatrick, J. Grefenstette, and D. Gucht, ''Image registration by genetic search,'' in *IEEE Southeast Conference*, (Louisville, EEUU), pp. 460–464, 1984.

[98] V. R. Mandava, J. M. Fitzpatrick, and D. R. Pickens, ''Adaptive search space scaling in digital image registration,'' *IEEE Transactions on Medical Imaging*, vol. 8, no. 3, pp. 251–262, 1989.

[99] P. W. M. Tsang, ''A genetic algorithm for aligning object shapes,'' *Image and Vision Computing*, vol. 15, pp. 819–831, 1997.

[100] L. J. Eshelman and J. D. Schaffer, ''Real-coded genetic algorithms and interval-schemata.,'' in *Foundation of Genetic Algorithms 2* (D. L. Whitley, ed.), (San Mateo, CA), pp. 187–202, Morgan Kaufmann., 1993.

[101] J. Demšar, ''Statistical comparisons of classifiers over multiple data sets,'' *Journal of Machine Learning Research*, vol. 7, pp. 1–30, Dec. 2006.

[102] P. Nemenyi, *Distribution-free multiple comparisons*. PhD thesis, Princeton University, 1963.

[103] M. Friedman, ''A comparison of alternative tests of significance for the problem of m rankings,'' *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.

[104] S. Holm, ''A simple sequentially rejective multiple test procedure,'' *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[105] 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.

[106] ''Elastix webpage.'' Accessed: 1 April 2013.

[107] L. Ibáñez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide*. Kitware, Inc. ISBN 1-930934-15-7, second ed., 2005.

[108] 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.

[109] 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.), pp. 299–307, 2002.

[110] 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), pp. 354–389, IEEE, 1992.

[111] B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard, ''Image registration via level-set motion: Applications to atlas-based segmentation,'' *Medical Image Analysis*, vol. 7, no. 1, pp. 1–20, 2003.

[112] C. Poupon, F. Poupon, L. Allirol, and J.-F. Mangin, ''A database dedicated to anatomo-functional study of human brain connectivity,'' in *12th Annual Meeting of the Organization for Human Brain Mapping*, no. 646, (Florence, Italy), 2006.

[113] D. Rueckert and J. A. Schnabel, ''Medical image registration,'' in *Biomedical Image Processing* (T. M. Deserno, ed.), Biological and Medical Physics, Biomedical Engineering, pp. 131–154, Springer Berlin Heidelberg, 2011.

[114] L. R. Dice, ''Measures of the amount of ecologic association between species,'' *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[115] F. Glover, M. Laguna, and R. Martí, ''Scatter search,'' in *Advances in Evolutionary Computation: Theory and Applications* (A. Ghosh and S. Tsutsui, eds.), pp. 519–537, New York: Springer-Verlag, 2003.

[116] M. Lozano, F. Herrera, N. Krasnogor, and D. Molina, ''Real-coded memetic algorithms with crossover hill-climbing,'' *Evolut. Comput.*, vol. 12, no. 3, pp. 273–302, 2004.

[117] F. Hutter, H. H. Hoos, K. Leyton-Brown, and T. Stützle, ''ParamILS: an automatic algorithm configuration framework,'' *Journal of Artificial Intelligence Research*, vol. 36, pp. 267–306, Oct. 2009.

[118] F. Hutter, H. H. Hoos, and K. Leyton-Brown, ''Sequential model-based optimization for general algorithm configuration,'' in *Learning and Intelligent Optimization, 5th International Conference, LION 5*, Lecture Notes in Computer Science, Springer, Heidelberg, Germany, 2011.

[119] T. Bartz-Beielstein, *Experimental Research in Evolutionary Computation: The New Experimentalism*. Berlin, Germany: Springer, 2006.

