Applied Soft Computing Journal 97 (2020) 106793

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Performance analysis of real-coded evolutionary algorithms under a computationally expensive optimization scenario: 3D–2D Comparative Radiography

Oscar Gómez [a,b,*], Oscar Ibáñez [a,b,c], Andrea Valsecchi [a,b,c], Enrique Bermejo [a,b,d], Daniel Molina [a,b], Oscar Cordón [a,b]

[a] Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain
[b] Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, Granada, Spain
[c] Panacea Cooperative Research S. Coop., Ponferrada, Spain
[d] National Research Institute of Police Science (NRIPS), Kashiwanoha, Kashiwa, Chiba 277-0882, Japan

ARTICLE INFO

Article history:
Received 21 May 2019
Received in revised form 22 September 2020
Accepted 7 October 2020
Available online 14 October 2020

Keywords:
Comparative radiography
Computer vision
Evolutionary computation
3D–2D evolutionary image registration

ABSTRACT

Real-coded evolutionary algorithms have solved numerous real-world optimization problems. In this work, we aim to analyze the behavior and robustness of several real-coded evolutionary algorithms from the state of the art in a challenging real world optimization problem. This optimization problem consists on the superimposition of 3D and 2D images of skeletal structures (i.e. bones and cavities) based on their silhouette. This task is required for the automation of a forensic identification technique known as comparative radiography, via the generation of the best projection of the 3D image with respect to the 2D image. This superimposition problem was tackled in a recent proposal using an evolutionary 3D–2D image registration method based on differential evolution. However, the results obtained were insufficient for its use in real scenarios, due to: (1) the complexity and multi-modality of search space, despite the reduced number of parameters to be optimized (7 in its simple version and 9 in a more complex one, proposed in this work); and (2) the high computational cost of generating and evaluating a superimposition. Particularly, we have performed a rigorous comparative study of six state-of-the-art real-coded evolutionary algorithms (DE, L-SHADE, CMA-ES, BIPOP-CMA-ES, CRO-SL, and MVMO-SH) with synthetic images of three forensic anatomical structures (frontal sinuses, clavicles, and patellae), showing that the best results are always obtained by MVMO-SH in terms of precision, robustness and computational cost. Furthermore, we have validated the quality of the superimpositions obtained by the evolutionary image registration method using MVMO-SH with real images of frontal sinuses. We have performed the comparison of 50 head radiographs and 50 3D images, resulting in 2,500 cross-comparisons (50 positive and 2,450 negatives). The obtained results are promising since the superimpositions obtained allowed us to filter out 88% of the possible candidates with 0 error rate in a fully automatic manner, showing the high quality of the superimposition obtained.

1. Introduction

Evolutionary algorithms (EAs) [1,2] are global optimization techniques inspired by biological evolution for solving optimization problems. These algorithms have been successfully applied in many real-world optimization problems with complex optimization functions [3] because they are able to obtain competitive results without requiring specific features to the problem to optimize, as well as they can tackle non-linear, non-differentiable, non-convex and multi-modal functions.

Among real-world problems, there are problems especially challenging since they require a significant amount of computational resources and time to evaluate just a single candidate solution. The solution procedure for these problems is called expensive optimization [4]. In these frameworks, the optimization algorithm must provide accurate solutions with a very reduced number of evaluations. Under these constraints the majority of EAs are not suitable because they need a significant higher number of evaluations to obtain a competitive solution. In recent years, numerous research works [5,6] and competitions [7,8] have studied fast convergence EAs, which have obtained a good trade-off between quality and computational time.

* Corresponding author at: Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain.
E-mail address: ogomez@decsai.ugr.es (O. Gómez).

https://doi.org/10.1016/j.asoc.2020.106793
1568-4946/

In this work, we tackle one of these complex and computationally expensive optimization problems, required for the automation of a forensic identification technique known as comparative radiography (CR) [9]. It involves the superimposition of *ante-mortem* (AM) radiographs and *post-mortem* (PM) 3D images of an skeletal structure (i.e. bones and cavities) based on its morphological silhouette (see Section 3 for further details). This superimposition task can be modeled as an image registration (IR) problem [10,11], where the best 2D projection of the PM 3D image, with respect to the AM 2D image, is searched by iteratively optimizing the parameters of the projective transformation. The superimposition problem is complex (see Section 3 for further details). Most 3D–2D IR approaches are designed for a controllable set-up, which is a common situation in many medical domains [11–13]. However, these assumptions are not suitable for CR since the AM radiograph is generally taken under unknown conditions (neither the acquisition set up, the pose nor the radiograph device are known in advance) as well as to the presence of occlusions in the skeletal structures to be registered. Therefore, the search for the optimal solution in the CR scenario is more challenging, despite the low number of parameters to be optimized (7 with a simple perspective transformation and 9 with a more complex one, see Section 4.1 for further details).

In [14], we studied both different numerical optimization methods and a simple real-coded evolutionary algorithm (RCEA), differential evolution (DE), to tackle this complex real-parameter IR optimization problem without any assumption about the initial perspective projection parameters (i.e. optimization parameters). While numerical methods' accuracy proved to be insufficient, DE showed a good performance, highlighting the potential of evolutionary IR approaches [15]. However, the robustness of the method was insufficient (due to the complexity of the search space landscape, see Section 4.3 for further details) and computationally expensive (due to the high computational cost of generating and evaluating a particular projection, see Section 4.1 for further details). Therefore, we aim to analyze the performance of several high performance RCEAs [2], with particular focus on those tested in complex real-world problems as well as in expensive optimization competitions from the IEEE Congress on Evolutionary Computation (CEC), in order to determine the influence of the considered RCEA to improve the accuracy, robustness, and time required to obtain a good superposition for the CR problem.

The goal of this paper is thus two-fold. Firstly, to study the performance, robustness and convergence speed of several state-of-the-art EAs in the superimposition problem using synthetic images of three skeletal structures (clavicles [16], patellae [17], and frontal sinuses [18]). These three skeletal structures have been utilized in the CR literature for forensic identification, as well as in the state-of-the-art evolutionary image registration for CR [14] allowing us a fair comparison. Secondly, to study if the quality of the superimpositions obtained by the best RCEA in real images of frontal sinuses is sufficient for identification and/or *sort listing* using the identification methodology proposed in [14]. With this aim, we have also proposed and validated a new projective transformation that can reproduce the perspective distortion of any kind of radiograph.

This paper is structured as follows. Section 2 briefly reviews the current state of the art in evolutionary IR and RCEAs, and justifies the choice of the algorithms studied in this work. Section 3 briefly reviews the CR technique and related works. Section 4 describes the proposed IR methodology to tackle the superimposition problem (including a new projective transformation) and the study of the complexity of the search space of the superimposition problem. Section 5 presents experiments and results. Section 6 details the conclusions.

2. Background and justification

The optimization problems underlying IR methods based on the direct search of the real-coded transformation parameters are complex. This is particularly true in real-world problems as CR, as stated in Section 1. RCEAs [1,19] have improved the results obtained by traditional methods in many IR problems [20–22]. In these works several studies are performed benchmarking traditional and evolutionary IR methods. Traditional IR methods have shown to be insufficient to tackle complex IR problems, especially those where an initial solution close to the ground truth (GT) one is not known, ending in non accurate local minimum. As a consequence, the interest on evolutionary IR approaches based on RCEAs have grown over the last two decades [15,23,24], with 575 works published until 2020.[1]

Among the classical RCEAs, DE [25] and Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [26] have shown a great behavior on global numerical optimization problems. They have performed extremely well in those tackled in the competitions handled within the IEEE CEC [27,28], making them useful for many real-world problems. In particular, DE is widely extended as a result of its reduced number of parameters to fine tune [29], its robustness, and its fast convergence. DE has demonstrated an excellent performance on many IR problems [30,31] and specifically with the only proposal dealing with the CR problem from an IR perspective [14]. Furthermore, several self-adaptive DE approaches proposed in the literature yielded better results than the classical DE in many different problems [29]. Among them, a self-adaptive DE approach with a linear reduction of population and an external memory of elite solutions (to enforce diversity in the mutation) called L-SHADE [32] has shown a very significant accuracy. L-SHADE ranked on the first positions at the IEEE CEC2014 competition on real-parameter single objective optimization [32]. In this competition, L-SHADE's results outperformed other state-of-the-art DE variants and are comparable to the state-of-the-art CMA-ES variants. However, some of the recent publications [33,34] have shown that DE and its variants face significant difficulties on non linearly separable functions and can be outperformed by CMA-ES. The latter method has advantageous convergence properties and performs well with small populations, which makes it even more promising when it comes to improve the computational time. In addition, it has already shown a good performance in some IR problems [21]. Furthermore, several modern CMA-ES variations have yielded better results than the classical CMA-ES in many different problems [28]. Among them, a restart CMA-ES with two interlaced restart strategies (one with an increasing population size and another with varying small population size) called BI-population-CMAES (BIPOP-CMA-ES) [35,36] has showed a very significant behavior outperforming the classic CMA-ES and other modern CMA-ES versions in the BBOB-2009 function testbed [35,36].

Recently, a powerful and versatile RCEA called Coral Reef Optimization with substrate layers (CRO-SL) was proposed in [37]. CRO-SL is inspired on the formation and reproduction of coral reefs. CRO-SL simulates the different phases that corals undergo during their lives, such as reproduction, larval settlement, or fight for a space in the reef. Furthermore, CRO-SL simulates the substrate layers in coral reefs. Substrate layers affect to the growth and development of the coral. These layers are modeled by using

[1] Search performed the 2nd June 2020 using the keywords (TITLE-ABS-KEY "image registration") AND (TITLE-ABS-KEY ("evolutionary algorithm") OR TITLE-ABS-KEY ("genetic algorithm") OR TITLE-ABS-KEY ("evolutionary algorithm") OR TITLE-ABS-KEY ("evolutionary") OR TITLE-ABS-KEY ("metaheuristic") OR TITLE-ABS-KEY ("metaheuristics") OR TITLE-ABS-KEY ("stochastic optimization") OR TITLE-ABS-KEY ("stochastic search") OR TITLE-ABS-KEY ("heuristic search")).

different exploration operators (e.g. DE search, Gaussian mutation, etc.) on different regions of the coral reef. Their simulation combines very different exploration operators within the competitive evolution rules of the coral reefs, providing a competitive grid-based co-evolutionary strategy to CRO-SL in just one population. Lastly, CRO-SL also improves the best solution using a local search (LS) method with a limited number of evaluations, making it become a powerful memetic algorithm [38].

There is a lot of controversy with the proposal of new bio-inspired algorithms [39] and their justification must be based on their actual performance beyond the metaphor. CRO-SL presents a high novelty since it provides an excellent exploration-exploitation trade-off and robustness as results of the combination of all the previous mentioned features, specially for its competitive environment and the incorporation of multiple search patterns. In addition, CRO-SL usually converges quickly to high quality solutions even in multi-modal search spaces, being suitable for computationally expensive optimization problems both satisfying quality and computation time constraints. However, its performance varies significantly depending on the CRO's parameters and the different substrates included in the simulated reef. In particular, CRO-SL has outperformed both classical and state-of-the-art evolutionary IR methods in many 3D–3D medical IR problems [22], making it a really promising RCEA for CR with the only drawback of the complex tuning of its parameters.

The best RCEA for solving computationally expensive optimization problems according to the IEEE CEC competitions is the mean–variance mapping optimization (MVMO) optimizer [40]. MVMO has ranked in top positions in expensive optimization competitions, such as IEEE CEC 2013 [41], 2014 [7], 2015 [8], 2016 [42], and 2018 [43], showing an excellent performance and robustness. MVMO is a novel single-individual RCEA that considers a best solution archive, but its novelty lies within a new mapping function employed for mutating the offspring. This mapping function is based on the mean and variance of the best solution archive. MVMO has been numerically compared to other enhanced RCEAs showing a better performance in many problems, especially in terms of convergence speed. For instance, a powerful variant called MVMO-SH (the "S" refers to the offspring approach based on single parent and multi-parent crossover, and the "H" for the hybridization of MVMO with the use of LS) improves the global search performance of the classical MVMO. MVMO-SH considers a set of solutions (i.e. particles of a swarm) instead of just one, each having its own best solution archive and mapping function, and allows the exchange of information and dynamic reduction of the swarm size.

