

The feasibility of manual parameter tuning for deformable breast MR image registration from a multi-objective optimization perspective

Kleopatra Pirpinia¹, Peter A N Bosman², Claudette E Loo³,
Gonneke Winter-Warnars³, Natasja N Y Janssen¹,
Astrid N Scholten¹, Jan-Jakob Sonke¹, Marcel van Herk⁴
and Tanja Alderliesten⁵

¹ Department of Radiation Oncology, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

² Centrum Wiskunde & Informatica, 1098 XG Amsterdam, The Netherlands

³ Department of Radiology, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

⁴ University of Manchester, Institute of Cancer Sciences, The Christie NHS Foundation Trust, M20 4BX Manchester, United Kingdom

⁵ Department of Radiation Oncology, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands

E-mail: k.pirpinia@nki.nl

Received 19 October 2016, revised 5 April 2017

Accepted for publication 24 April 2017

Published 23 June 2017



CrossMark

Abstract

Deformable image registration is typically formulated as an optimization problem involving a linearly weighted combination of terms that correspond to objectives of interest (e.g. similarity, deformation magnitude). The weights, along with multiple other parameters, need to be manually tuned for each application, a task currently addressed mainly via trial-and-error approaches. Such approaches can only be successful if there is a sensible interplay between parameters, objectives, and desired registration outcome. This, however, is not well established. To study this interplay, we use multi-objective optimization, where multiple solutions exist that represent the optimal trade-offs between the objectives, forming a so-called Pareto front. Here, we focus on weight tuning. To study the space a user has to navigate during manual weight tuning, we randomly sample multiple linear combinations. To understand how these combinations relate to desirability of registration outcome, we associate with each outcome a mean target registration error (TRE) based on expert-defined anatomical landmarks. Further, we employ a multi-objective

evolutionary algorithm that optimizes the weight combinations, yielding a Pareto front of solutions, which can be directly navigated by the user. To study how the complexity of manual weight tuning changes depending on the registration problem, we consider an easy problem, prone-to-prone breast MR image registration, and a hard problem, prone-to-supine breast MR image registration. Lastly, we investigate how guidance information as an additional objective influences the prone-to-supine registration outcome. Results show that the interplay between weights, objectives, and registration outcome makes manual weight tuning feasible for the prone-to-prone problem, but very challenging for the harder prone-to-supine problem. Here, patient-specific, multi-objective weight optimization is needed, obtaining a mean TRE of 13.6 mm without guidance information reduced to 7.3 mm with guidance information, but also providing a Pareto front that exhibits an intuitively sensible interplay between weights, objectives, and registration outcome, allowing outcome selection.

Keywords: parameter tuning, deformable image registration, multi-objective optimization, weight tuning, breast MRI

(Some figures may appear in colour only in the online journal)

1. Introduction

Deformable image registration (DIR) has the potential to be a key component of various processes in medical image analysis. It has particular potential in the field of radiotherapy, as it can be used for dose accumulation (Schaly *et al* 2004), response assessment (Boehler *et al* 2010), or treatment adaptation (Foskey *et al* 2005, Mencarelli *et al* 2014). Its use in clinical practice, however, remains limited.

For the majority of currently used registration methods, the first step when performing a registration is to set the values of a multitude of parameters (e.g. weights in cost function, number of image resolution levels, number of optimizer iterations), which are often tuned by hand. This can be a laborious process, as good settings can vary greatly depending on the type of registration problem, or even the specific problem instance at hand. For efficient iterative trial-and-error parameter adjustment to be feasible, a sensible relation between parameters, objectives (typically related to similarity and deformation magnitude), and registration outcome needs to hold. However, whether such a relation generally holds for currently widely used registration methods is not well-established.

We study the feasibility of trial-and-error approaches for DIR problems by investigating the relation between parameters, objectives, and desirability of registration outcomes. We specifically focus on a subset of parameters, i.e. the weights in the optimization function that determine the trade-offs between the objectives of interest in DIR. We propose to use multi-objective optimization (Deb 2001) as a weight-tuning strategy, to navigate the space of optimal registration outcomes. In multi-objective optimization, there is no unique optimal solution, but there are multiple equally good solutions that represent the optimal trade-offs between the objectives.

To see how the interplay between weights, objectives, and desirability of registration outcomes changes depending on the difficulty of the registration problem, we test the methodology on a relatively easy problem, i.e. with limited deformations, namely prone-to-prone breast magnetic resonance (MR) image registration of images acquired pre- and post-radiotherapy.

This can be helpful for treatment response assessment for patients that undergo pre-operative breast irradiation. We then test the methodology on a harder problem, prone-to-supine breast MR image registration, which is associated with large deformations and has been shown to be very challenging for intensity-based registration methods, e.g. (Lee *et al* 2010). Registration of contrast-enhanced prone MR imaging (MRI), currently routinely used for diagnostic purposes, to non-contrast-enhanced supine MRI, can be used for surgical planning, as it translates the diagnostic information to the patients treatment position during breast-conserving surgery. This can be particularly helpful for breast cancer patients with non-palpable tumors.

Preliminary results of this work have been reported in Pirpinia *et al* (2016a) and Pirpinia *et al* (2016b) on a limited number of test cases. In this work, we expand considerably our previous studies by increasing the number of datasets, and by including the following new elements; we investigate whether the addition of guidance information as a third objective, along with similarity and deformation magnitude, influences the quality of the registration outcome for the prone-to-supine registration problem. Further, we perform a quantitative comparison of the acquired Pareto fronts, by using two different comparison metrics. Lastly, we perform a more robust evaluation of the registration accuracy, based on the landmark annotations by two observers instead of one observer.

1.1. Related work

To address the challenge of parameter tuning, automated parameter tuning methods have been proposed for rigid image registration (Valsecchi *et al* 2013). Parameter optimization (including the weights) for DIR has been investigated and applied to computed tomography (CT) lung registration, showing that there can be large variations in the optimal values of parameters even for the same type of registration problem (Dou *et al* 2016). This indicates that parameter tuning is very challenging.

To specifically tackle the challenge of weight tuning in DIR, techniques for adaptively determining the weight associated with the smoothness term that regularizes the transformation during the optimization process have been proposed, Andrews *et al* (2014) and Cahill *et al* (2009). The focus of such approaches was more on accommodating the absence of information (e.g. edges) in parts of the images that could guide the optimization algorithm, and were tested on MR thigh images. Further, a machine learning approach to determine the weights for DIR was also introduced (Yeo *et al* 2010). This approach was validated on a specific class of problems, namely brain image registration. Moreover, all aforementioned approaches regard registration problems where little deformation is needed, in order to obtain an accurate match of the multiple anatomical structures present in the images. For such cases, optimization is typically not very challenging, making a trial-and-error approach likely viable. In contrast, in this work we focus on problems that involve (large) deformations.

2. Materials and methods

We first study the space of weights that a user has to navigate by randomly sampling sets of weight combinations and observing how those correspond to desirable registration outcomes. Further, we employ a multi-objective evolutionary algorithm (EA), in conjunction with open-source image registration software, that finds the weight combinations that lead to multiple, equally good DIR outcomes, in the multi-objective sense. To quantify desirability of registration outcome, we associate to each solution a mean target registration error (TRE). We assume

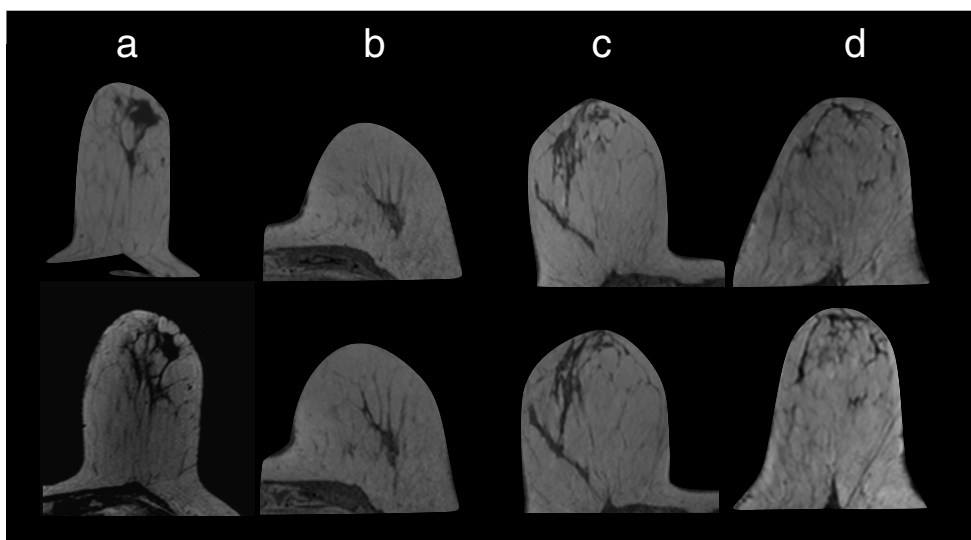


Figure 1. Examples of source (upper row) and target (lower row) images for the prone-to-prone registration problem. The source image is acquired pre-radiotherapy, the target image is acquired post-radiotherapy. From left to right: cases p_1, p_3, p_4, p_5 .

that the TRE aligns with a user's idea of a good registration outcome, hence making a solution with a low mean TRE a preferable solution.