[120] M. López-Ibáñez, J. Dubois-Lacoste, T. Stützle, and M. Birattari, ''The irace package, iterated race for automatic algorithm configuration,'' Tech. Rep. TR/IRIDIA/2011-004, IRIDIA, Université Libre de Bruxelles, Belgium, 2011.

[121] F. Hutter, H. H. Hoos, K. Leyton-Brown, and K. P. Murphy, ''An experimental investigation of model-based parameter optimisation: SPO and beyond,'' in *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2009* (F. Rothlauf, ed.), pp. 271–278, New York, NY: ACM Press, 2009.

[122] P. Balaprakash, M. Birattari, and T. Stützle, ''Improvement strategies for the F-race algorithm: Sampling design and iterative refinement,'' in *Hybrid Metaheuristics* (T. Bartz-Beielstein, M. J. Blesa, C. Blum, B. Naujoks, A. Roli, G. Rudolph, and M. Sampels, eds.), vol. 4771 of *Lecture Notes in Computer Science*, pp. 108–122, Springer, Heidelberg, Germany, 2007.

[123] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[124] M. Birattari, T. Stützle, L. Paquete, and K. Varrentrapp, ''A racing algorithm for configuring metaheuristics,'' in *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2002* (W. B. Langdon *et al.*, eds.), pp. 11–18, Morgan Kaufmann Publishers, San Francisco, CA, 2002.

[125] M. Birattari, *The Problem of Tuning Metaheuristics as Seen from a Machine Learning Perspective*. PhD thesis, Université Libre de Bruxelles, Brussels, Belgium, 2004.

[126] Birattari, *Tuning Metaheuristics: A Machine Learning Perspective*, vol. 197 of *Studies in Computational Intelligence*. Berlin/Heidelberg, Germany: Springer, 2009.

[127] O. Maron and A. W. Moore, ''The racing algorithm: Model selection for lazy learners,'' *Artificial Intelligence Research*, vol. 11, no. 1–5, pp. 193–225, 1997.

[128] M. Birattari, Z. Yuan, P. Balaprakash, and T. Stützle, ''F-race and iterated F-race: An overview,'' in *Experimental Methods for the Analysis of Optimization Algorithms* (T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuss, eds.), pp. 311–336, Berlin, Germany: Springer, 2010.

[129] J. Dubois-Lacoste, M. López-Ibáñez, and T. Stützle, ''Effective hybrid stochastic local search algorithms for biobjective permutation flowshop scheduling,'' in *Hybrid Metaheuristics* (M. J. Blesa, C. Blum, L. Di Gaspero, A. Roli, M. Sampels, and A. Schaerf, eds.), vol. 5818 of *Lecture Notes in Computer Science*, pp. 100–114, Springer, Heidelberg, Germany, 2009.

[130] J. Dubois-Lacoste, M. López-Ibáñez, and T. Stüetzle, ''A hybrid TP+PLS algorithm for bi-objective flow-shop scheduling problems,'' *Computers & Operations Research*, vol. 38, no. 8, pp. 1219–1236, 2011.

[131] J. Dubois-Lacoste, M. López-Ibáñez, and T. Stützle, ''Automatic configuration of state-of-the-art multi-objective optimizers using the TP+PLS framework,'' in *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2011* (N. Krasnogor and P. L. Lanzi, eds.), pp. 2019–2026, New York, NY: ACM Press, 2011.

[132] D. L. Pham, C. Xu, and J. L. Prince, ''Current Methods in Medical Image Segmentation,'' *Annual Review of Biomedical Engineering*, vol. 2, pp. 315–337, 2000.

[133] L. He, Z. Peng, B. Everding, X. Wang, C. Y. Han, K. L. Weiss, and W. G. Wee, ''A comparative study of deformable contour methods on medical image segmentation,'' *Image and Vision Computing*, vol. 26, pp. 141–163, 2008.

[134] T. Heimann and H.-P. Meinzer, ''Statistical shape models for 3d medical image segmentation: a review,'' *Medical Image Analysis*, vol. 13, pp. 543–563, 2009.