Thus, a rigorous comparison is needed to determine the influence of the RCEA in the proposed framework to automate the CR problem. Motivated by the analysis of the literature, the RCEAs to be studied in this paper are as follows: (1) DE, the RCEA used in state-of-the-art CR method [14]; (2) L-SHADE, one of the best self-adaptive variants of DE; (3) CMA-ES, a classic RCEA that has outperformed DE in many problems; (4) BIPOP-CMAES, one of the best modern variation of CMA-ES; (5) CRO-SL, a powerful RCEA that is the state-of-the-art method in 3D–3D IR problems but is complex to fine tune; and (6) MVMO-SH, a novel RCEA that has obtained groundbreaking results in many prestigious competitions such as those held within IEEE CEC, especially in costly optimization problems.

3. Comparative radiography basics and related works

CR consists of the comparison of skeletal structures (i.e. bones and cavities) in AM and PM radiographs to determinate the identity of a deceased (see Fig. 1). Depending on the number of available skeletal structures and their uniqueness, CR can be utilized either for positive identification or candidate short listing [44]. In the CR technique, we distinguished three consecutive tasks [14]:

1. Process the PM material (cleaning, stabilizing the skeletal remains, and scanning the "clean" bone with a laser range scanner or performing computed tomography (CT) scan) and ask the corresponding authorities for AM data of all the candidates. Image enhancement and/or segmentation.
2. Produce a PM radiograph that simulates the scope and projection of each of the AM radiographs.
3. Based on the superimpositions achieved, the identification decision is made by comparing the consistencies and inconsistencies in the bone or cavity morphology, together with other elements such as the quality of the AM radiograph, the visibility of bone or cavity, etc. Notice that the use of computers in this stage aims to support the final identification decision that will always be made by the forensic anthropologist.

Traditionally, the analyst manually performs the superimposition through a trial and error process, based only on its skills and experience. Thus, CR is time consuming and its results are subjective, hardly reproducible and suffer from errors related to the analyst's fatigue reducing the applicability of the CR technique. All these factors reduce the utility of the CR method and thus (semi) automatic methods to assist forensic experts in their identification endeavour are required.

The superimposition process can be automatized using a 3D–2D image registration (IR) approach [10,45]. Particularly, only IR methods based on features, such as segmentations of silhouettes, are fitting for CR, since intensities can significantly vary between the AM and PM images [46,47]. IR methods are based on an optimization process that searches for the best match between the silhouette of a skeletal structure in a AM radiograph and a 2D projection of the 3D PM skeletal structure (either obtained via the segmentation of a PM Computed Tomography (CT) or digitized with a 3D scanner). Most 3D–2D IR approaches are designed for a controllable set-up, which is a common situation in many medical domains. They can assume an initial pose nearby the GT and that the parameters associated with perspective distortions are known (i.e. an error of around 20 mm in translation and 20° in rotation in [11], a maximum target registration error [48] of 16 mm in [12], etc.). Feldman et al. [49] proposed a 3D–2D IR method based on silhouettes without any initialization assumptions by using free-form curves and surfaces, but still assuming the perspective distortions' parameters. However, the AM radiograph was taken in an uncontrollable set-up, where pose and radiograph device are unknown, and therefore none of these assumptions are appropriate for CR.

These drawbacks have been overcome by IR methods based on RCEAs, a.k.a. evolutionary IR methods, in several IR problems [15,23,24]. RCEAs are global optimization techniques with a robust performance, that enables them to tackle complex medical IR problems. In particular, in [14] Gómez et al. proposed the first evolutionary 3D–2D IR approach to automatize the superimposition process to compare the silhouette of any bone or cavity without any assumption on the initialization or the main parameters related to the perspective distortions in radiographs (i.e. the source to image distance, a.k.a. SID). The approach makes use of DE [25], a modification of the DICE metric [50] that considers occlusion regions (which are hard to segment regions either because of the fuzzy borders of the bone or occlusions caused by other overlapped structures), and a simple perspective transformation with 7 parameters: 3 translations; 3 rotations; and the SID. This method was tested with frontal sinuses, clavicles and

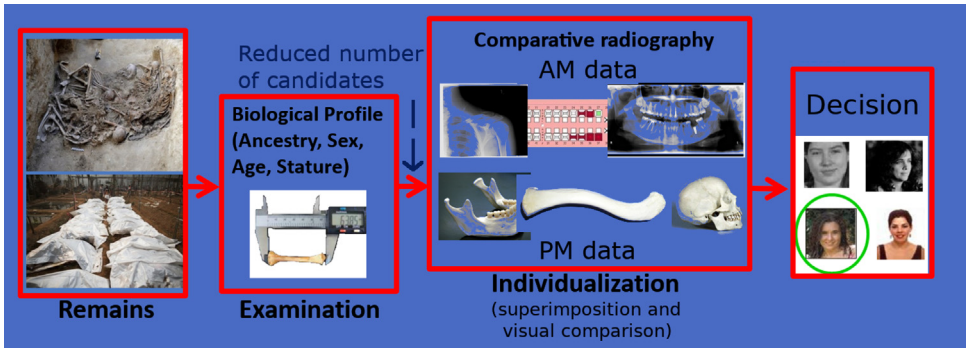

Fig. 1. The usual procedure utilized by forensic experts for CR-based identification is the following: (1) a biological profile (sex, age, stature, etc.) is obtained based on the PM remains of the deceased; (2) the candidates that do not match the biological profile are discarded; (3) all the possible AM records and medical images of the candidates are gathered and pre-processed; (4) the PM remains are superimposed and compared to the AM data through skeletal comparison techniques; and (5) an identification decision is taken based on the results.
Source: Extended version of our previous figure published in [14].

patellae obtaining a promising performance. However, it showed the following drawbacks: (1) the robustness of the DE algorithm, especially with clavicles and patellae, that in some runs leaded to bad superimpositions due to the stochastic nature of DE and the highly multimodal search space tackled (see [14] and Section 4.3 for further details of the landscape analysis); (2) the large amount of time required to obtain a superimposition with DE (1800 s on average). This high run time is motivated by the high computational cost required by each evaluation (on average, it takes 0.25 s to obtain projection of 1290 $\times$ 1050 pixels in a standard computer), uncovering the computationally expensive optimization nature of the CR problem as well as the high number of evaluations required by the optimizer to converge; and (3) none of the projective projections considered reproduced the perspective distortion of radiographs where the x-ray generator was not perpendicular to the image receptor (e.g. in the Water's projection of radiographs of frontal sinuses [9], see Fig. 3).

4. Evolutionary image registration for comparative radiography

The methodology proposed in this paper is based on [14] and depicted in Fig. 2. As said, the method proposed in [14] is the only existing method for the automation of the superimposition of 3D and 2D images for CR. This 3D–2D silhouette-based IR method is composed of the five following components:

1. A 3D image and a 2D image of a skeletal structure to be superimposed. The 3D image is a PM 3D surface model of the skeletal structure acquired using a 3D surface scanner or obtained by segmenting a CT. Meanwhile, the 2D image is the segmented silhouette of the skeletal structure in the AM radiograph together with its occlusion region (i.e. the region where the segmentation expert cannot distinguish among the target structure and other skeletal structures, organs and/or background).
2. A projective transformation [51], which produces 2D projections of the 3D image.
3. The expert knowledge that constrains the projective transformation's parameters. Radiographs are taken with the body in a known position (posteroanterior, anteroposterior, or lateral) following the radiographs acquisition protocols [52] with a certain margin of error (see Table 1).
4. A similarity metric, a.k.a. fitness function, that measures the similarity between the AM silhouette and a 2D projection. Both in the state-of-the-art work for CR [14] and in this work, the similarity metric utilized is the Masked DICE metric [14], which combines the DICE metric [50] with a occlusion region (see Eq. (1)). Masked DICE computes the overlap of the two silhouettes in the whole image except in the occlusion region where the information is not reliable.

$$\text{Masked DICE} = \frac{2 \cdot |(A \setminus M) \cap (B \setminus M)|}{|A \setminus M| + |B \setminus M|} \quad (1)$$

where A and B are sets of pixels of an object silhouette (i.e. the AM segmented skeletal structure and a PM projection of the 3D skeletal structure), and M is the occlusion region (of the AM radiograph).
5. An optimizer, which searches for the best projective transformation in terms of the similarity metric. The optimizer starts with a set of perspective transformations, randomly generated within the ranges of Table 1, that are iteratively improved. Notice that the optimizer cannot rely on an accurate initialization on the perspective transformation parameters since the radiograph was taken in an uncontrollable set-up and not all the forensic experts and practitioners have skills on IR.

The five former components are further detailed in [14]. The main contribution of the current work is in the proposal of a new projective transformation and the analysis of the optimizers (the second and fifth components, respectively) that will be detailed in the following subsections.

4.1. Projective transformation

The projective transformation [51] behind a radiograph image is, in most of the cases, a simple perspective transformation obtained using a pinhole camera model [53]. A simple perspective transformation considers 6 extrinsic parameters (3 translation and 3 rotations) and 1 intrinsic parameter (focal distance; assuming that the rest of intrinsic parameters of a complete perspective transformation are known: the principal point is located in the center of the image, pixels' aspect ratio is square, and no skewness). Particularly, in a radiograph, the focal distance is represented by the source to image receptor distance (SID) [53] (see Fig. 2). SID is also assumed as known in most works [11] since they are designed for a controllable set-up but that is not the case for the CR problem. Although the perspective distortion can be small in many radiographs because of the large distance between the x-ray generator and receptor (as in chest radiographs), its consideration has shown to be crucial in the IR endeavour. This has been shown in [14], where better results were obtained using the perspective transformation than the orthographic transformation, despite the more challenging optimization problem involved.

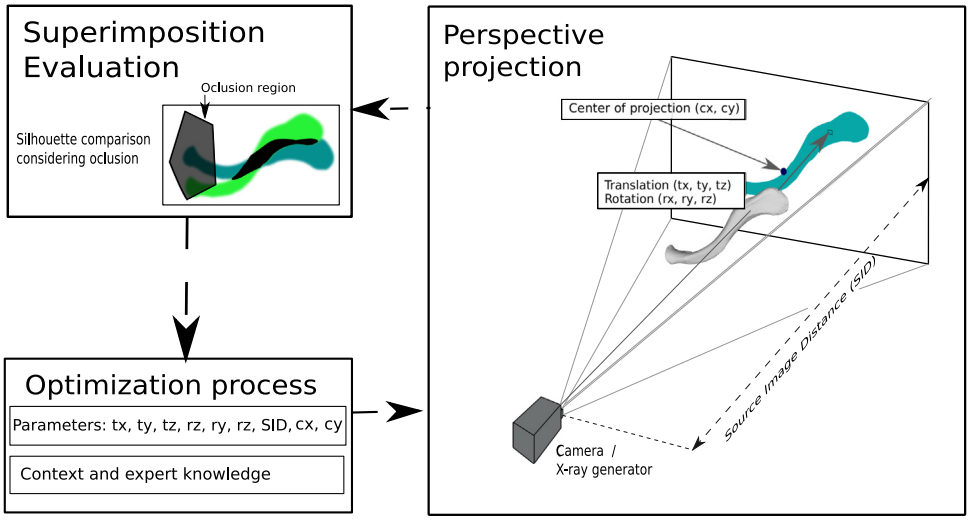

Fig. 2. Scheme of the proposal of 3D–2D IR for CR. Three main interconnected blocks are represented: (Right) the projective transformation to obtain a projection of the 3D model with 9 parameters: translation (t_x, t_y, and t_z), rotation (r_x, r_y, and r_z), and perspective distortions (SID, c_x, and c_y); (Top left) The similarity metrics that compares the PM projection (colored in blue) and the AM segmentation (colored in red) considering an occlusion region (colored in gray); (Bottom left) the optimization process to estimate the 9 parameters of the registration transformation that are only weakly limited by the context and expert knowledge from the x-ray acquisition protocol.
Source: Extended version of our previous figure published in [14].