2.1. Datasets

2.1.1. Prone-to-prone. Ten pairs of T1-weighted, non-contrast-enhanced MR images were retrospectively used. The images were acquired pre- and post-radiotherapy of breast cancer patients in prone orientation (figure 1), in a Philips 3 T scanner (Achieva Philips Medical Systems, Best, Netherlands). The patients (age ≥ 60 years, tumor size ≤ 30 mm on MR images) underwent partial breast radiotherapy between December 2012 and June 2015, prior to breast-conserving surgery (van der Leij *et al* 2015). The images have a reconstructed voxel size of $0.9 \times 0.9 \times 1.2$ mm³. The study for which the data was acquired was approved by the institutional review boards of the participating centers, and written informed consent was obtained prior to start of the treatment protocol.

2.1.2. Prone-to-supine. Ten volunteers (age 26–60 years) underwent a T1-weighted, non-contrast-enhanced imaging sequence (used for routine clinical tumor diagnosis in breast cancer patients) in a Philips 3T scanner (Achieva Philips Medical Systems, Best, The Netherlands) both in prone and supine orientation during the same scan session (figure 2). The images have a voxel size of $0.88 \times 0.88 \times 0.90$ mm³. In contrast to the prone-to-prone cases, where markers were not available (due to the retrospective use of the data from a study for which marker placement was not part of the protocol), here, nine self-adhesive MR-visible fiducial markers (MM3005, IZI Medical Products Corporation, Baltimore, MD, USA) were attached to the breast. They were placed in a symmetrical fashion in order to achieve consistent and reproducible positioning. The markers are donut-shaped with a diameter of 15 mm and a thickness of 3.5 mm. Details about the marker configuration are provided in section 2.3.2. Written informed consent was obtained from the volunteers for the MR acquisition.

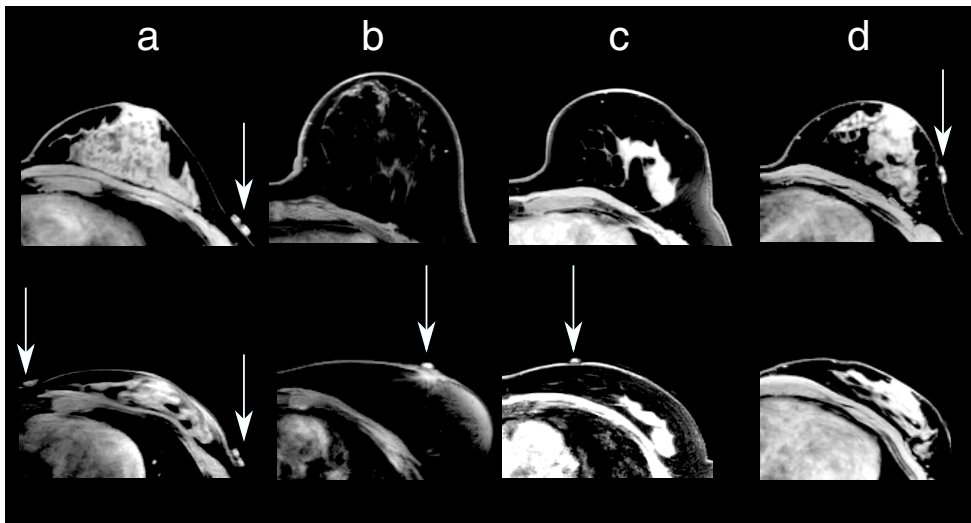


Figure 2. Examples of source (upper row) and target (lower row) images for the prone-to-supine registration problem. The source image is in prone orientation, the target image is in supine orientation. From left to right: cases v_1 , v_3 , v_7 , v_8 . Arrows indicate marker locations.

2.2. Multi-objective parameter tuning for DIR

DIR can naturally be viewed as a problem where multiple objectives are of interest. In DIR, not only a high degree of similarity between the target image and the transformed source image is desirable, but also an anatomically correct transformation T between the source image and the target image. Another objective of interest can be to minimize the distance between points with a known correspondence so as to aid the registration by use of guidance information. In current state-of-the-art registration methods, all m objectives of interest O_i are linearly combined into one cost function C that needs to be optimized, taking the following form:

$$C(T) = \lambda_0 O_0(T) + \lambda_1 O_1(T) + \lambda_2 O_2(T) + \dots + \lambda_{m-1} O_{m-1}(T). \quad (1)$$

The weights λ_i , $i \in \{0 \dots m-1\}$ determine the compromise between the objectives and need to be determined beforehand.

To compute such linear combinations, in this work we used `elastix`, an open-source image registration toolbox (Klein *et al* 2010). Within `elastix`, we used a transformation model based on B-splines (Rueckert *et al* 1999). We employed `elastix` to run linear combinations of two objectives: dissimilarity and deformation magnitude, as well as three objectives: dissimilarity, deformation magnitude, and guidance information.

2.2.1. Random sampling (RS) of linear combinations. We randomly sampled multiple linear weight combinations for $m = 2, 3$ objectives, which correspond to sets of 2 and 3 weights. The value of the weights was chosen to be sampled in $[0,1]$.

2.2.2. Optimized linear combinations (OLC). The preferred registration outcome encompasses the optimal trade-off between all the objectives, which, as expressed by the relative weighting of the objectives of interest, can vary greatly depending on the registration problem as well as the specific set of images to be registered. Hence, by design, the use of

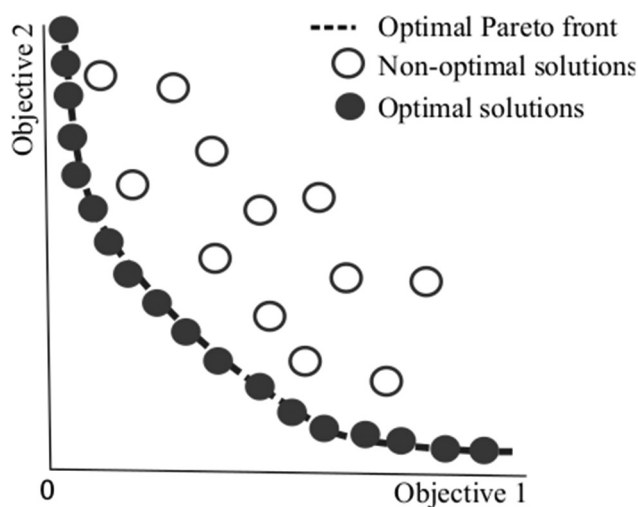


Figure 3. A convex optimal Pareto front with non-dominated solutions (filled) and dominated solutions (non-filled). Here, both objectives are to be minimized.

single-objective optimization for these problems puts the user potentially at a disadvantage. Multi-objective optimization (see appendix A.1) focuses on such problems, where there exist multiple objectives of interest and decisions need to be taken in the presence of trade-offs between the objectives. Consequently, there exist multiple solutions that can be considered equally good, forming the Pareto front (figure 3).

The goal is to find linear combinations that correspond to registration outcomes (i.e. after running `elastix`) that form a Pareto front, but also are as good as possible in objective space. To this end, we employed a multi-objective EA (Deb 2001) to optimize the weight combinations. EAs are population-based optimization methods. Being able to simultaneously advance an entire population of solutions, and thereby approximating the Pareto front in one run, makes EAs state-of-the-art in multi-objective optimization.

The specific EA we used is an estimation-of-distribution algorithm (Lozano *et al* 2006) known as iMAMaLGaM (incremental multi-objective adapted maximum likelihood Gaussian model) (Rodrigues *et al* 2014). To evolve the population, in every generation 35% of the best solutions are selected based on a well-known domination ranking scheme (Deb *et al* 2002) and are subsequently separated into clusters on the basis of their objective values, distributing the search bias across the Pareto front. An l -dimensional Gaussian mixture distribution is subsequently estimated (l is the number of parameters) and then sampled to generate new candidate solutions. iMAMaLGaM has been shown to perform very well on benchmark as well as real-world problems (Bosman and Alderliesten 2012).

