

Simplex-based navigation tool for a posteriori selection of the preferred deformable image registration outcome from a set of trade-off solutions obtained with multiobjective optimization for the case of breast MRI

Kleopatra Pirpinia
Peter A. N. Bosman
Claudette E. Loo
Nicola S. Russell
Marcel B. van Herk
Tanja Alderliesten

Kleopatra Pirpinia, Peter A. N. Bosman, Claudette E. Loo, Nicola S. Russell, Marcel B. van Herk, Tanja Alderliesten, "Simplex-based navigation tool for a posteriori selection of the preferred deformable image registration outcome from a set of trade-off solutions obtained with multiobjective optimization for the case of breast MRI," *J. Med. Imag.* **5**(4), 045501 (2018), doi: 10.1117/1.JMI.5.4.045501.

Simplex-based navigation tool for a posteriori selection of the preferred deformable image registration outcome from a set of trade-off solutions obtained with multiobjective optimization for the case of breast MRI

Kleopatra Pirpinia,^{a,*} Peter A. N. Bosman,^b Claudette E. Loo,^c Nicola S. Russell,^a Marcel B. van Herk,^d and Tanja Alderliesten^e

^aNetherlands Cancer Institute, Department of Radiation Oncology, Amsterdam, The Netherlands

^bLife Sciences and Health Group, Centrum Wiskunde and Informatica, Amsterdam, The Netherlands

^cNetherlands Cancer Institute, Department of Radiology, Amsterdam, The Netherlands

^dUniversity of Manchester, School of Medical Sciences, Manchester Cancer Research Centre,

Manchester Academic Health Sciences Centre, Division of Cancer Science, Faculty of Biology, Medicine and Health, Manchester, United Kingdom

^eUniversity of Amsterdam, Amsterdam UMC, Department of Radiation Oncology, Amsterdam, The Netherlands

Abstract. Multiobjective optimization approaches for deformable image registration (DIR) remove the need for manual adjustment of key parameters and provide a set of solutions that represent high-quality trade-offs between objectives of interest. Choosing a desired outcome a posteriori is potentially far more insightful as differences between solutions can be immediately visualized. The purpose of this work is to investigate whether such an approach allows clinical experts to intuitively select their preferred DIR outcome. To this end, we developed a simplex-based tool for solution navigation and asked 10 clinical experts to use it to choose their preferred DIR outcome from sets of trade-off solutions obtained for 10 breast magnetic resonance DIR cases of low (prone-prone DIR; $n = 5$) and high (prone-supine DIR; $n = 5$) difficulty, of patients and volunteers, respectively. The usability of the software is subsequently evaluated by the observers using the system usability scale. Further, the quality of the selected DIR outcomes is evaluated using the mean target registration error. Results show that the users are able to identify and select high-quality DIR outcomes, and attested to high learnability and usability of our software, supporting the validity of the presumed added value of taking a multiobjective perspective on DIR in clinical practice. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/JMI.5.4.045501](https://doi.org/10.1117/JMI.5.4.045501)]

Keywords: deformable image registration; breast MRI; observer study; solution visualization; decision support software; multiobjective optimization.

Paper 18015RR received Jan. 21, 2018; accepted for publication Oct. 10, 2018; published online Oct. 30, 2018.

1 Introduction

Deformable image registration (DIR),¹ i.e., the process of deforming one image to match another image, is a medical image processing problem of potentially high impact in the field of radiation treatment.² Nonetheless, its use remains limited in clinical practice, since performing DIR using current state-of-the-art software can still be challenging. This is partially due to the large number of parameters that need to be determined separately for each DIR application, combined with the lack of insight to successfully tune such parameters, often resulting in time-consuming manual parameter adjustments, which can lead to suboptimal results.

Although typically approached as a single-objective optimization problem (e.g., Refs. 3–6), in DIR actually multiple, conflicting objectives are of interest, making a multiobjective optimization approach a much more natural fit. These objectives typically describe, e.g., the degree of similarity between the images, which needs to be high, but also the deformation needed for a good alignment, that needs to be sufficiently smooth, in

order to ensure an anatomically realistic correspondence between the images while avoiding overfitting. The concept of multiobjective optimization for DIR was introduced⁷ to remove the need to combine these objectives into one cost function and to provide a set of DIR outcomes that can be considered a set of superior choices in terms of these key objectives of interest. Using a multiobjective approach, the need for parameter tuning is removed, and a set of trade-off solutions is obtained, containing solutions representing high-quality trade-offs between the objectives of interest, and giving insight into the interplay between the objectives. From this set, a solution needs to then be manually selected. Providing this set of trade-off solutions to the user can enable insightful selection of a desired DIR outcome a posteriori, i.e., after the multiobjective DIR algorithm has been terminated, and thereby ensure that the user gets the most out of the DIR method. Recent results indicate that to solve more complex DIR problems, such as a posteriori approach is really needed as the space of parameters to be tuned by hand is too complex to navigate.⁸ It was moreover shown that by using a multiobjective approach, potentially clinically acceptable results

*Address all correspondence to: Kleopatra Pirpinia, E-mail: k.pirpinia@nki.nl

could be obtained for an easy as well as a hard breast DIR problem.⁸ The easy problem is prone-prone breast DIR, i.e., registration between images of a patient lying face down, and can be used, for example, to monitor treatment response. The harder problem is prone-supine breast DIR, where during the supine acquisition the patient is lying face up. This could be used to improve tumor localization during surgery as well as radiation treatment by translating pretreatment information to the intra-treatment setting. However, whether the clinically acceptable results can be a posteriori identified by a user, i.e., a clinical expert, by navigating the set of trade-off solutions, is not yet studied.