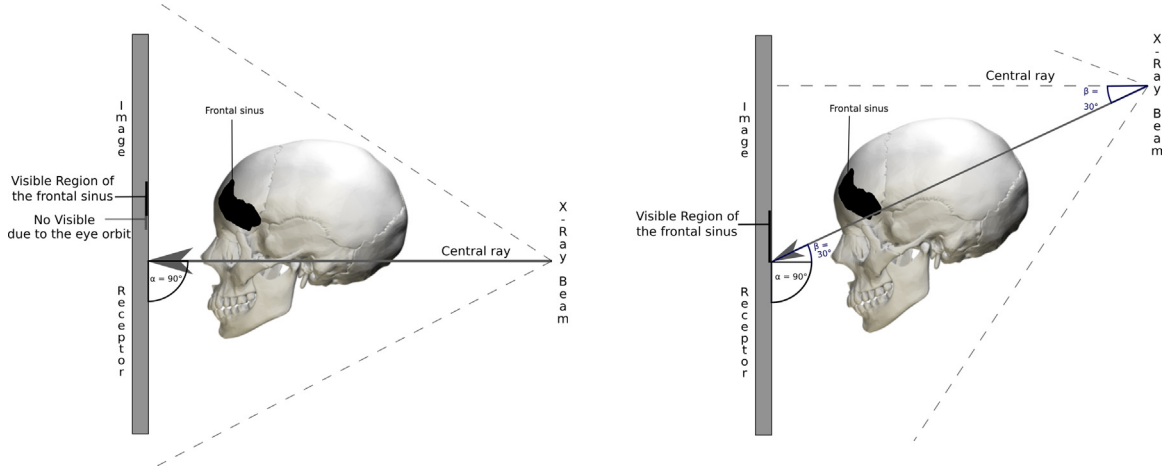

Fig. 3. (Left) Diagram of a frontal sinus radiograph with a posteroanterior view, where the ray between x-ray generator and the center of the image receptor is perpendicular. (Right) Diagram of a frontal sinus radiograph with a Water's view, where the ray between x-ray generator and the center of the image receptor is not perpendicular.

However, radiographs acquired with procedures where the ray that joins the x-ray generator and the center of the image receptor is not perpendicular cannot be modeled with a simple perspective transformation. That is the case of frontal sinuses radiographs taken in one of the acquisition protocols of the Water's view (see Fig. 3 for a graphical example). In these radiographs, the acquisition protocols [52] establish that the x-ray bean is angled at β to the center of the receptor (see Fig. 3). It causes that the principal point of the image is not located at the center of the images (as can be seen in Fig. 3) and can be located even outside the image limits. Thus, to model these radiographs, a more complex perspective transformation, that also models changes in the principal points is needed (resulting in 9 parameters to be optimized). The movement of the principal point in an axis can be calculated according to the following equation:

$$c_i = SID \cdot \frac{sin(90 - \beta_i)}{sin(\beta_i)} \tag{2}$$

where c_i is the principal point displacement in the axis i, and β_i is the angle of the ray that joins the center of the image receptor and x-ray generator in the axis i.

Furthermore, even radiographs taken in conventional views as the posterioanterior can be affected by this distortion (although with a minor effect), due to the small alignment errors between the image receptor and x-ray generator and the modeling of changes in the principal point can also be beneficial for them.

To sum up, two projective transformations are considered in this contribution, aiming to improve the performance of the automatic CR-method: the simple perspective projection with 7 parameters (t_x, t_y, t_z, r_x, r_y, r_z, and SID) from [14] and a new more complex perspective projection with 9 parameters (t_x, t_y, t_z, r_x, r_y, r_z, SID, β_x, and β_y). The two transformations will be referred from now on as P7 and P9, respectively. Their parameters' range is stated in Section 5.1.

4.2. Real-coded evolutionary algorithms for the image registration optimizer

This subsection is devoted to the description of the six RCEAs to be benchmarked on the CR problem in the current contribution. All the RCEAs are designed for the optimization of the real coded parameters of the P7 and P9 transformations, defined in Section 4.1, using the Masked DICE metric, introduced in Section 4, as fitness function. The candidate solutions of the different RCEAs are randomly initialized within the ranges showed in Table 1 since a closer one would be unrealistic as stated in Section 1.

4.2.1. Differential evolution

DE [25] is a variant of an evolution strategy [54]. It begins with a random initialization of a population of n candidate solutions. Afterward, DE searches for better solutions by combining the candidate solutions' parameters using a crossover operator along a limited number of generations. The crossover operator combines the parameters of three random candidate solutions from the previous generation (detailed equations can be reviewed in [25]). Lastly, DE also has an elitism mechanism which maintains the best candidate solution so far into the next generation. In summary, DE has the following parameters: the population size p, the differential weight F, and the crossover probability P_c.

4.2.2. L-SHADE

L-SHADE is a self-adaptive DE approach proposed by Tanabe et al. in 2014 [32] based on a previous adaptive DE optimizer called SHADE [55]. Its main addition was a linear reduction of the population size (which it is initially set to p_{init}) thought the generations. L-SHADE maintains the automatic adjust of the differential weight F and crossover probability P_c parameters in each generation of SHADE. To this end, it keeps a historical memory with H entries for both F and P_c. Furthermore, it also conserves its mutation strategy, to-pbest/1, where the greediness is adjustable using a parameter pb, and the use of an external archive for maintaining diversity, its size equal to p_{init} plus r^{arc}. The goal is to adjust the optimizer behavior during the first generation to promote the search space exploration and subsequently to reinforce its exploitation. To sum up, the parameters to be tuned for L-SHADE are: p_{init}, H, pb, and r^{arc}. Their recommended ranges are reported in [32].

4.2.3. CMA-ES

CMA-ES [26] has been largely considered as the state of the art in RCEAs and has outperformed DE and its variants in many optimization problems, as stated in Section 2. CMA-ES is based on updating the covariance matrix of the multivariate normal distribution along the algorithm's generations to focus the exploration on the most promising regions. Afterward, CMA-ES performs the following two steps in each generation: (1) λ candidate solutions are generated according to the multi-variable normal distribution, the covariance matrix, and the step size σ; and (2) the distribution center and the covariance matrix are updated based on the μ best candidate solutions and σ is updated based on the improvement achieved (detailed equations can be reviewed in [26]).

CMA-ES only requires to set three parameters μ, λ (number of best solutions considered to update the distribution center and number of individuals of the population, respectively), and initial step size σ. Their default value in function of the number of variables n according to the authors is: $\lambda = 4 + \lfloor 3\ln(n) \rfloor$ and $\mu = \lambda/2$. However, some works have shown that a higher value for λ and a modification of the value of μ can lead to make CMA-ES more robust and/or exploitative on multimodal problems [21].

4.2.4. BIPOP-CMA-ES

BIPOP-CMA-ES [35,36] is a restart CMA-ES with two interlaced restart strategies, that modifies the values of the number of candidate solutions λ and the number of best solutions utilized for updating the covariance matrix μ in each restart. The first restart strategy consists of increasing the population size λ by a factor of 2. Meanwhile, the second restart strategy involves decreasing the population size λ based on the previous and the default values of λ (detailed equations can be reviewed in [35]). In both restart strategies, the new value of μ is obtained by halving the new value of λ. Performing the first or second restart strategy depends on which restart strategy's budget value is smaller. Nevertheless, the first and last restarts always utilize the first's strategy. Lastly, the maximum number of restarts that can be performed is nine. To sum up, BIPOP-CMA-ES requires to set the three same parameters than CMA-ES (λ, μ, and the step size σ). The only difference is that the values of λ and μ given to BIPOP-CMA-ES are only their initial values since they are adapted in each restart.

4.2.5. CRO-SL

CRO-SL [22] is based on natural processes occurring in coral reefs. The coral reef R is represented as a bi-dimensional grid of p positions (population size), where each position stands for solutions to the current optimization problem. At the beginning, p_0 positions (given as a percentage of the total population) are randomly initialized with candidate solutions to the problem tackled while the rest are empty, reserved to allow other corals to grow. For each generation, the following stages will be applied to the coral reef sequentially (these stages are further detailed in [22]): (1) Broadcast spawning: it consists of generating new larvae from a pair of candidate solutions using a crossover operator; (2) Brooding: new larva are generated via a mutation mechanism that is applied to a fraction of corals 1 - F_b; (3) Larvae setting: each larvae will try to set in a random position of the coral reef, they will only set if it the location is free or the larvae has a better fitness value than the solution occupying that position; (4) Depredation: a fraction (F_d) of the corals with the worst fitness are removed from the population with very small probability (P_d).

CRO-SL is an extension of the basic algorithm that also simulates the substrate layers in coral reefs. It divides equally the coral reef R into several substrate's layers, and the crossover operator in step 2 will vary depending in which layer the larvae falls. The choice of the operators (or substrate layers) to be used has a significant effect in the optimizer's behavior. In particular, the operators (or substrate layers) considered for IR in [22] are: Harmony search, DE, Gaussian mutation, Cauchy mutation, Simulated Binary Crossover (SBX), and Blend Crossover(BLX)-α. Furthermore, CRO-SL (as stated in Section 2) also has a LS to improve the larvae with the Bound Optimization BY Quadratic Approximation (BOBYQA) optimizer [56] using a maximum of n_{LS} evaluations.

To sum up, the parameters to be tuned for CRO-SL are as follows: reef size p, number of coral reef positions initialized p_0, number of generations g, number of LS evaluations n_{LS}, deprecation fraction F_d, deprecation probability P_d, asexual reproduction proportion F_a, mutation fraction F_b, mutation probability, the set of substrate layers utilized, and the parameters from the operators (e.g. F for DE and δ for harmony search).

4.2.6. MVMO-SH

MVMO-SH [41] begins with a initialization stage where the p particles (candidate solutions) of the swarm are randomly generated. The particles are normalized to the range [0, 1], which is a necessary condition to the latter mutation via mapping function (a key element in MVMO) and are only de-normalized for their

fitness evaluation. Afterward, the following steps are performed for each generation (these are detailed in depth in [41]): (1) LS optimization of the particles with a probability p_{LS}; (2) If a particle founds a better solution in terms of fitness than those in its solution archive, the new solution is added to the particle's solution archive (notice that if the archive has reached its maximum size A_s the worst solution is removed); (3) Particles are sorted and divided into two groups according to their fitness value, the *GP* best ones are classified as "good particles" and the rest as "bad particles" (*GP* is adapted along the process taking values between the 20% and 70% of p). The good particles are modified via a custom single parent crossover operation based on local best [7] and bad particles via a custom multi-parent crossover operation based on a subset of good particles [7]; (4) the particles are mutated using a mapping function. This mapping function is based on the mean and variance of each particle's solution archive and a scaling factor f_s that modulates the function's shape. The scaling factor usually begins with a small value f_{start} and progressively increases until reaching its maximum value f_{end} to progressively increase the algorithm's accuracy.

To sum up, the parameters to be tuned are: number of particles p (the recommended value is $15 * number\_variables$; if the number of particles chosen is equal to 1, MVMO-SH will perform as the standard MVMO), LS probability p_{LS}, archive size A_s, scaling factor start (f_{start}) and end values (f_{end}), initial value of the shape of all the variables at the beginning of the optimization d_r (values around 1–5 are suitable to guarantee good initial performance. In practice, it is usually set to 1), and parent selection method (random, neighbor group selection in single step or block steps, or sequential selection of the first variable and the rest randomly).

4.3. Problem landscape complexity

Apart from the high computation requirements of CR, the numerical optimization problem underlying the superimposition process is complex. This can be confirmed by studying the fitness' landscape of the CR problem by using the fitness-distance correlation [57] (see Eq. (3) for the distance function) with synthetic data of a clavicle, a patella, and a frontal sinus (see Section 5.1).

$$\text{Dist} = \frac{\sum_{i=1}^{n} \left| \frac{t_i - min_i}{max_i - min_i} - \frac{GT_i - min_i}{max_i - min_i} \right|}{n} \tag{3}$$

where n is the number of parameters, t_i is the ith parameter of a registration transformation t, GT_i is the ith parameter of the GT transformation GT, min_i is the minimum possible value of the ith parameter, and max_i is the maximum possible value of the ith parameter.

However, the fitness landscape varies significantly with each CR identification problem, due to factors such as the singularity and discriminatory power of each skeletal structure, the segmentation of the AM and PM images, the occlusions present within these images, the projective transformation utilized, etc. Nevertheless, the complexity of the CR problem can be uncovered by studying its simplest scenario, i.e. synthetic data without occlusions or segmentation errors. To analyze the simplest optimization scenario, a sample of 200,000 random transformations near to the GT transformation have been generated and evaluated for each skeletal structure (clavicles, patellae and frontal sinuses) and perspective transformation (P7 and P9), as shown in Fig. 4. Fig. 4 shows many bad superimpositions with a small distance to the GT transformation, as well as good superimpositions with a big distance to the GT transformation. It hints the multimodality of the search space. Furthermore, the fitness distance correlation according to the Pearson's correlation coefficient [58] is 0.47 for P7 and 0.42 for P9, both weakly correlated confirming the complexity and multimodality of the search space even in its simplest scenario.