For the optimization problem at hand, each candidate solution is a set of weights which is given as input to `elastix`, which yields a value for each objective (after `elastix` has converged to a solution). The values of the objectives are then passed back to iMAMaLGaM.

After testing different settings on benchmark problems, we allowed iMAMaLGaM to run for 100 generations, or equivalently use approximately 4000 evaluations per objective. The remainder of the EA-specific settings were set according to guidelines that were shown to work well on various benchmark problems (Bosman and Alderliesten 2012).

2.3. Registration experiments and setup

For both registration problems, the images were first rigidly aligned using `elastix`. For DIR, default `elastix` parameter settings were used, excluding the number of resolution levels, the final control point grid spacing, and the number of optimizer iterations per resolution, which were determined after testing different settings on a subset of images included in this study and following `elastix` guidelines. For both datasets, the fixed number of iterations of the optimizer per resolution was set to 200, and the final control point grid spacing was set to $8 \times 8 \times 8$ mm. Furthermore, we used the adaptive stochastic gradient descent algorithm (Klein *et al* 2008) available in `elastix` as the optimizer. For the prone-to-prone registration problem, we used four resolutions, whereas for the prone-to-supine problem, five resolutions.

2.3.1. Prone-to-prone. We needed to define the dissimilarity objective as well as the objective that describes the magnitude of the deformation. For the first objective we used the negative normalized correlation coefficient (NCC). To quantify the deformation magnitude, we used the bending energy penalty (Wahba 1990).

2.3.2. Prone-to-supine. First, the positions of the external markers were annotated on all pairs of images. For each pair, a region of interest was manually defined on the target image. The region of interest included the volume of the breast to be registered up to and including the pectoral muscle. We then pre-processed the images in three different ways, resulting in three different subsets. In the first subset, we removed the MR-visible markers from the images using ImageJ (Schneider *et al* 2012) (by replacing them with a uniform intensity region of gray value 0) so as to investigate the performance of the registration algorithm without any guidance information. In this case, we used two objectives, dissimilarity and deformation magnitude, as in the prone-to-prone registration experiments. In the second subset, we removed 5 out of 9 markers from the images, in a way that results in a more sparse but still uniform distribution of markers on the breast. In the third subset, we did not remove the markers (figure 4).

For the second and third subsets we used three objectives: dissimilarity, deformation magnitude, as defined previously, and guidance error. Guidance error is expressed as the mean Euclidean distance between the locations of the center of the external markers in the transformed source and target image that needs to be minimized (Baiker *et al* 2011). The definitions of all objectives can be found in the `elastix` manual.

2.4. Evaluation

2.4.1. Pareto front comparison. A measure that accurately describes the quality of a Pareto front should capture several aspects, such as number of solutions on the Pareto front, proximity to the optimal Pareto front (if known), as well as diversity of the solutions; i.e. how well the solutions are distributed along the Pareto front. The hypervolume indicator (Zitzler and Thiele 1998) (see appendix A.2) is a unary measure that encapsulates these properties in one number. The larger the hypervolume, the higher the quality of the Pareto front. To be able to calculate the hypervolume for the RS approach, we derived Pareto fronts from the sets of the found solutions.

Another measure for Pareto front comparison is the percentage of solutions found by, e.g. algorithm A that are dominated by the solutions found by algorithm B. If this percentage is (close to) 100%, whereas the percentage of solutions of algorithm B dominated by A is small, then algorithm B can be said to be truly superior under the tested circumstances.

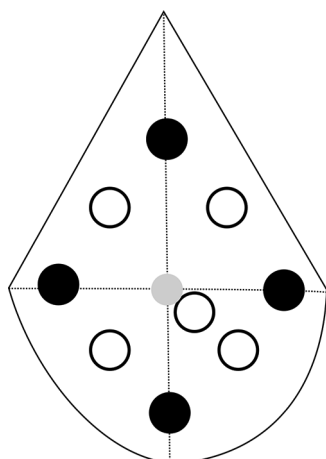


Figure 4. Illustration of the marker configuration on the breast. For the first subset of experiments, all markers were removed. For the second subset of experiments, the non-filled markers were removed. For the third subset of experiments, no markers were removed.

2.4.2. Registration accuracy. To evaluate the quality of the registration outcomes, an experienced breast radiologist annotated 8 to 12 internal anatomical landmarks in the source image and their corresponding locations in the target image (see example in figure 5). We calculated the mean TRE as the mean Euclidean distance between the landmark locations in the target image and their locations in the transformed source image. We calculated the mean TRE after rigid registration (which was used as our baseline), as well as after DIR.

Furthermore, for the subsets of experiments where the external markers were removed, we used the removed markers (i.e. those not used as guidance information) to calculate the mean external marker-based TRE. Lastly, a second breast radiologist re-annotated the internal landmarks on the target image based on those that were annotated in the source image by the first radiologist. In this way, we could determine if the registration error falls into the range of the inter-observer variability of the landmark annotations by the two radiologists.

To compare the performance of RS and OLC on these problems, we used the Wilcoxon signed rank test (non-normal, paired data), with $p < 0.05$ as the significance level.

3. Results

3.1. Prone-to-prone

For the prone-to-prone registration problem, 86.8–96.7% of the solutions found by the RS approach are dominated by those found by OLC, with the inverse being 3.7–12.6%. Furthermore, the OLC approach obtains larger hypervolume indicator values (table 1) for a comparable number of evaluations (approximately 4000 random samples) ($p = 0.002$). The OLC obtains solutions with a larger TRE distribution (figure 6), as it looks for solutions all along the Pareto front, including those with little to no deformation. Nonetheless, both approaches obtain solutions with a low mean TRE (and thereby preferable solutions) (table 2). After rigid registration, mean TRE ranged from 3.8–7.9 mm and after DIR it was reduced to ≤ 2.9 mm for all cases and both approaches (figure 6).

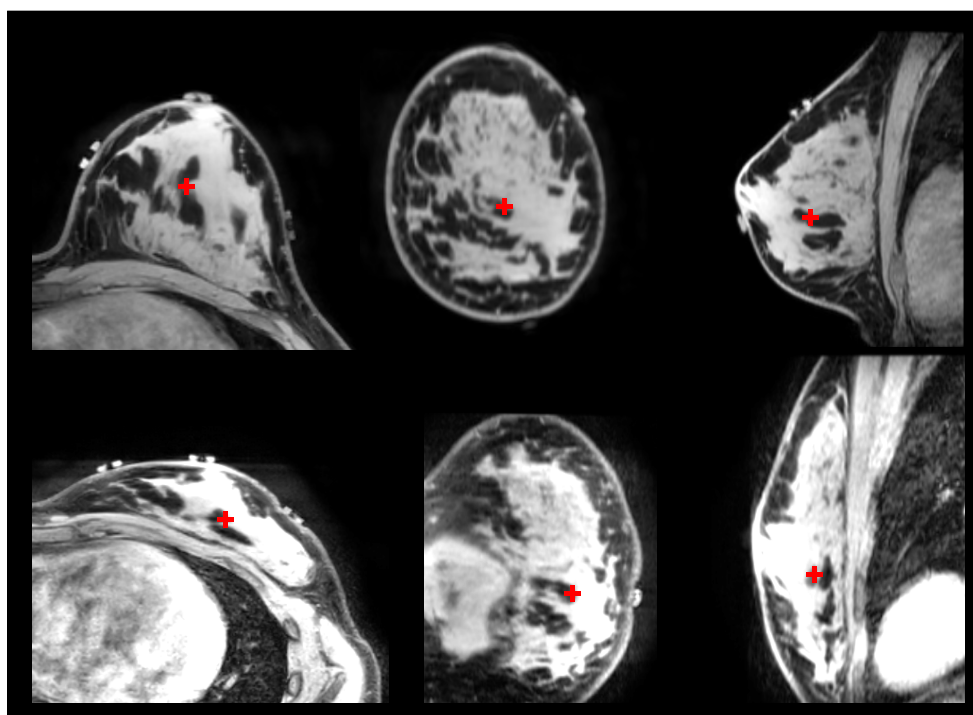


Figure 5. Example (case v_6) of an annotated landmark in a prone (upper row) and supine (lower row) image.