[135] J. Malik, S. Belongie, T. Leung, and J. Shi, ''Contour and texture analysis for image segmentation,'' *International Journal of Computer Vision*, vol. 43, pp. 7–27, June 2001.

[136] Y. Zhang, B. J. Matuszewski, L.-K. Shark, and C. J. Moore, ''Medical Image Segmentation Using New Hybrid Level-Set Method,'' in *Procs. of the International Conference BioMedical Visualization: Information Visualization in Medical and Biomedical Informatics*, pp. 71–76, 2008.

[137] M. Chupin, A. Hammers, R. S. N. Liu, O. Colliot, J. Burdett, E. Bardinet, J. S. Duncan, L. Garnero, and L. Lemieux, ''Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: Method and validation,'' *NeuroImage*, vol. 46, no. 3, pp. 749–761, 2009.

[138] D. Terzopoulos and K. Fleischer, ''Deformable models,'' *The Visual Computer*, vol. 4, pp. 306–331, 1988.

[139] M. Kass, A. Witkin, and D. Terzopoulos, ''Snakes: Active contour models,'' *International Journal of Computer Vision*, vol. 1, pp. 321–331, 1988.

[140] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, ''Active shape models-their training and application,'' *Computer Vision and Image Understanding*, vol. 61, pp. 38–59, 1995.

[141] T. F. Cootes, G. J. Edwards, and C. J. Taylor, ''Active Appearance Models,'' in *Proc. of the European Conference on Computer Vision*, vol. 2, pp. 484–498, 1998.

[142] T. F. Cootes, G. Edwards, and C. Taylor, ''Comparing active shape models with active appearance models,'' in *Procs. of British Machine Vision Conference*, pp. 173–182, 1999.

[143] M. Bro-Nielsen, ''Active nets and cubes,'' tech. rep., 1994.

[144] B. B. Kimia, A. R. Tannenbaum, and S. W. Zucker, ''Shapes, shocks, and deformations i: The components of two-dimensional shape and the reaction-diffusion space,'' *International Journal of Computer Vision*, vol. 15, pp. 189–224, 1994.

[145] R. Kimmel, A. Amir, and A. M. Bruckstein, ''Finding shortest paths on surfaces using level sets propagation,'' *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, no. 6, pp. 635–640, 1995.

[146] G. Sapiro and A. Tannenbaum, ''Affine invariant scale-space,'' *International Journal of Computer Vision*, vol. 11, no. 1, pp. 25–44, 1993.

[147] S. Osher and J. Sethian, ''Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations,'' *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988.

[148] J. Sethian, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 1999.

[149] S. J. Osher and R. P. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2002.

[150] P. Mesejo, R. Ugolotti, F. D. Cunto, M. Giacobini, and S. Cagnoni, ''Automatic hippocampus localization in histological images using differential evolution-based deformable models,'' *Pattern Recognition Letters*, vol. 34, no. 3, pp. 299 – 307, 2013.

[151] M. Heydarian, M. Noseworthy, M. Kamath, C. Boylan, and W. Poehlman, ''Optimizing the level set algorithm for detecting object edges in mr and ct images,'' *IEEE Trans. on Nuclear Science*, vol. 56, no. 1, pp. 156 –166, 2009.

[152] L. Ballerini, ''Genetic snakes for medical images segmentation,'' in *Evolutionary Image Analysis, Signal Processing and Telecommunications*, vol. 1596, pp. 59–73, 1999.

[153] D.-H. Chen and Y.-N. Sun, ''A self-learning segmentation framework—the taguchi approach,'' *Computerized Medical Imaging and Graphics*, vol. 24, no. 5, pp. 283 – 296, 2000.

[154] C. McIntosh and G. Hamarneh, ''Medial-based deformable models in non-convex shape-spaces for medical image segmentation using genetic algorithms,'' *IEEE Trans. on Medical Imaging*, vol. 31, no. 1, pp. 33–50, 2012.