5. Experiments

The experimental study is divided into three parts. The first experiment is devoted to fine tune the different RCEAs to find their best configuration in terms of accuracy and robustness. For this experiment, only simulated CR problems (positive cases, i.e. the AM and PM data belong to the same person) of frontal sinuses are considered, since these are of great forensic interest and result in the most complex optimization scenario (as it has to model both posterioanterior and Water's views). Furthermore, it is computationally unaffordable (because of its high computational cost) to perform this experimentation also with clavicles and patellae. Meanwhile, the second experiment is devoted to compare the best configuration of each RCEA with simulated CR problems of frontal sinuses, clavicles, and patellae with P7 and P9 in order to find the best RCEA in terms of accuracy and robustness. Finally, the third experiment is devoted to studying the identification capability of the proposed IR framework using P9 and the best resulting RCEA, in turn MVMO-SH, in real images of frontal sinuses.

The same stop criteria is established for all the RCEAs to allow a fair comparison in terms of computational resources. The optimization process ends when at least one of the following three conditions holds: (1) the maximum number of evaluations is reached. This value is set to 50,000 evaluations (it includes the evaluations performed by the LS methods); (2) the optimization process has got stuck. It is considered that the optimization process has stagnated when it has performed 10,000 evaluations without improving the fitness of the best solution; and (3) the optimization process has achieved a good solution/superimposition. A solution is considered of good quality when it shows an error lower than 0.001 in terms of fitness.

All the experiments (I, II and III) have been performed on the high performance computing server Alhambra from the University of Granada composed of 1808 cores Fujitsu PRIMERGY CX250/ RX350/RX500 nodes running Red Hat Enterprise 6.4, although on average only 50 cores were available for this experimentation. Furthermore, several preliminary experiments were performed in the supercomputing center of Galicia (CESGA). It is important to remark the large computational cost of the experimentation following a rigorous experimental design of a computationally expensive optimization problem as CR. Overall, around 1314 computation hours (55 days) were required to perform Experiments I, II and III when the 50 cores were available uninterruptedly. Notice that the reported computational time is for the entire experimentation and it would require a significantly smaller computational time for its use in real forensic scenarios. For instance, the comparison of a 3D surface model against a radiograph only requires 1000 s (0.27 computational hours in one core) and the comparison of a 3D module against a set of 50 radiographs of possible candidates requires 50,000 s (14 computational hours in one core). Nevertheless, each superimposition can easily run in parallel reducing significantly the required time for all the comparisons.

5.1. Simulated data set

The dataset employed in Experiments I and II is formed by 900 simulated CR problems, 300[2] for each skeletal structure to

[2] According to the sample estimation equation presented in [59] ($n = (4 \cdot Z^2 \cdot \sigma^2)/(W^2)$, where n is the required sample, z the z-score value, σ the expected standard deviation, and W the margin of error), a sample size of 300 is sufficient to guarantee an error lower than ± 0.03 ($W = 0.06$) in the mean with a confidence level of 99% (z-score $Z = 2.4$) and expected standard deviation of 0.2. The value of σ was estimated in a preliminary extermination and, latter, confirmed in our experimental study, see Table 4. Furthermore, it gives a good trade-off between the computational cost of the experimentation and the number of CR scenarios captured in the sample.

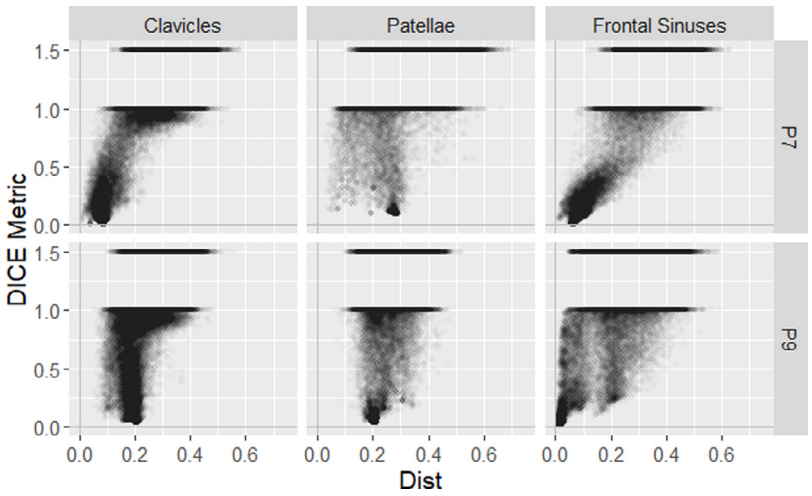

Fig. 4. Scatter plots of DICE metric of a transformation versus its distance to the ground truth transformation according to bone/cavity, and projective transformation.

be studied. Each of simulated CR problems is composed of a 3D surface model and a random 2D perspective projection of the 3D model with occlusions using the projective transformations P7 and P9. The random parameters, that generated the 2D perspective projection of a simulated CR problem, composed the **GT transformation**, i.e. the transformation parameters that our method aims to find. Furthermore, we have also generated a second 2D perspective projection using the GT transformation but without considering occlusions. The GT transformation and the GT projection without occlusions allow us to objectively measure the quality of the superimpositions archived.

In [14], the optimizers showed a different behavior depending on the singularity of the skeletal structure and occlusions, among other factors. Thus, we have composed a simulated dataset that captures the diversity of the CR problem to guarantee that the evolutionary image registration method performs properly in as many different scenarios as possible. We utilized 300 problems per skeletal structure since this value gives a good trade-off between diversity and computational cost of the experimentation.

The dataset has been generated using 30 3D surface models (10 of each skeletal structure studied in this work, i.e. 10 frontal sinuses, 10 clavicles, 10 patellae) obtained as in [14]. Particularly, frontal sinueses' models were obtained by manually segmenting CTs (provided by the Hospital de Castilla la Mancha, Spain) using 3D Slicer 4.5.0-1 [60]. Meanwhile, clavicles and patellae' models were obtained by scanning bones (from the bone collection of the Physical Anthropology Lab at the University of Granada) using a laser range scanner (Artec *Spider*^{TM} 3D scanner). All these 3D models were placed in their respective most frequent positions in a radiograph [52] (a frontal position for frontal sinus and clavicle's models, and a lateral one for patella's models). For each 3D surface model, 10 perspective projections (5 with P7 and 5 with P9) were randomly generated within the ranges showed in Table 1 (these ranges have been set based on international acquisition protocols [52] and are detailed in [14] with the exception of the new parameters β_x and β_y. Notice that these parameters are set to 0 with the P7 transformation). The parameters β_x and β_y have been added to model small alignment errors in the posterioranterior view for clavicles and patellae, and model posterioranterior and Water's views for frontal sinuses (as stated in Section 4). With frontal sinuses, the parameter β_y has a larger range to allow the optimizer to adapt automatically to both posterioranterior and Water's views. In addition, the rotation range has been increased to $[-40, 40]$ to study the robustness of the RCEA to a greater uncertainty on the initial pose of the 3D model. These projections are generated with a resolution of 2 pixels per mm, resulting in images of 480×600 pixels for frontal sinuses and patellae, and 860×700 pixels for clavicles. Lastly, in order to model the occlusions present in real radiographs, two additional projections were generated with occlusion on the skeletal structure of 20% and 40% for each of the previous projective projections. The occlusion ranges are greater than in [14] to test the RCEAs in a more complex optimization scenario.

5.2. Real dataset

The dataset employed in Experiment III was provided by the *Hospital de Castilla-La Mancha*, Spain, and is composed of 50 CTs and 50 radiographs where the frontal sinuses are visible. The data were segmented by two forensic anthropology MSc students from the Physical Anthropology lab (PAL) of the University of Granada. All CTs were segmented by the forensic student *A* (Andrea Cerezo Vallecillo), and all radiographs were segmented by forensic students *B* (José Manuel Pérez Jiménez).

5.3. Performance metrics

Two GT metrics are employed to objectively measure the quality of the superimpositions archived by RCEAs: GT DICE [50] and the mean reprojection distance error (mRPD) [48]. The GT DICE metric measures the overlap between the GT projection's silhouette (equal to the simulated AM projection but without any occlusion) and the 2D projection's silhouette archived by the RCEA. However, the GT DICE metric and the fitness function (i.e. Masked DICE, see Section 4) are highly correlated (e.g. they are equal in cases without occlusions) and thus, to avoid any possible bias, the mRPD metric is also employed. mRPD is an standardized metric for the evaluation of 3D–2D IR methods by computing the retroprojection error between the transformation obtained by the RCEA and the GT transformation (see [14] for further details of the utilization of mRPD in the CR problem). Notice that these metrics can be employed only in simulated CR problems since in real CR problems the GT projection and the GT transformation are unknown.

5.4. Experiment I: Fine tuning of the evolutionary algorithms for the CR problem

5.4.1. Experimental set up

This experimentation involves the application of six different RCEAs (DE, L-SHADE, CMA-ES, BIPOP-CMA-ES, CRO-SL, and MVMO-SH) and two kinds of projective transformations (P7 and

Table 1
Range of the parameters of the projective transformations for each skeletal structure according to international acquisition protocols [52] and expert knowledge.

Parameter	Bone/Cavity		
	Frontal sinuses	Patellae	Clavicles
Image receptor dimension (mm)	240 × 300	240 × 300	430 × 350
t_x (mm)	[−125, 125]	[−125, 125]	[−210, 210]
t_y (mm)	[−150, 150]	[−150, 150]	[−175, 175]
t_z (mm)	[900 − 200, 900 + 200]	[900 − 200, 900 + 200]	[900 − 200, 1700 + 200]
r_x, r_y, and r_z (degrees)	[−40°, 40°]	[−40°, 40°]	[−40°, 40°]
SID (mm)	[1000 − 100, 1000 + 100]	[1000 − 100, 1000 + 100]	[1800 − 100, 1800 + 100]
β_x (degrees)	[−10°, 10°]	[−10°, 10°]	[−10°, 10°]
β_y (degrees)	[−50°, 10°]	[−10°, 10°]	[−10°, 10°]

Table 2
Summary of all the parameters of the different RCEAs and their studied values in Experimentation I.

Fixed parameters				
General par.	Number of evaluations: 50,000			
DE	$p = 100$	$F = 0.5$	$P_c = 0.5$	
L-SHADE	$r^{arc} = 2$[a]	None		
CMA-ES	None			
BIPOP-CMA-ES	None			
CRO-SL	$p_0 = 0.4$[b]	$n_{LS} = 50$[c]	$F = 0.5$	
	Substrates = (Harmony search, DE, Cauchy Mutation[d], SBX, and BLX-α)			
MVMO-SH	f_{start} =1	$d_r = 1$	$GP = 5$	$p_{LS} = 0.015$
	Parent selection strategy = sequential selection of the 1st variable, and the rest randomly[e]			
Parameters to fine tune				N° conf.
DE	None. DE's parameters were already fine tuned in [14])			1
L-SHADE	$p_{init} = (15, 20, 25)$	$pb = (0.05, 0.1, 0.15)$	$H = (2, 5, 10)$	27
CMA-ES	$\lambda\&\mu$ = (100 & 15, 40 & 15, d[f] & d[g]).		$\sigma = (0.01, 0.1, 0.3)$	9
BIPOP-CMA-ES	$\lambda\&\mu$ = (100 & 15, 40 & 15, d[f] & d[g]).		$\sigma = (0.01, 0.1, 0.3)$	9
CRO-SL	$p = (25, 50, 100)$	$\delta = (0.1, 0.25, 0.4)$		9
MVMO-SH	$p = (1, 25,$ d[h])	$A_s = (5, 10, 25)$	$f_{end} = (1.5, 2.5)$	12

[a]Other values (1, 3) were also studied in a preliminary experimentation with worst performance results.
[b]Other values (0.15, 0.65) were also studied in a preliminary experimentation with worst performance results.
[c]Other values (0, 100) were also studied in a preliminary experimentation with worst performance results.
[d]The Gaussian Mut. was studied as alternative to the Cauchy Mut. in preliminary experiments with worst results.
[e]The rest of selection strategies were tested in preliminary experiments with worst performance results.
[f]d = default value calculated according to the following equation: $\lambda = 4 + \lfloor 3\ln(n) \rfloor$. Thus, it is equal to 9 and 10 for P7 and P9, respectively.
[g]d = default value calculated according to the following equation: $\mu = \lambda/2$. Thus, it is equal to 4 and 5 for P7 and P9, respectively.
[h]d = default value calculated according to the following equation: 15*number_variables. Thus, it is equal to 105 and 135 for P7 and P9, respectively.