The two radiologists were in agreement about the landmark annotations, with a mean distance of approximately 1.0 mm. Furthermore, the distribution of the TRE in objective space is smooth (mean TRE decreases as deformation, described by the deformation magnitude objective, increases), for the set of solutions obtained by the RS as well as the Pareto front obtained by the OLC approach (figure 7).

3.2. Prone-to-supine

Here, the algorithmic performance of the OLC approach is overall superior; the domination percentage for OLC over RS ranges from 65.6–100% (the inverse ranges from 0.0–32.4%) and hypervolume indicator values are larger (table 1) (no markers: $p = 0.0156$, four markers: $p = 0.0039$, nine markers: $p = 0.0039$). When there is no guidance information, for a comparable number of evaluations, again this did not result in large differences between the two approaches in lowest mean TRE for most cases (figure 8). Furthermore, in 4 out of 10 cases (v_3, v_5, v_7, v_{10}) it was not possible to find acceptable registration outcomes, due to the very difficult and large breast deformation as well as some intensity inhomogeneities present in the MR images (see e.g. figure 2(b)). This is reflected in the lowest mean TRE (table 2) as well as the lowest mean external marker-based TRE (table 3).

Including guidance in the optimization as a third objective leads to improved results for multiple cases; the overall mean TRE (table 2) decreases from 13.6 mm to 10.4 mm with four markers to 7.3 mm with 9 markers for the OLC approach, which performs better than the RS approach ($p < 0.05$); registration examples found by OLC can be seen in figure 9. Also

Table 1. Hypervolume indicator values of Pareto fronts found by RS and OLC approach for cases p_1 – p_{10} and cases v_1 – v_{10} .

	Prone-to-prone		Prone-to-supine						
	—		0 markers		4 markers		9 markers		
	RS	OLC	RS	OLC	RS	OLC	RS	OLC	
p_1	2.09×10^{-4}	2.10×10^{-4}	v_1	3.8×10^{-3}	3.8×10^{-3}	4.1×10^{-2}	4.5×10^{-2}	4.3×10^{-2}	4.7×10^{-2}
p_2	3.92×10^{-4}	3.97×10^{-4}	v_2	2.2×10^{-3}	2.2×10^{-3}	8.8×10^{-2}	1.1×10^{-1}	9.1×10^{-3}	1.0×10^{-2}
p_3	3.52×10^{-4}	3.54×10^{-4}	v_3	10.8×10^{-4}	10.9×10^{-4}	3.1×10^{-1}	3.7×10^{-1}	3.0×10^{-1}	4.7×10^{-1}
p_4	2.75×10^{-4}	2.79×10^{-4}	v_4	5.0×10^{-3}	5.2×10^{-3}	2.6×10^{-2}	2.9×10^{-2}	3.0×10^{-1}	3.3×10^{-1}
p_5	4.32×10^{-4}	4.37×10^{-4}	v_5	2.9×10^{-4}	3.1×10^{-4}	5.0×10^{-1}	5.8×10^{-1}	1.9×10^{-1}	2.2×10^{-1}
p_6	6.25×10^{-4}	6.28×10^{-4}	v_6	2.7×10^{-5}	3.6×10^{-5}	7.6×10^{-2}	8.8×10^{-2}	10.2×10^{-2}	11.4×10^{-2}
p_7	6.71×10^{-4}	6.77×10^{-4}	v_7	4.2×10^{-4}	4.5×10^{-4}	1.8×10^{-1}	2.1×10^{-1}	4.0×10^{-1}	4.4×10^{-1}
p_8	6.72×10^{-4}	6.91×10^{-4}	v_8	37.3×10^{-4}	37.8×10^{-4}	5.5×10^{-2}	7.2×10^{-2}	6.0×10^{-2}	6.8×10^{-2}
p_9	3.59×10^{-4}	3.68×10^{-4}	v_9	4.6×10^{-4}	4.8×10^{-4}	3.5×10^{-2}	3.9×10^{-2}	2.3×10^{-2}	2.5×10^{-2}
p_{10}	4.76×10^{-4}	5.00×10^{-4}	v_{10}	3.1×10^{-4}	3.4×10^{-4}	2.5×10^{-1}	3.3×10^{-1}	6.2×10^{-1}	7.8×10^{-1}

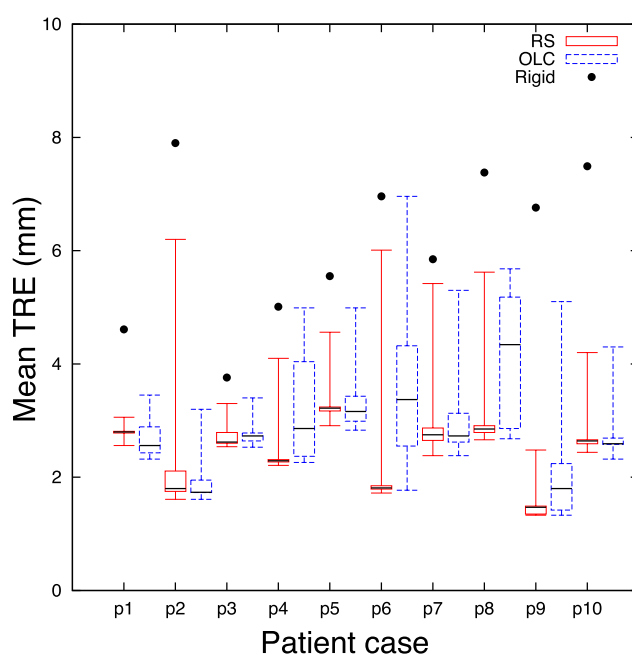


Figure 6. Mean internal landmark-based TRE distribution of solutions found by RS (red) and OLC (blue) for prone-to-prone cases p_1 – p_{10} . Mean TRE after rigid registration shown in black.

here the observers were in agreement about the landmark annotations, with a mean distance of 1.5 mm. Further, the OLC approach finds a larger range of superior solutions for most test cases, as reflected in the distribution of mean TRE (figure 8), see e.g. figure 10. Here, the weights are color-coded in RGB space by straightforwardly placing $(\lambda_0, \lambda_1, \lambda_2)$ in the unit RGB cube. There is no smooth distribution of the TRE in neither parameter or objective space for the RS approach, whereas this holds for the OLC. In the case of the OLC approach, navigating the Pareto front directly, i.e. *a posteriori*, no longer requires the weight combinations to have an intuitive logical ordering along the Pareto front (i.e. the weight color-codings in figures 10(c) and (d) A2(c) and (d) A3(c) and (d)). Instead, only the TRE distribution (the TRE color-codings in figures 7(b), 10(b), A2(b) and A3(b), which should correspond to a user's intuition of registration desirability, should have an intuitive logical ordering. Lastly, the weight distribution as illustrated in the color-coding of figures 10(d), A2(d) and A3(d), in combination with the solution distribution in objective space varies considerably between test cases.

4. Discussion

In this work, we studied DIR from a multi-objective perspective, to better understand the feasibility of trial-and-error approaches for parameter tuning of existing single-objective DIR software. To our knowledge, it is the first study to do so. We studied the space of trade-offs that a human has to manually navigate by means of iterative weight adjustment and how the complexity of such a task changes depending on the difficulty of the registration problem. Results suggest that for relatively simple problems, iterative weight adjustment via trial and

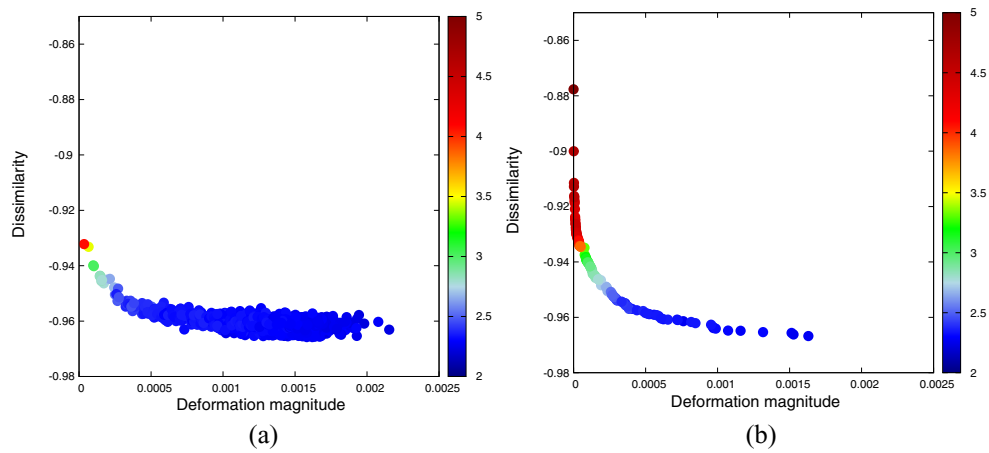


Figure 7. Solutions acquired via RS (left) and OLC (right) for prone-to-prone registration case p_4 . Color-coded internal landmark-based mean TRE in mm.