[155] C.-Y. Hsu, C.-Y. Liu, and C.-M. Chen, ''Automatic segmentation of liver pet images,'' *Computerized Medical Imaging and Graphics*, vol. 32, no. 7, pp. 601 – 610, 2008.

[156] O. Ibáñez, N. Barreira, J. Santos, and M. G. Penedo, ''Genetic approaches for topological active nets optimization,'' *Pattern Recognition*, vol. 42, pp. 907–917, May 2009.

[157] J. Novo, J. Santos, and M. G. Penedo, ''Topological active models optimization with differential evolution,'' *Expert Systems with Applications*, vol. 39, no. 15, pp. 12165–12176, 2012.

[158] N. Bova, Óscar Ibáñez, and O. Cordón, ''Image segmentation using extended topological active nets optimized by scatter search,'' *IEEE Computational Intelligence Magazine*, vol. 8, no. 1, pp. 16–32, 2013.

[159] P. Ghosh, M. Mitchell, J. A. Tanyi, and A. Hung, ''A genetic algorithm-based level set curve evolution for prostate segmentation on pelvic ct and mri images,'' in *Biomedical Image Analysis and Machine Learning Technologies*, pp. 127–149, 2010.

[160] D. Feltell and L. Bai, ''3d level set image segmentation refined by intelligent agent swarm,'' in *Procs. of IEEE Congress on Evolutionary Computation*, pp. 1–8, 2010.

[161] P. Mesejo and S. Cagnoni, ''An experimental study on the automatic segmentation of in situ hybridization-derived images,'' in *Proc. on 1st International Conference on Medical Imaging using Bio-Inspired and Soft Computing (MIBISOC'13)*, 2013. In Press.

[162] T. F. Chan and L. A. Vese, ''Active contours without edges,'' *IEEE Trans. on Image Processing*, vol. 10, pp. 266–277, Feb. 2001.

[163] B. Li, S. Member, S. T. Acton, and S. Member, ''Active contour external force using vector field convolution for image segmentation,'' *IEEE Trans. on Image Processing*, vol. 16, pp. 2096–2106, 2007.

[164] C. Xu and J. L. Prince, ''Gradient Vector Flow: A New External Force for Snakes,'' in *Procs. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 66–71, 1997.

[165] Y. Shi and W. C. Karl, ''A real-time algorithm for the approximation of level-set-based curve evolution,'' *IEEE Trans. on Image Processing*, vol. 17, no. 5, pp. 645–656, 2008.

[166] Allen Institute for Brain Science, ''Allen Reference Atlases.'' http://mouse.brain-map.org, 2004-2006.

[167] P. Mesejo, R. Ugolotti, S. Cagnoni, F. Di Cunto, and M. Giacobini, ''Automatic Segmentation of Hippocampus in Histological Images of Mouse Brains using Deformable Models and Random Forest,'' in *Procs. of Symposium on Computer-Based Medical Systems*, 2012.

[168] N. Otsu, ''A threshold selection method from gray-level histograms,'' *IEEE Trans. on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62 –66, 1979.

[169] L. Breiman, ''Random forests,'' *Maching Learning*, vol. 45, pp. 5–32, 2001.

[170] S. Aja-Fernandez, G. Vegas-Sanchez-Ferrero, and M. Martin Fernandez, ''Soft thresholding for medical image segmentation,'' in *Proc. International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4752 –4755, 2010.

[171] V. Caselles, R. Kimmel, and G. Sapiro, ''Geodesic active contours,'' *International Journal of Computer Vision*, vol. 22, pp. 61–79, 1997.

[172] W. H. Kruskal, ''Historical notes on the wilcoxon unpaired two-sample test,'' *Journal of the American Statistical Association*, vol. 52, no. 279, pp. 356–360, 1957.