P9) for each of the 300 CR cases of frontal sinuses to achieve our goal of determining the influence of the evolutionary optimizer used by the automatic CR method. As mentioned above, this experiment is meant to fine tune the six RCEAs. While there are unsupervised methods for parameter tuning [61], they tend to evaluate a very large number of parameter configurations, making them infeasible for an expensive optimization problem as CR (since each configuration should be tested over the 300 CR problems to compare them rigorously). Therefore, we have utilized a grid search where the parameter values are chosen based on the recommendations present on the RCEA's original paper and on expert knowledge about its behavior. Taking these considerations into account, the parameter grid shown in Table 2 was designed. Lastly, in order to allow for a fair comparison, every RCEA will have the same computational resources with maximum number of 50,000 evaluations.

In summary, a total of 67 parameter configurations were considered, resulting in 20,100 executions. For each of these executions, 10 independent runs were performed to study the robustness of the RCEAs for solving the CR problem due to their stochastic component. Thus, 201,000 runs (i.e. superimpositions) were performed. Each superposition takes 1000 s on average, resulting in an experimentation of around 55,833 computation hours (2326 computation days) that performed on the 50 available cores of Alhambra required "only" around 1100 computation hours (45 computation days).

5.4.2. Results

Fig. 5 shows the results obtained by the different RCEAs and their configurations according to the GT DICE metric. The performance varies significantly depending on the RCEA and projective transformation in terms of mean and standard deviation values. Better results are always obtained with P7 proving that P9 is significantly more complex as stated in Section 3, which is confirmed by the Wilcoxon's test [62] obtaining a p-value lower than $1 \cdot 10^{-15}$ with both metrics. CMA-ES is an exception obtaining better results with P9 but its results are still significantly worse than those provided by the other RCEAs with both P7 and P9. Nevertheless, P9 holds a greater forensic interest since it allows to model radiographical scenarios that P7 cannot model.

Studying the influence of the different parameters, it can be observed large differences for each RCEA, especially with respect

Table 3
Summary of the results according to projective transformation, skeletal structure type, and RCEA optimizer (Experiment II).

Bone	Opt.	Proj. Tran.	Masked DICE		GT DICE		mRPD	
			Mean	Sd	Mean	Sd	Mean	Sd
Frontal Sinus	CMA-ES	P7	0.414	0.140	0.446	0.130	8.736	4.767
		P9	0.272	0.067	0.307	0.078	46.306	36.856
	BIPOP-CMA-ES	P7	0.011	0.054	0.015	0.061	0.595	1.998
		P9	0.016	0.044	0.029	0.070	15.453	28.762
	CRO-SL	P7	0.073	0.069	0.111	0.100	2.458	2.930
		P9	0.198	0.075	0.249	0.084	44.723	35.268
	DE	P7	0.008	0.034	0.015	0.048	0.307	1.396
		P9	0.048	0.040	0.076	0.066	29.024	31.458
	L-SHADE	P7	0.079	0.085	0.113	0.110	2.553	3.213
		P9	0.147	0.071	0.202	0.091	49.439	31.324
	MVMO-SH	P7	**0.001**	**0.009**	**0.002**	**0.009**	**0.047**	**0.369**
		P9	**0.011**	**0.020**	**0.021**	**0.042**	**14.778**	**29.968**
Clavicle	CMA-ES	P7	0.542	0.130	0.564	0.140	22.695	16.785
		P9	0.519	0.139	0.537	0.149	32.044	17.304
	BIPOP-CMA-ES	P7	0.089	0.186	0.109	0.220	7.063	17.111
		P9	0.132	0.220	0.155	0.246	30.573	29.887
	CRO-SL	P7	0.107	0.134	0.149	0.178	10.092	18.339
		P9	0.133	0.122	0.176	0.153	27.121	16.946
	DE	P7	0.005	0.021	0.010	0.036	0.461	3.116
		P9	0.028	0.053	0.046	0.077	23.024	15.253
	L-SHADE	P7	0.105	0.142	0.129	0.159	7.396	15.648
		P9	0.111	0.149	0.136	0.164	33.862	20.716
	MVMO-SH	P7	**0.001**	**0.000**	**0.002**	**0.002**	**0.065**	**0.051**
		P9	**0.004**	**0.004**	**0.009**	**0.009**	**19.383**	**14.008**
Patella	CMA-ES	P7	0.273	0.117	0.330	0.116	15.318	15.878
		P9	0.268	0.136	0.326	0.122	22.001	15.865
	BIPOP-CMA-ES	P7	0.016	0.024	0.045	0.063	9.486	19.605
		P9	0.019	0.028	0.053	0.072	22.163	25.054
	CRO-SL	P7	0.043	0.033	0.096	0.070	12.395	19.434
		P9	0.080	0.054	0.152	0.092	22.558	19.350
	DE	P7	0.014	0.022	0.045	0.057	7.057	16.320
		P9	0.025	0.026	0.073	0.073	21.411	20.970
	L-SHADE	P7	0.096	0.083	0.143	0.089	14.228	23.969
		P9	0.146	0.144	0.184	0.134	28.048	23.771
	MVMO-SH	P7	**0.003**	**0.010**	**0.009**	**0.026**	**2.650**	**13.130**
		P9	**0.006**	**0.011**	**0.026**	**0.044**	**17.151**	**18.216**

to their sensibility to the parameter choice. L-SHADE presents the more robust behavior since the results are similar for the different parameter values for each one of the problems. On the contrary, CMA-ES gives very different results in P7 depending on the parameter values used (in P9 there are very similar). More in detail, the most influential parameter in CMA-ES seems to be sigma, σ, obtaining better results with higher σ values. In BIPOP-CMA-ES this tendency is increased, corroborating that σ parameter is clearly more influential in both problems. By setting an appropriate σ value, BIPOP-CMA-ES obtains for both problems better results than the majority of the remaining RCEAs but DE and MVMO-SH. MVMO-SH is very sensitive to the number of particles, p. In P7, results are clearly different with p = 1 and p = 25, obtaining two performance levels based on that parameter value. For P9, results show three very different performance levels, for p = 1, p = 25, and p = d. In both problems, the results provided by MVMO-SH with p = 1 are worse than the other RCEAs but with p = 25 it outperforms the majority of algorithms, and with p = d, MVMO-SH achieves the best results overall.

In terms of accuracy and robustness of the best configuration of each RCEA, the worst RCEA (i.e. the sixth position) is CMA-ES (best configuration: $\lambda = 100$, $\mu = 25$, and $\sigma = 0.3$). It is followed by L-SHADE ($p_{init} = 25$, $pb = 0.15$, $H = 2$, and $r^{arc} = 2$) and CRO-SL ($p = 100$, and $\delta = 0.25$) in the fifth and fourth positions, respectively, closely tied. Neither CMA-ES, L-SHADE nor CRO-SL can obtain better results than DE, the state-of-the-art RCEA for CR [14], either with P7 and P9. BIPOP-CMA-ES ($\lambda = 100$, $\mu = 25$, and $\sigma = 0.3$) and DE are also closely tied (taking the third and second positions). Finally, the best RCEA in terms of average and standard deviation values, and confirmed by the Wilcoxon's test with p-values lower than $1 \cdot 10^{-7}$ in the comparison with all the other RCEAs, is MVMO-SH ($p =$ d, $A_s = 4$, and $F_{end} = 2.5$).

MVMO-SH has greatly improved the state-of-the-art results both in terms of accuracy and robustness with P7 [14]. MVMO-SH has also successfully solved a more complex version of the CR problem based on the projective transformation P9, that allows to model both posterioanterior and Water's views, as well as, being robust to occlusions up to the 40% of their silhouettes and rotation ranges of up to 80° ([−40°, 40°]) in the three axis.

5.5. Experiment II: Comparison of the RCEAs over all the CR problems

5.5.1. Experimental set up

This experimentation involves the application of the best configuration of the six different RCEAs (DE, L-SHADE, CMA-ES, BIPOP-CMA-ES, CRO-SL, and MVMO-SH) from Experiment I to all the 900 CR cases (300 frontal sinuses, 300 clavicles, and 300 patellae) using the two kinds of projective transformations (P7 and P9). The best configuration of the parameters in Table 2 are as follows:

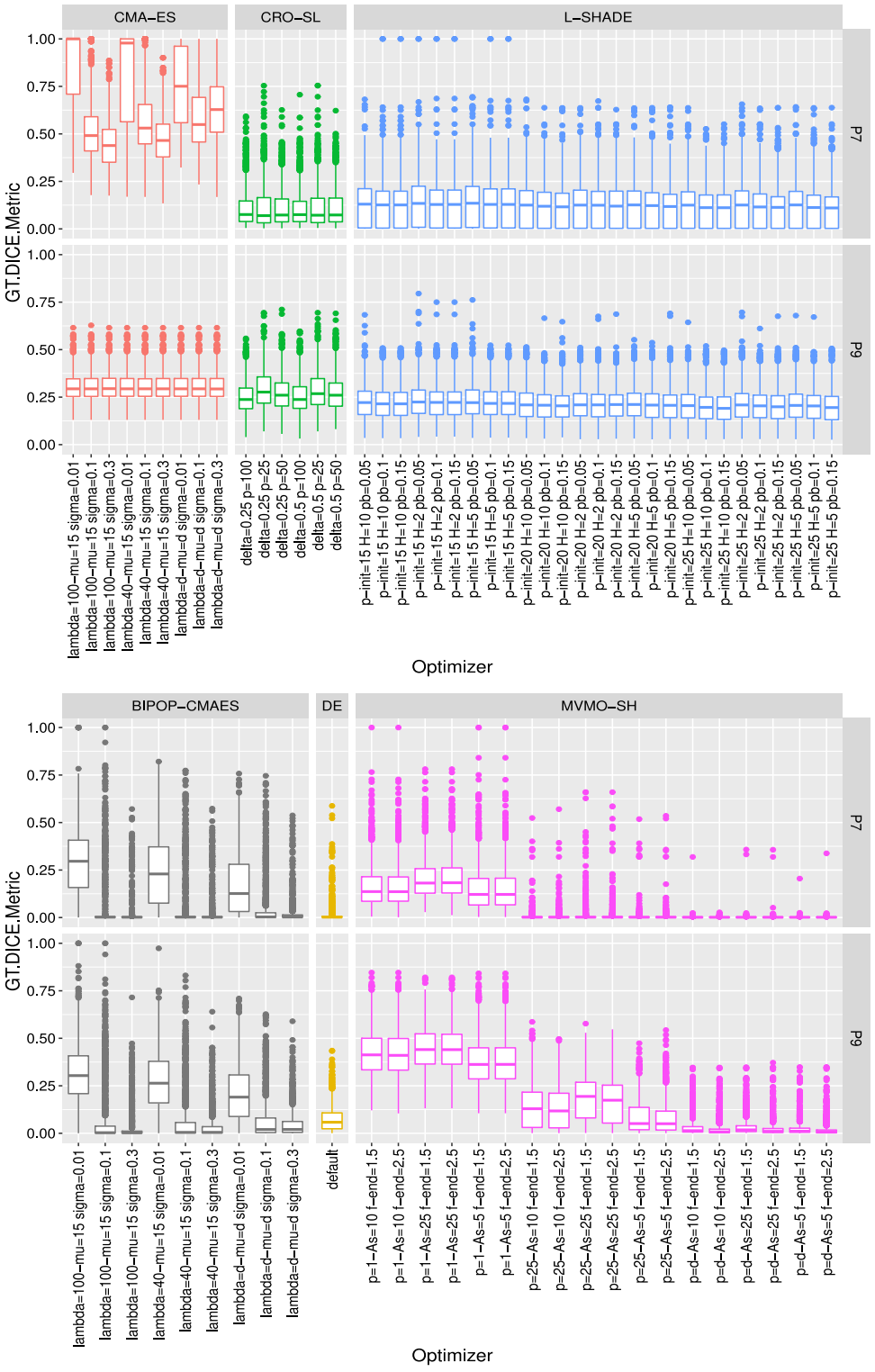

Fig. 5. Boxplot of the minimum errors according to projective transformation, and optimizer for the GT DICE metric (Experiment I). Notice that good superimpositions obtain a Masked DICE value close to 0, while bad ones are close to 1.

- DE: $p = 100$, $F = 0.5$, and $P_c = 0.5$ (fine tuned in [14]).
- L-SHADE: $p_{init} = 25$, $pb = 0.15$, $H = 2$, and $r^{arc} = 2$.
- CMA-ES: $\lambda = 100$, $\mu = 25$, and $\sigma = 0.3$.
- BIPOP-CMA-ES: $\lambda = 100$, $\mu = 25$, and $\sigma = 0.3$.
- CRO-SL: $p = 100$, and $\delta = 0.25$.
- MVMO-SH: $p = \mathrm{d}$, $A_s = 4$, and $F_{end} = 2.5$.