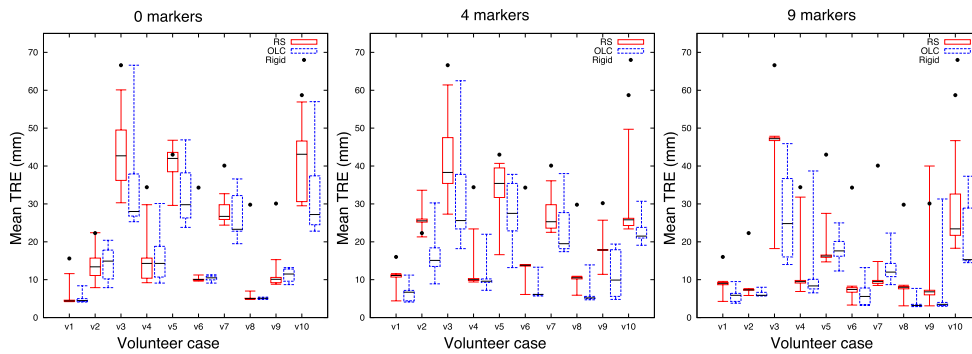


Figure 8. Mean internal landmark-based TRE distribution of solutions after rigid registration (black), after DIR by RS (red) and OLC (blue) for 0 markers (left), 4 markers (middle) and 9 markers (right).

Table 2. Average (and standard deviation) over all cases of the lowest internal landmark-based mean TRE in mm found by RS and OLC.

	RS	OLC
Prone-to-prone	2.3 (0.5)	2.2 (0.5)
Prone-to-supine (0 markers)	15.8 (11.1)	13.6 (8.3)
Prone-to-supine (4 markers)	14.7 (8.6)	10.4 (6.3)
Prone-to-supine (9 markers)	8.6 (6.2)	7.3 (4.4)

error can be sufficient. This is indicated by the objective space shown by the RS approach on the prone-to-prone registration problem, which shows a sensible interplay between objectives and registration quality. Therefore for this particular registration problem, although the algorithmic performance of the OLC approach is superior for the majority of the cases, this does not always result in clinically relevant differences. This is not the case for the prone-to-supine registration problem with guidance information as the third objective. Here, the space shown

Table 3. Lowest external marker-based mean TRE found by RS and OLC for prone-to-supine datasets with 4 and 0 markers. Note that for the case of 4 remaining markers, their guidance error is close to 0, since it is being explicitly minimized as a third objective, but as seen below this is not representative of registration accuracy.

	Prone-to-supine				
	Rigid	4 markers		0 markers	
		RS	OLC	RS	OLC
v_1	16.0	5.1	5.2	8.6	8.1
v_2	22.3	15.9	5.9	9.2	8.9
v_3	52.3	23.5	19.4	25.6	20.3
v_4	34.5	10.0	6.4	8.6	7.9
v_5	48.2	13.2	14.5	13.1	13.1
v_6	34.3	7.5	7.3	10.6	10.4
v_7	40.1	15.5	12.1	23.1	23.3
v_8	20.2	4.9	4.2	8.5	8.4
v_9	28.2	11.9	4.5	9.2	9.0
v_{10}	55.7	18.9	15.8	24.3	23.6
Mean		12.6	9.5	14.1	13.3
SD		6.0	5.5	7.2	6.2

by the RS approach does not exhibit an intuitively simple and sensible interplay neither in parameter or objective space, making trial-and-error approaches very challenging. For this hard problem, a space with a sensible interplay between parameters, objectives and registration outcomes is obtained only when a powerful optimizer such as the EA is used. Assuming that the user's intuition about a good outcome aligns with the TRE, navigating the Pareto front obtained by the OLC should be far easier. Moreover, for these prone-to-supine cases the OLC approach obtains a larger range of registration outcomes that are also superior. It should however be noted that given an optimizer of a non-local nature, the results of the RS approach would potentially be equivalent to the results of the OLC approach for all cases, provided that a mechanism is used to ensure a good, even spread along the Pareto front, provided that the Pareto front is convex. Lastly, the variation in the weight and solution distribution between different test cases indicate that deriving task-optimal parameter settings may be very challenging, and hence, patient-specific tuning is needed.

Independently of the difficulty of the problem, choosing the right parameter settings is always a non-trivial task. When users need to define, e.g. the weights, they have to take into account not only the relative importance of one objective versus the other objective, but also for example the difference in the magnitude of the objectives, in order to come up with an appropriate linear combination. For these reasons, even the range from which the weights can be sampled can be a non-trivial task to determine, as was observed for the three-objective case in this article. This issue is further exemplified by additional experiments that we performed using different combinations of weights (e.g. linear combinations versus convex combinations). For the three-objective case, we restricted the search space to the unit simplex (i.e. ensuring that all weights sum up to 1), which, mathematically, describes the same optimal solutions, but led to inferior results compared to those without the simplex restriction. This is indicative that the more compact search space in the simplex is either more complex to search, or that due to implementation details, the weight combinations do not describe exactly the same solution obtained by *elastix* (e.g. a set of weights (0.3, 0.7) does not give the same solution as (0.03, 0.07)). We therefore additionally performed experiments with linear

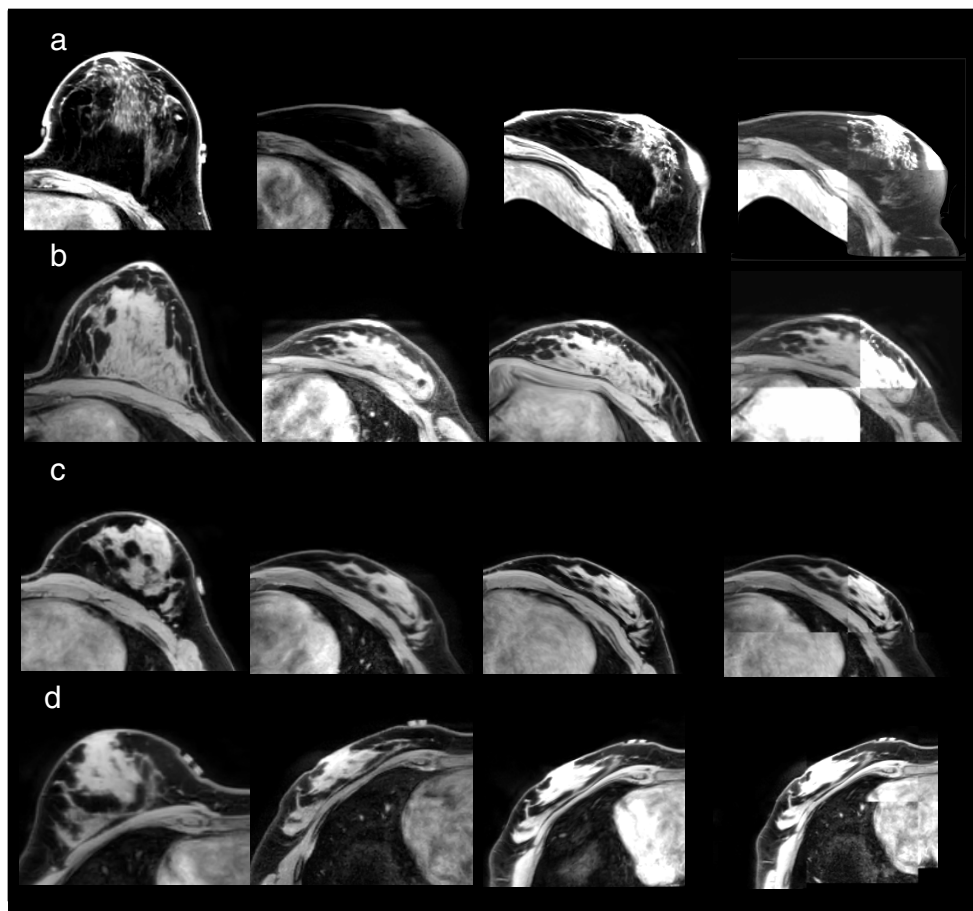


Figure 9. Examples of prone-to-supine registration outcomes (associated to a low mean TRE) with 9 markers, for cases v_3 (a), v_6 (b), v_8 (c) and v_9 (d). From left to right: source image, target image, transformed source image, checkerboard image of transformed source and target. Structures such as the heart are not well registered because they were not included in the region of interest.

combinations rescaled by a constant factor, which resulted in slightly different values for the objectives. This indicates that there are also optimizer-specific details that differ per registration algorithm that the user may be unaware of and that may also affect the way in which each linear combination is being optimized.