In summary, the six RCEAs are applied to the 900 CR cases resulting in 3000 executions. As in the first experiment, 10 independent runs are performed to avoid any possible bias caused by the stochastic component of the RCEA, resulting in 30,000 runs/superimpositions and around 200 computation hours (8 days) when performed using the 50 cores.

Table 4
Summary of the Masked DICE metric results according to projective transformation, and RCEA optimizer at 5000, 10,000 and 50,000 evaluations (Experiment II).

Opt.	N. Ev.	P7		P9	
		Mean	Sd	Mean	Sd
CMA-ES	5000	0.429	0.168	0.418	0.163
	10,000	0.422	0.169	0.391	0.167
	50,000	0.410	0.169	0.351	0.169
BIPOP-CMA-ES	5000	0.174	0.180	0.221	0.200
	10,000	0.079	0.159	0.109	0.186
	50,000	0.053	0.137	0.075	0.167
CRO-SL	5000	**0.087**	0.098	0.168	0.120
	10,000	0.078	0.094	0.146	0.109
	50,000	0.072	0.092	0.134	0.102
DE	5000	0.096	**0.061**	0.152	**0.080**
	10,000	**0.036**	**0.036**	**0.078**	**0.052**
	50,000	0.009	0.027	0.034	0.042
L-SHADE	5000	0.094	0.109	**0.142**	0.130
	10,000	0.093	0.108	0.135	0.128
	50,000	0.093	0.108	0.135	0.128
MVMO-SH	5000	0.241	0.138	0.338	0.171
	10,000	0.157	0.096	0.269	0.145
	50,000	**0.001**	**0.007**	**0.006**	**0.013**

5.5.2. Results

Table 3 shows the results obtained by the different RCEAs according to Masked DICE, GT DICE, and mRPD metrics. In view of those results, the impact of the considered skeletal structure on the RCEA's performance depicted in [14] has been reduced but not eliminated. When a P7 transformation is considered, the best results are still obtained with frontal sinuses, followed by clavicles and patellae. This is probably due to the frontal sinus' silhouettes are more singular than those from clavicles and patellae. In fact, frontal sinuses are usually employed for identification [18], while clavicles and patellae are mainly employed for short listing [16,17]. However, when P9 is considered, clavicles achieve the best results since the optimization problem to solve with frontal sinuses is more complex (notice that β_y has a range of 50° compared with the 20° of clavicles and patellae). Nevertheless, frontal sinuses are able to obtain significant results with a mean error of 0.02 (i.e. an error of only the 2% of the pixels of the silhouette) and 14 mm according to GT DICE and mRPD metrics, respectively. They also show a low standard deviation of 0.009 and 29 mm for GT DICE and mRPD metrics, respectively. As in P7, patellae had the last position due to their lower singularity.

In this experiment, MVMO is again the best RCEA for CR in terms of average and standard deviation values, as confirmed by the Wilcoxon's test [62] obtaining a p-value equal or lower than $2 \cdot 10^{-16}$ in the comparison with the other RCEAs considering the two metrics and the three bones. The rest of the optimizers are ranked as follows: the second best is DE, the state-of-the-art optimizer for CR; the third best is BIPOP-CMA-ES, that also outperforms to DE in some particular scenarios (e.g. with patellae and P9); followed by L-SHADE and CRO-SL with no significant differences between them (-value of 0.166 according to the Wilcoxon's test [62]); and the worst results are obtained by CMA-ES.

Table 4 shows the mean and standard deviation, according to the Masked DICE metric, of each RCEA and projective transformation after 5,000, 10,000 and 50,000 evaluations. Meanwhile, Fig. 6 reports the average time required by the RCEAs to reach a stop condition and the average results obtained according to the GT DICE metric. In view of the results collected in Table 4, the convergence speed of MVMO-SH is lower than that of the other RCEAs, needing almost all the 50,000 evaluations to obtain significant results in terms of accuracy and robustness. On the contrary, the other RCEAs have a similar performance with 10,000 and 50,000 evaluations, and the only one showing acceptable results with only 10,000 evaluations is DE. However, after the 50,000 evaluations limit, the best RCEA in terms of time is also MVMO-SH (as can also be seen in Fig. 6). In that figure, we can also observe that every algorithm but CMA-ES, and sometimes DE, does do not stop due to the maximum number of evaluations condition, but to the premature convergence (worse case) or good superimposition (best case) stop conditions. CRO-SL and L-SHADE stops more than 90% of times for premature convergence, while BIPOP-CMA-ES stops for good superimposition more than half of times. The main stopping condition of MVMO-SH is the good superposition (especially in problem P7, in which almost all runs stop for that good condition). The most frequent stopping condition reached by MVMO-SH is the good superimposition in 92% of all executions, while the converged condition arises in 7%, and the maximum number of evaluation condition only in 1% of runs (see Fig. 7). Thus, MVMO-SH has obtained an improvement in accuracy, robustness, and convergence, as well as run time (see Table 5), in the solution of the CR problem. In general, every RCEA (but CMA-ES for P7 and P9 and DE for P9) is not limited by the maximum number of evaluations and thus no further improvements are to be expected with further run times.

5.6. Experiment III: Testing the identification capability of our 3D-2D IR-based CR framework with frontal sinuses

5.6.1. Experimental set up

This experimentation is aimed to evaluate the identification capability of the proposed 3D–2D IR-based CR framework using frontal sinuses and the best RCEA configuration (MVMO-SH with $p = \mathrm{d}, A_s = 4, F_{end} = 2.5$, and P9). To this end, we confront 50 manually segmented radiographs against 50 manually segmented CTs, resulting in a total of 2500 CR problems (50 positive and 2450 negative cases). Since previous experiments have already shown the robustness of MVMO-SH, and due to the large computational cost of employing again 10 repetitions, a single run is performed. Each of the 2500 runs takes on average 1000 s, resulting in 695 h of computation (or 29 computation days) that, performed on the 50 available cores of computing server Alhambra, required only around 14 computation hours.

5.6.2. Results

High quality superimpositions have been obtained without any initial pose (see Fig. 8) and regardless if the identification case is positive or negative, as desired. Notice that, the IR algorithm must provide the best possible superimposition of the two skeletal structures. The final identification decision is taken later over that superimposition by comparing the consistency in the bone or cavity morphology. In the reliability study, positive and negative cases have shown important differences in terms of fitness according to the Masked DICE Metric (see Fig. 9). However, this metric alone is not sufficient to precisely distinguish between positive and negative cases.

Therefore, the results are reported using CMC curves to study the identification capabilities of the proposal, as done in [14]. A CMC curve measures the probability that the correct match for a identification case is present in a candidate list of the r best matches, where r denotes the position in the rank. For example, rank 5 identification accuracy denotes the probability that the correct match is one of the subjects in a list of the top 5 matches. The results of the reliability study are significant (see Fig. 10). The positive case ranks in the first position in 88% of the cross-comparisons (out of 50 candidates, 2% of the total sample), and a confidence level of 100% of success is reached when the first six

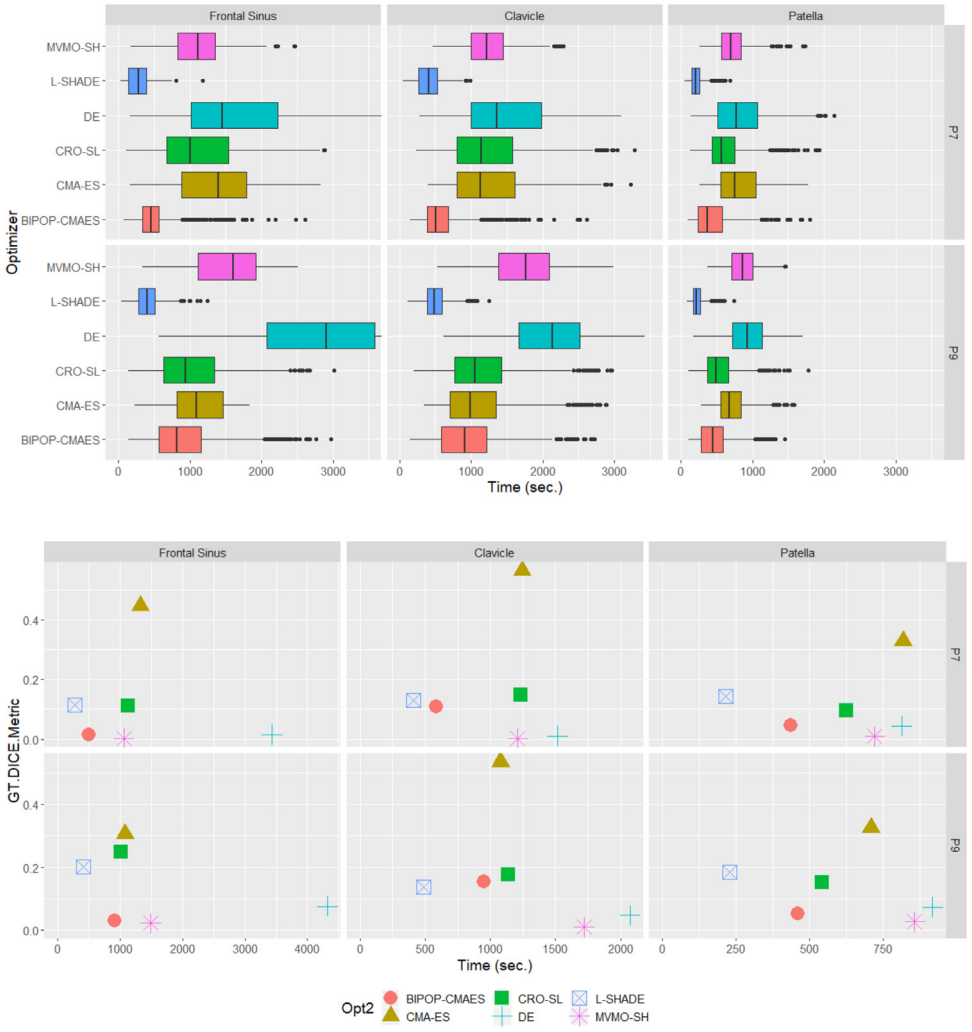

Fig. 6. (Top) Boxplots of the time required to perform a superimposition according to projective transformation and RCEA optimizer. (Bottom) Relation between the average time (seconds) and the GT DICE metric according to projective transformation and RCEA optimizer (Experiment II).

Table 5
Summary of the computational time, in seconds, according to skeletal structure, projective transformation, and RCEA optimizer (Experiment II).

Opt.	Proj. Tran.	Frontal sinuses		Clavicles		Patellae	
		Mean	Sd	Mean	Sd	Mean	Sd
BIPOP-CMAES	P7	497	285	586	311	435	260
BIPOP-CMAES	P9	912	478	951	450	460	221
CMA-ES	P7	1321	592	1243	565	818	333
CMA-ES	P9	1080	439	1080	500	709	217
CRO-SL	P7	1127	598	1232	578	624	282
CRO-SL	P9	1014	513	1135	506	542	232
DE	P7	1665	913	1515	652	814	369
DE	P9	2730	1035	2073	586	919	306
L-SHADE	P7	278	149	408	176	214	90
L-SHADE	P9	405	172	489	166	228	80
MVMO-SH	P7	1056	396	1211	354	722	234
MVMO-SH	P9	1487	548	1720	498	858	232

positions are considered. Notice that these results are obtained with the simplest decision-making method based only on the registration error. Therefore, a more complex decision-making method based on multiple forensic criteria and metrics, as the one depicted in [63] for the craniofacial identification technique, can further improve the identification capabilities of the proposal.

6. Conclusions and future developments

6.1. Regarding the 3D-2D comparative radiography problem

In this work, we have tackled the superimposition problem within the CR task using an evolutionary 3D–2D IR approach based on the silhouette of the skeletal structure without any

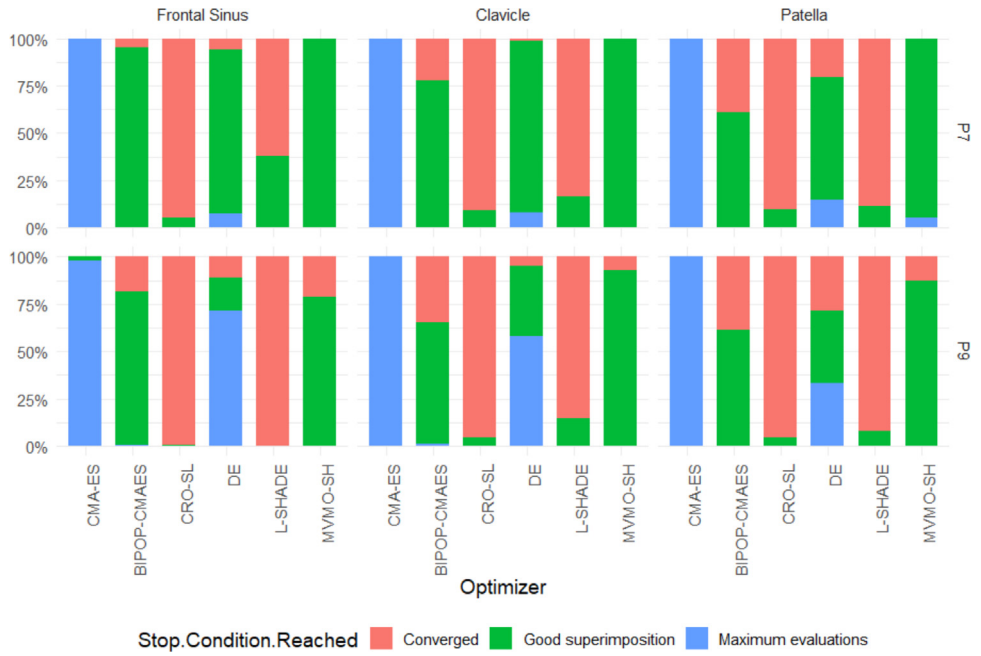

Fig. 7. Boxplots of stop condition (defined in Section 5) reached by the optimization process according to skeletal structure, projective transformation and RCEA optimizer (Experiment II).

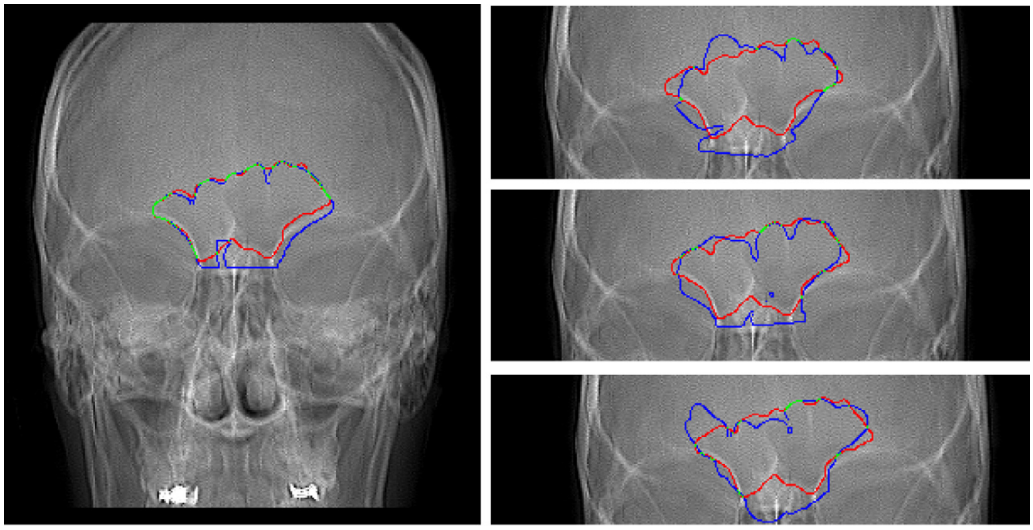

Fig. 8. (Left) An example of a positive case, radiograph *A* compared against CT *A*; (Right) Example of negative cases, radiograph *A* compared against CTs *A*, *B* and *C*.

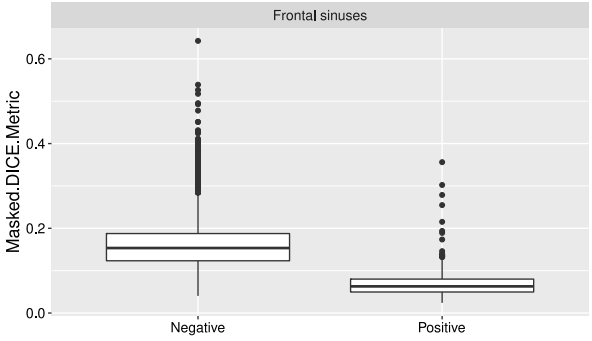

Fig. 9. Boxplots of the minimum error of positive and negative cases according to the Masked DICE metric (Experiment III).

initialization of perspective projection parameters (i.e. the initial parameters are randomly chosen within the range of the parameters of the projective transformations for each skeletal structure according to international acquisition protocols [52]). It considers a completely realistic scenario and thus a more complex version of the CR problem than the one studied in [14] using a new perspective transformation that also models perspective distortions related to angled radiographs. Our aim was to analyze the influence of the RCEA optimizer estimating the registration transformation parameters in the CR solution. To this end, six different state-of-the-art RCEAs (DE, L-SHADE, CMA-ES, BIPOP-CMA-ES, CRO-SL, and MVMO-SH) have been fine-tuned and studied to deal with this challenging and computationally expensive optimization scenario.

In summary, after a detailed analysis of the results obtained by the different RCEAs, we can conclude that the underlying optimization problem within CR is really complex for reasons such

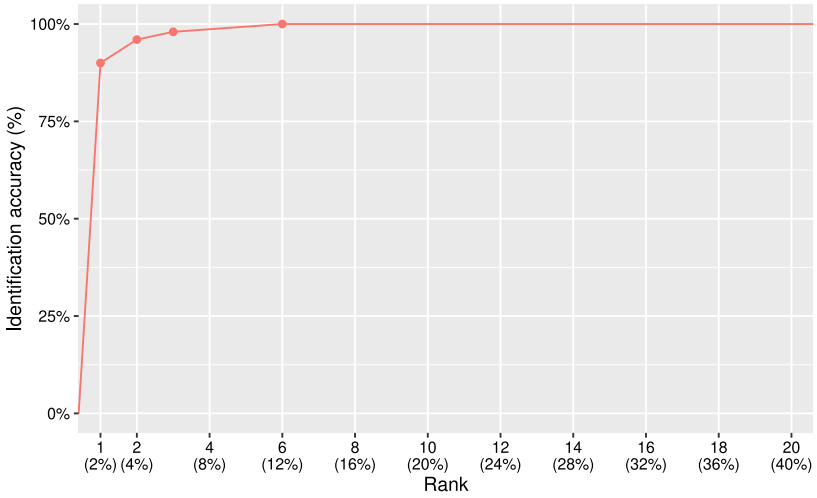

Fig. 10. CMC curve of the comparison of 50 radiographs against 50 CTs (Experiment III).

as the strong correlation among the parameters, their order of magnitude, and the high computational cost. We also confirmed that there is a strong influence of the kind of RCEA considered in a complex problem like CR. Advanced RCEAs such as CMA-ES, L-SHADE, and CRO-SL have not been able to obtain accurate results despite their good behavior in other real-world optimization problems. Nonetheless, promising results have been obtained with MVMO-SH overcoming BIPOP-CMA-ES and the state-of-the-art RCEA for CR, DE. The fine-tuned MVMO-SH (with the following optimizer's "hyper" parameters: $p = \mathrm{d}$, $A_s = 4$, and $F_{end} = 2.5$) allowed us to obtain accurate superimpositions with an average error of 0.001 and 0.006 according to the Masked DICE metric for P7 and P9, respectively, in all the studied bones/cavities (frontal sinuses, clavicles, and patellae). Despite of its stochastic nature, it also showed a robust behavior with a low standard deviation (frontal sinuses, 0.009 for P7 and 0.042 for P9; clavicles 0.002 with P7 and 0.009 with P9; and patellae, 0.026 for P7 and 0.044 for P9) according to GT DICE metric. The results in terms of the mRPD metric with P7 were lower than 1 mm for frontal sinuses and clavicles, and lower than 3 mm for patellae, but were around 15 mm when P9 was considered. Furthermore, by using MVMO-SH, the strong dependency on the kind of bone or cavity was greatly reduced, obtaining accurate results with every bone under study. The main drawback is the computation time required to obtain the superimpositions, that despite having been reduced is still high. Notice that all RCEAs were fine-tuned only with the synthetic CR images of frontal sinuses. Nonetheless, the 3D–2D IR method obtains robust and high quality results with the synthetic CR images of the two other skeletal structures under study (clavicles and patellae), hinting that the proposed automatic 3D–2D IR method with the fine-tuned RCEAs could have a similar performance with other skeletal structures. However, this 3D–2D IR method was designed based on the specific characteristics of the CR superimposition problem and thus cannot be directly used in other 3D–2D IR problems without an specific adaption. Nevertheless, even if our results cannot be directly extrapolated to other IR problems, the importance of performing rigorous optimization comparative studies of EAs as a crucial part of the design and development of new evolutionary IR methods must be highlighted and can serve as a guideline for future research in the area, especially to that dealing with real-world problems.

Lastly, we have validated the evolutionary IR method using the best RCEA, MVMO-SH, for solving real CR problems of frontal sinuses. We have compared 50 skull radiographs against 50 skull CTs, where the frontal sinuses were segmented by forensic anthropology master students at the Physical Anthropology lab (PAL) of the University of Granada. The positive case ranks in the first position in 88% of the cross-comparisons (out of 50 candidates, 2% of the total sample), and a confidence level of 100% of success is reached when the first six positions are considered. Thus, if we compare 50 AM radiographs of possible candidates against the PM 3D image, our proposal with a very preliminary version of the decision making stage, based only on the value of the Masked DICE metric, is able to filter out 88% of the possible candidates with 0 error rate in a fully automatic manner.

Future research is planned to reduce the run time required by studying evolutionary multiresolution IR approaches, surrogate assisted approaches [64], and computation on GPUs. We also plan to study the identification potential of different bones and cavities (both separately and combined) for the CR task [65] through a collaboration with the Israel National Center of Forensic Medicine and the Hebrew University of Jerusalem. Lastly, we plan to develop and validate a hierarchical decision support system that will analyze frontal sinuses superimpositions using multiple criteria (morphological, intensities, etc.), each of them measured using multiple metrics.

6.2. Regarding the soft computing field

Theoretical benchmarks are usually utilized as a means to validate and compare RCEAs. However, these optimization problems are often not able to reflect the full complexity of a real-world optimization problem. In this work, we have performed a rigorous comparative study of several state-of-the-art RCEAs (DE, L-SHADE, CMA-ES, BIPOP-CMA-ES, CRO-SL, and MVMO-SH) in a complex real-world optimization problem, 3D–2D superimposition for CR. The underlying optimization task is computationally expensive, limiting the maximum number of evaluations that can be performed by the optimizer to satisfy the time constrains of the real-world problem. Furthermore, it also establishes certain limitations on the experimental study, since the computational resources and time are limited.

We think that the methodology employed in this paper can be utilized as a guideline for tackling other computationally expensive real-world optimization problems. As a brief summary, the guidelines that have been most relevant to this work are: the utilization of syntactic data with GT solutions for finetuning the RCEAs, the selection of a dataset that represents as many different real scenarios as possible, the study of the stop criteria, the analysis of the convergence of the RCEAs with different number of evaluations, and the validation of the results with real data not utilized in the selection of the best RCEA and its configuration.

In addition, we have been able to conclude that some of the RCEAs which showed to be competitive in some of the competitions developed, did not provide a good performance when applied on a real-world complex problem. On the opposite, MVMO-SH showed up as the best performing optimizer when provided with a sufficient number of evaluations to converge (50,000), confirming its good results in these competitions. We should say that for a lower number of evaluations (5000 and 10,000) the basic DE was the most competitive method, showing that classical RCEAs are still promising in complex real-world problems. It would be interesting to analyze if this behavior could also happen in other kinds of real-world optimization problems.

CRediT authorship contribution statement

Oscar Gómez: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. **Oscar Ibáñez:** Conceptualization, Methodology, Validation, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Andrea Valsecchi:** Conceptualization, Methodology, Validation, Writing - review & editing. **Enrique Bermejo:** Software, Writing - review & editing. **Daniel Molina:** Methodology, Software, Writing - review & editing. **Oscar Cordón:** Conceptualization, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Spanish Ministry of Science, Innovation and Universities, and European Regional Development Funds (ERDF) under grant EXASOCO (PGC2018-101216-B-I00), and by the Regional Government of Andalusia, Spain under grant EXAISFI (P18-FR-4262). Dr. Gómez's work was supported by Spanish MECD FPU grant [grant number FPU14/02380]. Dr. Ibáñez's work is funded by Spanish Ministry of Science, Innovation and Universities-CDTI, Neotec program 2019 [reference EXP-00122609/SNEO-20191236]. We would also like to acknowledge the high performance computing server Alhambra from the University of Granada and the supercomputing center of Galicia (CESGA), that were utilized in this work.