Interestingly, the multi-objective approach used here to gain insight into the feasibility of manual parameter tuning could itself be considered as a tuning strategy, as it does not suffer from these hidden implementation details but rather exploits this information to obtain the Pareto front. It may well be better then to use less restrictive parameter ranges to allow the EA-based parameter tuner to obtain the best results, albeit potentially at the cost of a longer runtime. Further, in this context the approach can straightforwardly be used to optimize more parameters that are important for registration (e.g. number of resolutions, step size of the gradient optimizer, final control point spacing) in order to further improve the registration outcomes. Lastly, although in this work *elastix* was chosen as it provides the freedom to

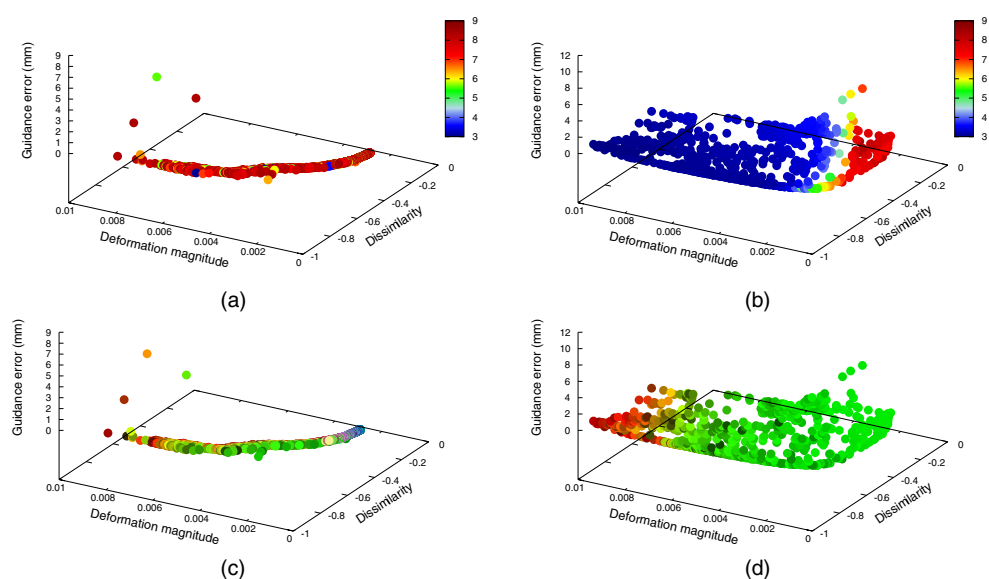


Figure 10. Solutions acquired via RS (left) and OLC (right) for prone-to-supine registration case v_8 (9 markers). Upper row: color-coded internal landmark-based mean TRE in mm, lower row: color-coded weight distribution in RGB space.

adjust parameters, in principle, the approach can be integrated with any commercially available registration framework and applied to any DIR problem.

Some of the prone-to-supine registration problems are just too hard to solve (with or without guidance information) due to the limitations of the transformation model as well as the intensity-based similarity metric. However, for a subset of the DIR problems with guidance information, results comparable to the current state-of-the-art in literature were obtained; namely, mean external marker-based TRE in the order of 8.4 mm found by Han *et al* (2014), which was obtained by using biomechanical modelling in combination with intensity-based registration. Biomechanical modelling methods (Hipwell *et al* 2016) as well as the use of a dual-dynamic transformation model (Alderliesten *et al* 2013) have been helpful in capturing the deformation for prone-to-supine registration. Recently, for four test cases, a mean internal-landmark based TRE of 3.7 mm was reported using symmetric biomechanical modelling (Eiben *et al* 2016). In our study, comparable results were achieved only with the use of guidance information, which is not currently present in a clinical setting. Nonetheless, this could easily be introduced. Moreover, it is also possible that, with more insightful weight tuning for intensity-based registration, the combination of biomechanical modeling and intensity-based registration may lead to even better results.

A limitation of the approach presented in this article is the susceptibility of the selected underlying optimization method used by `elastix` for the registration to the inherent incapability of linear scalarization methods of capturing certain parts of a Pareto front (e.g. concave parts) (Das and Dennis 1997) as well as to the local-search nature of the underlying optimization method. Nonetheless, potentially clinically useful solutions were obtained for both types of registration problems. Whether the best registration outcomes in terms of lowest mean TRE, however, will be actually identified and chosen as the preferred ones by an expert who will navigate the solution space of the Pareto front without having knowledge of the TRE, still needs to be investigated, as part of future work. It should be noted that the EA can be used

not only as a parameter tuner, but also as the optimization component of a registration method (Alderliesten *et al* 2013), which removes the limitation of not being able to capture concave parts of Pareto fronts.

Further, in some cases where the images displayed intensity inhomogeneities (inherent to the imaging sequence), the choice of NCC as a similarity metric might not be optimal, or preprocessing might be needed in the last resolution step, such as applying an unsharp mask filter (Dekker *et al* 2003).

The runtime, approximately 1 day per registration case, as needed to obtain the results presented in this study, is too long for most clinical applications. It should be noted, however, that the current implementation was not optimized for speed. The long runtime can be mitigated by using, e.g. the GPU implementation of `elastix` (Shamonin *et al* 2014). We should also note the EA has negligible overhead, resulting in (almost) the same calculation time for RS and OLC. Further, the EA can be easily parallelized, potentially significantly reducing runtime requirements. Lastly, it should be noted that the proposed approach could be a first step towards derivation of task-optimal parameter settings for at least some DIR problems, possibly removing the need for applying it for every DIR task.

5. Conclusions

In this work, we investigated the feasibility of manual parameter tuning for DIR from a multi-objective optimization perspective. We thereby obtained insight into the interplay between parameters, objectives of interest in DIR and quality of registration outcomes. By using random linear combinations as a surrogate for manual navigation of the search space by a user, we were able to give insight into why such a task can be very challenging for hard DIR problems, but feasible for easier DIR problems. By considering DIR from a multi-objective perspective, the user can navigate the space of optimized registration trade-offs in the form of a Pareto front which contains high-quality solutions, getting the most out of the registration software being used. Moreover, we found that the aforementioned interplay along this Pareto front does align with intuition in all tested cases, facilitating result selection.

Acknowledgments

This work was funded by the Dutch Cancer Society (KWF; Grant No. KWF 2012-5716).

Appendix

A.1. Multi-objective optimization

In multi-objective optimization we assume to have m objectives $f_i(\mathbf{x})$, $i \in \{0, 1, \dots, m-1\}$ that need to be optimized simultaneously. Without loss of generality, we assume that the goal is to *minimize* all objectives. Here, optimality is based on the notion of (Pareto) domination. A solution \mathbf{x}^1 is said to (Pareto) dominate a solution \mathbf{x}^2 (denoted $\mathbf{x}^1 \succ \mathbf{x}^2$) if and only if $f_i(\mathbf{x}^1) \leq f_i(\mathbf{x}^2)$ holds for all $i \in \{0, 1, \dots, m-1\}$ and $f_i(\mathbf{x}^1) < f_i(\mathbf{x}^2)$ holds for at least one $i \in \{0, 1, \dots, m-1\}$. A *Pareto set* of size n is a set of solutions \mathbf{x}^j , $j \in \{0, 1, \dots, n-1\}$ for which no solution dominates any other solution, i.e. there are no $j, k \in \{0, 1, \dots, n-1\}$ such that $\mathbf{x}^j \succ \mathbf{x}^k$ holds. A *Pareto front* corresponding to a Pareto set is the set of all m -dimensional objective values corresponding to the solutions, i.e. the set of all $\mathbf{f}(\mathbf{x}^j)$, $j \in \{0, 1, \dots, n-1\}$. A solution \mathbf{x}^1 is said to be *Pareto optimal* if and only if there is no other \mathbf{x}^2 such that $\mathbf{x}^2 \succ \mathbf{x}^1$

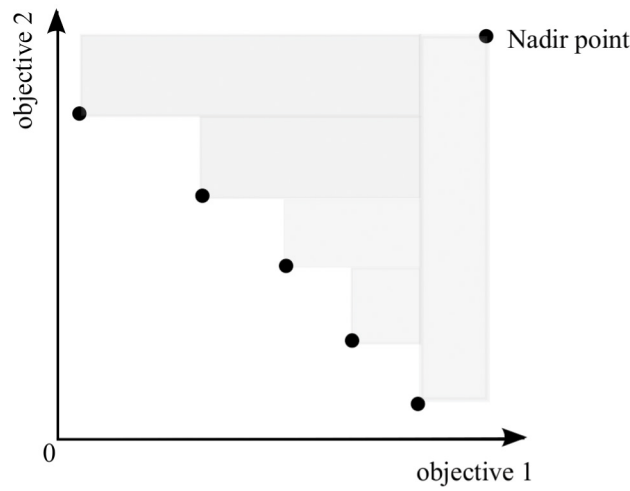


Figure A1. Visualization of the hypervolume (grey) contained by a set of points with respect to the Nadir point.