References

[1] T. Bäck, D.B. Fogel, Z. Michalewicz, Handbook of Evolutionary Computation, CRC Press, 1997.
[2] D. Molina, A. LaTorre, F. Herrera, An insight into bio-inspired and evolutionary algorithms for global optimization: Review, analysis, and lessons learnt over a decade of competitions, Cogn. Comput. 10 (4) (2018) 517–544.
[3] U. Chakraborty, Advances in Differential Evolution, Vol. 143, Springer, 2008.
[4] Y. Tenne, C.-K. Goh, Computational Intelligence in Expensive Optimization Problems, Vol. 2, Springer Science & Business Media, 2010.
[5] I.B. Aydilek, A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems, Appl. Soft Comput. 66 (2018) 232–249.
[6] X. Cai, L. Gao, F. Li, Sequential approximation optimization assisted particle swarm optimization for expensive problems, Appl. Soft Comput. 83 (2019) 105659.
[7] I. Erlich, J.L. Rueda, S. Wildenhues, F. Shewarega, Solving the IEEE-CEC 2014 expensive optimization test problems by using single-particle MVMO, in: 2014 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2014, pp. 1084–1091.
[8] J.L. Rueda, I. Erlich, MVMO for bound constrained single-objective computationally expensive numerical optimization, in: 2015 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2015, pp. 1011–1017.
[9] M.J. Thali, B. Brogdon, M.D. Viner, Forensic Radiology, CRC Press, 2002.
[10] F.P. Oliveira, J.M.R. Tavares, Medical image registration: a review, Comput. Methods Biomech. Biomed. Eng. 17 (2) (2014) 73–93.
[11] J. Jomier, E. Bullitt, M. Van Horn, C. Pathak, S.R. Aylward, 3D/2d model-to-image registration applied to TIPS surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2006, pp. 662–669.
[12] D.B. Russakoff, T. Rohlfing, K. Mori, D. Rueckert, A. Ho, J.R. Adler, C.R. Maurer, Fast generation of digitally reconstructed radiographs using attenuation fields with application to 2d-3D image registration, IEEE Trans. Med. Imaging 24 (11) (2005) 1441–1454.
[13] J. Spoerk, C. Gendrin, C. Weber, M. Figl, S.A. Pawiro, H. Furtado, D. Fabri, C. Bloch, H. Bergmann, E. Gröller, et al., High-performance GPU-based rendering for real-time, rigid 2d/3D-image registration and motion prediction in radiation oncology, Z. Med. Phys. 22 (1) (2012) 13–20.
[14] O. Gómez, O.I. nez, A. Valsecchi, O. Cordón, T. Kahana, 3D-2d silhouette-based image registration for comparative radiography-based forensic identification, Pattern Recognit. 83 (2018) 469–480.
[15] J. Santamaría, O. Cordón, S. Damas, A comparative study of state-of-the-art evolutionary image registration methods for 3D modeling, Comput. Vis. Image Underst. 115 (9) (2011) 1340–1354.
[16] C. Stephan, A. Winburn, A. Christensen, A. Tyrrell, Skeletal identification by radiographic comparison: Blind tests of a morphoscopic method using antemortem chest radiographs, J. Forensic Sci. 56 (2) (2011) 320–332.
[17] E. Niespodziewanski, C.N. Stephan, P. Guyomarc'h, T.W. Fenton, Human identification via lateral patella radiographs: A validation study, J. Forensic Sci. 61 (1) (2016) 134–140.
[18] G. Quatrehomme, P. Fronty, M. Sapanet, G. Grévin, P. Bailet, A. Ollier, Identification by frontal sinus pattern in forensic anthropology, Forensic Sci. Int. 83 (2) (1996) 147–153.
[19] T. Back, Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms, Oxford university press, 1996.
[20] O. Cordón, S. Damas, J. Santamaría, A practical review on the applicability of different EAs to 3D feature-based registration, in: Genetic and Evolutionary Computation in Image Processing and Computer Vision. EURASIP Book Series on SP&C, 2007, pp. 247–269.
[21] O. Ibáñez, L. Ballerini, O. Cordón, S. Damas, J. Santamaría, An experimental study on the applicability of evolutionary algorithms to craniofacial superimposition in forensic identification, Inform. Sci. 179 (23) (2009) 3998–4028.
[22] E. Bermejo, M. Chica, S. Damas, S. Salcedo-Sanz, O. Cordón, Coral reef optimization with substrate layers for medical image registration, Swarm Evol. Comput. 42 (2018) 138–159.
[23] S. Damas, O. Cordón, J. Santamaría, Medical image registration using evolutionary computation: An experimental survey, IEEE Comput. Intell. Mag. 6 (4) (2011) 26–42.
[24] A. Valsecchi, E. Bermejo, S. Damas, O. Cordón, Metaheuristics for medical image registration, Handb. Heuristics (2018) 1079–1101.
[25] R. Storn, K. Price, Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (4) (1997) 341–359.
[26] N. Hansen, S.D. Müller, P. Koumoutsakos, Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), Evol. Comput. 11 (1) (2003) 1–18.
[27] A.K. Qin, X. Li, Differential evolution on the CEC-2013 single-objective continuous optimization testbed, in: 2013 IEEE Congress on Evolutionary Computation (CEC), 2013, pp. 1099–1106.
[28] I. Loshchilov, CMA-ES with restarts for solving CEC 2013 benchmark problems, in: 2013 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2013, pp. 369–376.
[29] S. Das, P.N. Suganthan, Differential evolution: A survey of the state-of-the-art, IEEE Trans. Evol. Comput. 15 (1) (2011) 4–31.
[30] J. Santamaría, S. Damas, J.M. García-Torres, O. Cordón, Self-adaptive evolutionary image registration using differential evolution and artificial immune systems, Pattern Recognit. Lett. 33 (16) (2012) 2065–2070.
[31] I. De Falco, A. Della Cioppa, D. Maisto, E. Tarantino, Differential evolution as a viable tool for satellite image registration, Appl. Soft Comput. 8 (4) (2008) 1453–1462.
[32] R. Tanabe, A.S. Fukunaga, Improving the search performance of shade using linear population size reduction, in: 2014 IEEE Congress on Evolutionary Computation (CEC), 2014, pp. 1658–1665.
[33] J. Rönkkönen, S. Kukkonen, K.V. Price, Real-parameter optimization with differential evolution, in: Congress on Evolutionary Computation, 2005, pp. 506–513.
[34] N. Hansen, S. Kern, Evaluating the CMA evolution strategy on multimodal test functions, in: International Conference on Parallel Problem Solving from Nature, Springer, 2004, pp. 282–291.

[35] N. Hansen, Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed, in: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, ACM, 2009, pp. 2389–2396.
[36] N. Hansen, Benchmarking a BI-population CMA-ES on the BBOB-2009 noisy testbed, in: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, ACM, 2009, pp. 2397–2402.
[37] S. Salcedo-Sanz, J. Muñoz-Bulnes, M.J. Vermeij, New coral reefs-based approaches for the model type selection problem: a novel method to predict a nation's future energy demand, Int. J. Bio-inspired Comput. 10 (3) (2017) 145–158.
[38] Y. Ong, M.H. Lim, X. Chen, Memetic computation—Past, present amp; future [research frontier], IEEE Comput. Intell. Mag. 5 (2) (2010) 24–31.
[39] K. Sörensen, Metaheuristics—the metaphor exposed, Int. Trans. Oper. Res. 22 (1) (2015) 3–18.
[40] I. Erlich, G.K. Venayagamoorthy, N. Worawat, A mean-variance optimization algorithm, in: 2010 IEEE Congress on Evolutionary Computation (CEC), 2010, pp. 1–6.
[41] J.L. Rueda, I. Erlich, Hybrid mean-variance mapping optimization for solving the IEEE-CEC 2013 competition problems, in: 2013 IEEE Congress on Evolutionary Computation (CEC), 2013, pp. 1664–1671.
[42] J. Rueda, I. Erlich, Solving the CEC2016 real-parameter single objective optimization problems through MVMO-PHM, in: 2016 IEEE World Congress on Computational Intelligence, 2016, pp. 1–10.
[43] J.L. Rueda, I. Erlich, Hybrid single parent-offspring MVMO for solving CEC2018 computationally expensive problems, in: 2018 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2018, pp. 1–8.
[44] T. Kahana, J. Hiss, Identification of human remains: forensic radiology, J. Clin. Forensic Med. 4 (1) (1997) 7–15.
[45] P. Markelj, D. Tomaževič, B. Likar, F. Pernuš, A review of 3D/2d registration methods for image-guided interventions, Med. Image Anal. 16 (3) (2012) 642–661.
[46] H.L. Runolfsdottir, G. Sigurdsson, L. Franzson, O.S. Indridason, Gender comparison of factors associated with age-related differences in bone mineral density, Arch. Osteoporos. 10 (1) (2015) 1–9.
[47] L.S. Bell, M.F. Skinner, S.J. Jones, The speed of post mortem change to the human skeleton and its taphonomic significance, Forensic Sci. Int. 82 (2) (1996) 129–140.
[48] E.B. van de Kraats, G.P. Penney, D. Tomazevic, T. Van Walsum, W.J. Niessen, Standardized evaluation methodology for 2-d-3-d registration, IEEE Trans. Med. Imaging 24 (9) (2005) 1177–1189.
[49] J. Feldmar, N. Ayache, F. Betting, 3D-2d projective registration of free-form curves and surfaces, in: Computer Vision, 1995. Proceedings., Fifth International Conference on, IEEE, 1995, pp. 549–556.
[50] T. Sørensen, A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons, K. Dan. Vidensk. Selsk. 5 (1948) 1–34.
[51] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.
[52] K.L. Bontrager, J. Lampignano, Textbook of Radiographic Positioning and Related Anatomy, Elsevier Health Sciences, 2013.
[53] D. Mery, X-ray Testing, in: Computer Vision for X-Ray Testing: Imaging, Systems, Image Databases, and Algorithms, Springer International Publishing, Cham, 2015, pp. 1–33, Chapter 1.
[54] H.-G. Beyer, The Theory of Evolution Strategies, Springer Science & Business Media, 2013.
[55] R. Tanabe, A. Fukunaga, Success-history based parameter adaptation for differential evolution, in: 2013 IEEE Congress on Evolutionary Computation (CEC), 2013, pp. 71–78.
[56] M.J. Powell, The BOBYQA Algorithm for Bound Constrained Optimization Without Derivatives, Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, UK, 2009.
[57] M. Tomassini, L. Vanneschi, P. Collard, M. Clergue, A study of fitness distance correlation as a difficulty measure in genetic programming, Evol. Comput. 13 (2) (2005) 213–239.
[58] K. Pearson, Note on regression and inheritance in the case of two parents, Proc. R. Soc. Lond. 58 (1895) 240–242.
[59] G.D. Israel, Determining Sample Size, US Department of Agriculture, Cooperative Extension Service, University of Florida, Institute of Food and Agricultural Sciences, 1992.
[60] S. Pieper, M. Halle, R. Kikinis, 3D slicer, in: IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004, IEEE, 2004, pp. 632–635.
[61] M. López-Ibáñez, J. Dubois-Lacoste, L.P. Cáceres, M. Birattari, T. Stützle, The irace package: Iterated racing for automatic algorithm configuration, Oper. Res. Perspect. 3 (2016) 43–58.
[62] E.A. Gehan, A generalized wilcoxon test for comparing arbitrarily singly-censored samples, Biometrika 52 (1–2) (1965) 203–223.
[63] C. Campomanes-Alvarez, O. Ibáñez, O. Cordón, C. Wilkinson, Hierarchical information fusion for decision making in craniofacial superimposition, Inf. Fusion 39 (Supplement C) (2018) 25–40.
[64] R.T. Haftka, D. Villanueva, A. Chaudhuri, Parallel surrogate-assisted global optimization with expensive functions–a survey, Struct. Multidiscip. Optim. 54 (1) (2016) 3–13.
[65] M. Page, J. Taylor, M. Blenkin, Uniqueness in the forensic identification sciences–fact or fiction?, Forensic Sci. Int. 206 (1) (2011) 12–18.