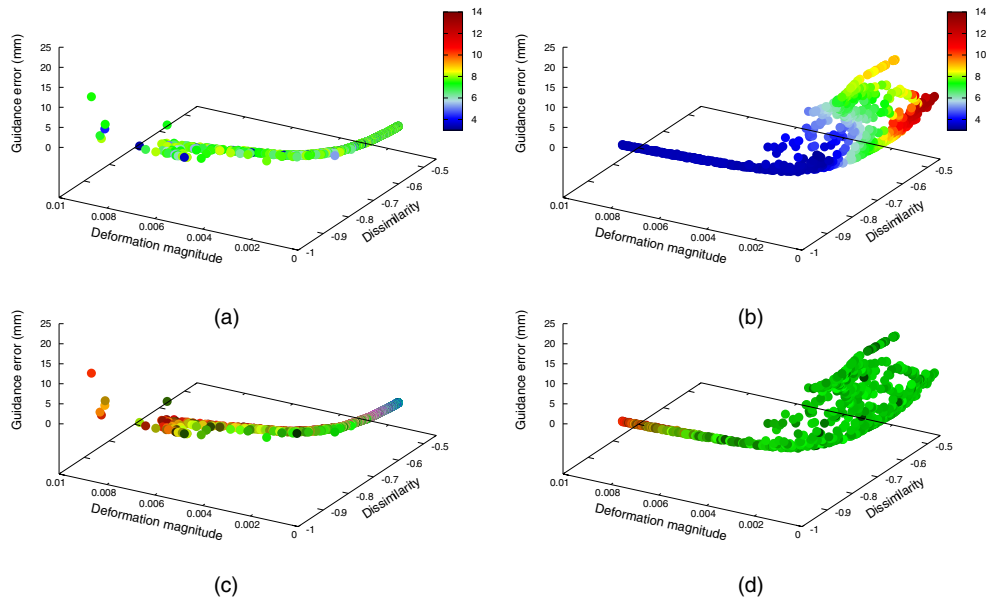


Figure A2. Solutions acquired via RS (left) and OLC (right) for prone-to-supine registration case v_6 (9 markers). Upper row: color-coded internal landmark-based mean TRE in mm, lower row: color-coded weight distribution in RGB space.

holds. Further, the *optimal Pareto set* is the set of all Pareto-optimal solutions and the *optimal Pareto front* is the Pareto front that corresponds to the optimal Pareto set. If the optimization problem is convex (i.e. the objective functions as well as the Pareto front are convex), then the optimization of any linear combination can potentially yield a solution on the optimal Pareto front (Das and Dennis 1997). This, however, holds only when the optimizer solves the linear combination problem to global optimality. Furthermore, no linear combination can lead to

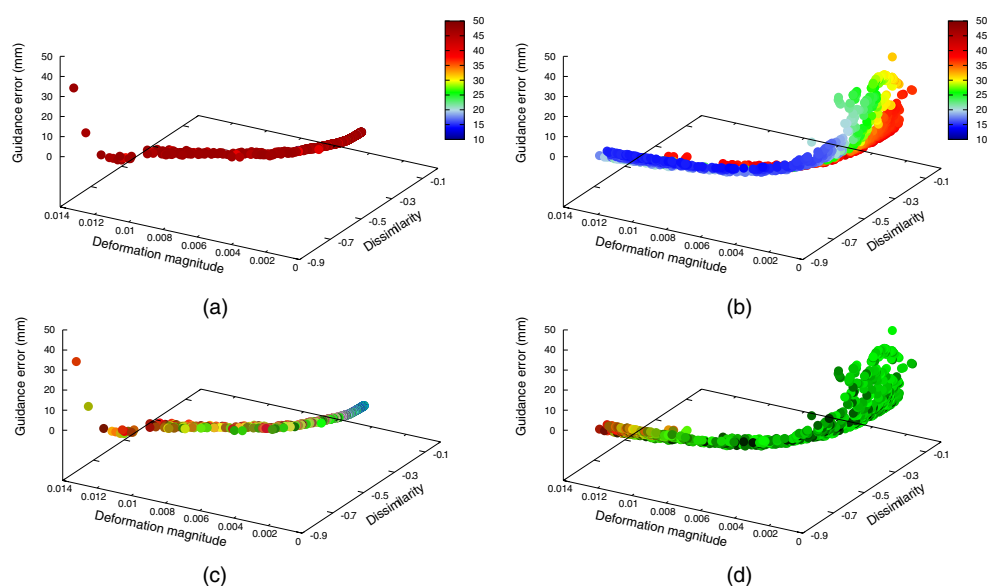


Figure A3. Solutions acquired via RS (left) and OLC (right) for prone-to-supine registration case v_3 (9 markers). Upper row: color-coded internal landmark-based mean TRE in mm, lower row: color-coded weight distribution in RGB space.

solutions on concave parts of the Pareto front. Lastly, a uniform sampling of linear combinations does not necessarily lead to a uniformly spaced set of solutions in objective space.

A.2. Hypervolume

The hypervolume is the size of the n -dimensional space (where n is the number of objectives) contained by a set of points, bounded by one reference point, called the Nadir point (figure A1). The Nadir point is defined as the point composed by taking the ‘worst’ (in our case, the highest) values for all objectives (as found by either RS or OLC), and multiplying by 1.1. Hypervolumes per registration case are comparable, however, hypervolumes cannot be compared between cases, since each one depends on the shape of the Pareto front of the specific registration problem. More details about the hypervolume indicator as well as how to compute it can be found in Fonseca *et al* (2006) and Auger *et al* (2009).

A.3. Examples of sets of solutions and Pareto fronts

Different examples of sets of solutions obtained by RS and Pareto fronts obtained by OLC can be found, for registrations with various degrees of success, based on the mean TRE (figures A2 and A3).

A.4. Registration parameters

A.4.1. Prone-to-prone.

(FixedInternalImagePixelType ‘float’)
 (MovingInternalImagePixelType ‘float’)
 (FixedImageDimension 3)

(MovingImageDimension 3)
(UseDirectionCosines 'true')
(Registration 'MultiMetricMultiResolutionRegistration')
(Interpolator 'BSplineInterpolator')
(ResampleInterpolator 'FinalBSplineInterpolator')
(Resampler 'DefaultResampler')
(FixedImagePyramid 'FixedRecursiveImagePyramid')
(MovingImagePyramid 'MovingRecursiveImagePyramid')
(Optimizer 'AdaptiveStochasticGradientDescent')
(Transform 'BSplineTransform')
(Metric 'AdvancedNormalizedCorrelation' 'TransformBendingEnergyPenalty')
(Metric0Weight x)
(Metric1Weight y)
(FinalGridSpacingInPhysicalUnits 8)
(HowToCombineTransforms 'Compose')
(ErodeMask 'false')
(NumberOfResolutions 4)
(MaximumNumberOfIterations 200)
(NumberOfSpatialSamples 2048)
(NewSamplesEveryIteration 'true')
(ImageSampler 'Random')
(BSplineInterpolationOrder 1)
(FinalBSplineInterpolationOrder 3)
(DefaultPixelValue 0)
(WriteResultImage 'false')

A.4.2. *Prone-to-supine.*

(FixedInternalImagePixelFormat 'float')
(MovingInternalImagePixelFormat 'float')
(FixedImageDimension 3)
(MovingImageDimension 3)
(UseDirectionCosines 'true')
(Registration 'MultiMetricMultiResolutionRegistration')
(Interpolator 'BSplineInterpolator')
(ResampleInterpolator 'FinalBSplineInterpolator')
(Resampler 'DefaultResampler')
(FixedImagePyramid 'FixedRecursiveImagePyramid')
(MovingImagePyramid 'MovingRecursiveImagePyramid')
(Optimizer 'AdaptiveStochasticGradientDescent')
(Transform 'BSplineTransform')
(Metric 'AdvancedNormalizedCorrelation' 'TransformBendingEnergyPenalty'
'CorrespondingPointsEuclideanDistanceMetric')
(Metric0Weight x)
(Metric1Weight y)
(Metric2Weight z)
(FinalGridSpacingInPhysicalUnits 8)
(HowToCombineTransforms 'Compose')
(ErodeMask 'false')
(NumberOfResolutions 5)

(MaximumNumberOfIterations 200)
 (NumberOfSpatialSamples 2048)
 (NewSamplesEveryIteration 'true')
 (ImageSampler 'Random')
 (BSplineInterpolationOrder 1)
 (FinalBSplineInterpolationOrder 3)
 (DefaultPixelValue 0)
 (WriteResultImage 'false')

References

- Alderliesten T, Sonke J J and Bosman P A N 2013 Deformable image registration by multi-objective optimization using a dual-dynamic transformation model to account for large anatomical differences *Proc. SPIE Medical Imaging* ed S Ourselin and D R Haynor, vol 8669 p [866910](#)
- Andrews S, Tang L and Hamarneh G 2014 Topology preservation and anatomical feasibility in random walker image registration *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* vol 17 pp [210–17](#)
- Auger A, Bader J, Brockhoff D and Zitzler E 2009 Theory of the hypervolume indicator: optimal distributions and the choice of the reference point *Proc. ACM SIGEVO Workshop on Foundations of Genetic Algorithms* (New York: ACM) pp [87–102](#)
- Baiker M, Staring M, Löwik C W G M, Reiber J H C and Lelieveldt B P F 2011 *Automated Registration of Whole-Body Follow-Up MicroCT Data of Mice* (Berlin: Springer) (https://doi.org/10.1007%2F978-3-642-23629-7_63)
- Boehler T, Schilling K, Bick U and Hahn H K 2010 Deformable image registration of follow-up breast magnetic resonance images *Proc. Int. Workshop on Biomedical Image Registration* (Berlin: Springer) pp [13–24](#)
- Bosman P A N and Alderliesten T 2012 Incremental Gaussian model-building in multi-objective EDAs with an application to deformable image registration *Proc. Int. Conf. on Genetic and Evolutionary Computation* (New York: ACM) pp [241–48](#)
- Cahill N D, Noble J A and Hawkes D J 2009 A Demons algorithm for image registration with locally adaptive regularization *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* vol 12 pp [574–81](#)
- Das I and Dennis J E 1997 A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems *Struct. Optim.* **14** [63–9](#)
- Deb K 2001 *Multi-Objective Optimization Using Evolutionary Algorithms* (New York: Wiley)
- Deb K, Pratap A, Agarwal S and Meyarivan T 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II *IEEE Trans. Evol. Comput.* **6** [182–97](#)
- Dekker N, Ploeger L S and van Herk M 2003 Evaluation of cost functions for gray value matching of two-dimensional images in radiotherapy *Med. Phys.* **30** [778–84](#)
- Dou T H, Min Y, Neylon J, Thomas D, Kupelian P and Santhanam A P 2016 Fast simulated annealing and adaptive Monte Carlo sampling based parameter optimization for dense optical-flow deformable image registration of 4DCT lung anatomy *Proc. SPIE Medical Imaging* ed R J Webster and Z R Yaniv, vol 9786 pp [97860N–18](#)
- Eiben B et al 2016 Symmetric biomechanically guided prone-to-supine breast image registration *Ann. Biomed. Eng.* **44** [154–73](#)
- Fonseca C M, Paquete L and López-Ibáñez M 2006 An improved dimension-sweep algorithm for the hypervolume indicator *Proc. Congress on Evolutionary Computation* (Piscataway, NJ: IEEE) pp [1157–63](#)
- Foskey M, Davis B, Goyal L, Chang S, Chaney E, Strehl N, Tomei S, Rosenman J and Joshi S 2005 Large deformation three-dimensional image registration in image-guided radiation therapy *Phys. Med. Biol.* **50** [5869–92](#)
- Han L, Hipwell J H, Eiben B, Barratt D, Modat M, Ourselin S and Hawkes D J 2014 A nonlinear biomechanical model based registration method for aligning prone and supine MR breast images *IEEE Trans. Med. Imaging* **33** [682–94](#)

- Hipwell J H, Vavourakis V, Han L, Mertzaniidou T, Eiben B and Hawkes D J 2016 A review of biomechanically informed breast image registration *Phys. Med. Biol.* **61** R1–31
- Klein S, Pluim J P W, Staring M and Viergever M A 2008 Adaptive stochastic gradient descent optimisation for image registration *Int. J. Comput. Vis.* **81** 227–39
- Klein S, Staring M, Murphy K, Viergever M A and Pluim J P W 2010 Elastix: a toolbox for intensity-based medical image registration *IEEE Trans. Med. Imaging* **29** 196–205
- Lee A W, Schnabel J A, Rajagopal V, Nielsen P M and Nash M P 2010 Breast image registration by combining finite elements and free-form deformations *Int. Workshop on Digital Mammography* (Berlin: Springer) pp 736–43
- Lozano J A, Larranaga P, Inza I, Bengoetxea E (ed) 2006 *Towards A New Evolutionary Computation Studies in Fuzziness and Soft Computing* vol 192 (Berlin: Springer) (<https://doi.org/10.1007%2F11007937>)
- Mencarelli A, van Kranen S and Sonke J J 2014 Deformable image registration for adaptive radiation therapy of head and neck cancer: accuracy and precision in the presence of tumor changes *Int. J. Radiat. Oncol. Biol. Phys.* **90** 680–87
- Pirpinia K, Bosman P A N, Loo C, Scholten A, Sonke J J, van Herk M and Alderliesten T 2016a Multi-objective optimization as a novel weight-tuning strategy applied to pre-operative partial-breast radiotherapy *Proc. Int. Conf. on the Use of Computers in Radiation therapy*
- Pirpinia K, Bosman P A N, Sonke J J, van Herk M and Alderliesten T 2016b A first step toward uncovering the truth about weight tuning in deformable image registration *Proc. SPIE Medical Imaging* ed M A Styner and E D Angelini p 978445
- Rodrigues S, Bauer P and Bosman P A N 2014 A novel population-based multi-objective CMA-ES and the impact of different constraint handling techniques *Proc. Conf. on Genetic and Evolutionary Computation* (New York: ACM) pp 991–98
- Rueckert D, Sonoda L I, Hayes C, Hill D L, Leach M O and Hawkes D J 1999 Nonrigid registration using free-form deformations: application to breast MR images *IEEE Trans. Med. Imaging* **18** 712–21
- Schalj B, Kempe J A, Bauman G S, Battista J J and Van Dyk J 2004 Tracking the dose distribution in radiation therapy by accounting for variable anatomy *Phys. Med. Biol.* **49** 791–805
- Schneider C A, Rasband W S and Eliceiri K W 2012 NIH image to ImageJ: 25 years of image analysis *Nat. Methods* **9** 671–75
- Shamonin D, Bron E, Lelieveldt B, Smits M, Klein S and Staring M 2014 Fast parallel image registration on cpu and gpu for diagnostic classification of alzheimer's disease *Front. Neuroinform.* **7** 1–15
- Valsecchi A, Dubois-Lacoste J, Stutzle T, Damas S, Santamaria J and Marrakchi-Kacem L 2013 Evolutionary medical image registration using automatic parameter tuning *Proc. IEEE Congress on Evolutionary Computation* (IEEE) pp 1326–33
- van der Leij F et al 2015 First results of the preoperative accelerated partial breast irradiation (PAPBI) trial *Radiother. Oncol.* **114** 322–27
- Wahba G 1990 *Spline Models for Observational Data* vol 59 (Philadelphia, PA: Society for Industrial and Applied Mathematics)
- Yeo B T T, Sabuncu M R, Vercauteren T, Holt D J, Amunts K, Zilles K, Golland P and Fischl B 2010 Learning task-optimal registration cost functions for localizing cytoarchitecture and function in the cerebral cortex *IEEE Trans. Med. Imaging* **29** 1424–41
- Zitzler E and Thiele L 1998 Multiobjective optimization using evolutionary algorithms—a comparative case study *Int. Conf. on Parallel Problem Solving from Nature* (Springer) pp 292–301