In this work, we investigate the usability of a navigation tool for multiobjective DIR in a clinical setting. To this end, we provided 10 clinical experts with DIR outcomes sets of trade-off solutions of five prone-prone breast DIR cases, and five prone-supine breast DIR cases, which they could navigate by using a tool that we specifically developed for this study. The set of low-difficulty cases consisted of data acquired from breast cancer patients before and after radiation treatment. The set of high-difficulty cases consisted of data acquired from healthy volunteers. We assessed the quality of the observers' preferred DIR outcomes, by calculating the mean target registration error (TRE)⁹ for each outcome based on expert-annotated anatomical landmarks. Moreover, the observers' experience with the software was assessed using the system usability scale (SUS)¹⁰ as well as a software-specific questionnaire.

2 Materials and Methods

2.1 Multiobjective Optimization

In multiobjective optimization,¹¹ we assume to have m objectives $O_i(\mathbf{x})$, $i \in \{0, 1, \dots, m-1\}$ which need to be minimized. A solution \mathbf{x}^1 is said to (Pareto) dominate a solution \mathbf{x}^2 (denoted as $\mathbf{x}^1 \succ \mathbf{x}^2$) if and only if $O_i(\mathbf{x}^1) \leq O_i(\mathbf{x}^2)$ holds for all $i \in \{0, 1, \dots, m-1\}$ and $O_i(\mathbf{x}^1) < O_i(\mathbf{x}^2)$ holds for at least one $i \in \{0, 1, \dots, m-1\}$. A nondominated 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 nondominated front corresponding to a non-dominated set is the set of all m -dimensional objective function values corresponding to the solutions, i.e., the set of all $\mathbf{O}(\mathbf{x}^j) = [O_0(\mathbf{x}^j), O_1(\mathbf{x}^j), \dots, O_{m-1}(\mathbf{x}^j)]$, $j \in \{0, 1, \dots, n-1\}$. A solution \mathbf{x}^1 is said to be Pareto optimal if and only if there is no other \mathbf{x}^2 such that $\mathbf{x}^2 \succ \mathbf{x}^1$ holds. Further, the Pareto set is the set of all Pareto-optimal solutions and the Pareto front is the set of objective function values that corresponds to the Pareto set. As for real-world problems the Pareto front is typically unknown, the set of solutions obtained by a multiobjective optimization algorithm is a nondominated front, or, equivalently, a set of trade-off solutions (that do not dominate each other in objective space) that approximates the Pareto front.

To solve multiobjective optimization problems, evolutionary algorithms (EAs)¹² are frequently adopted. Their capacity to approximate the Pareto front in one run by evolving a population (i.e., a set) of solutions simultaneously as well as their good performance on benchmark and real-world problems make them state of the art in multiobjective optimization. In this work, we used EAs to solve two multiobjective optimization problems. First, we employed an EA to optimize the parameters of a single-objective DIR method, providing us with nondominated

fronts of DIR outcomes that exhibit the best trade-offs in terms of key objectives of interest. Second, an EA was employed to solve a multiobjective optimization problem that enables the intuitive visualization and navigation of the aforementioned nondominated fronts. More details follow in Secs. 2.2 and 2.3, respectively.

2.2 Multiobjective Deformable Image Registration

In DIR, the aim is to find the optimal transformation that deforms the so-called source image to the target image. To do so, current state-of-the-art DIR methods optimize a cost function that consists of a linear combination of terms that describe objectives of interest, where the weights in this linear combination need to be determined beforehand. By formulating DIR as a multiobjective problem, we aim to optimize these objectives separately and simultaneously. Specifically, in previous work,⁸ we used an EA to find the weights that result in nondominated objective values when used within a well-known single-objective registration method called elastix.⁴ The EA we employed is an estimation-of-distribution algorithm called iMAMaLGaM (incremental multiobjective adapted maximum-likelihood Gaussian model).¹³ Such EAs evolve and generate solutions by estimating a probability distribution from a selection of high-quality solutions in the population and subsequently sampling the estimated distribution to generate new solutions. In iMAMaLGaM, the selected solutions are grouped into clusters in the objective space, because it is known that clustering can play an important role in dealing with the complexity of multi-objective optimization problems.¹⁴ For each cluster, an l -dimensional (where l is the number of parameters of the optimization problem) Gaussian distribution is estimated. Subsequently, iMAMaLGaM samples the Gaussian distributions to generate new solutions.

In this metaoptimization process, a set of candidate weight vectors (m weights for m objectives) is first generated by iMAMaLGaM, which is then passed to elastix, which performs single-objective DIR with each weight vector, and calculates the objective values. The objective values are then passed back to iMAMaLGaM, which then subsequently generates new candidate weight vectors. As within elastix we used linear combinations of $m = 2$ and $m = 3$ objectives, we obtained two- and three-dimensional (2-D and 3-D) nondominated fronts of DIR outcomes, for the low- and high-difficulty DIR cases, respectively. For more details on this combination of algorithms, see Ref. 8. There, we showed that to get the most out of DIR software for hard DIR cases, manual tuning is not sufficient, and metaoptimization is necessary.

2.3 Solution Exploration for Deformable Image Registration

For a 2-D nondominated front, a simple slider was used. However, it was essential to develop an intuitive visualization tool for 3-D nondominated front exploration. To this end, we used trade-off sliders combined with a 2-D unit simplex. However, large differences between the objectives of DIR in scale and optimization difficulty, especially in the 3-D case, resulted in irregular nondominated fronts, which, when mapped to the unit simplex, resulted in highly nonuniform distributions across it. An example of such a 3-D DIR non-dominated front can be seen in Fig. 1. Navigating 2-D and 3-D fronts with typical tools such as trade-off sliders (which are already used in clinical

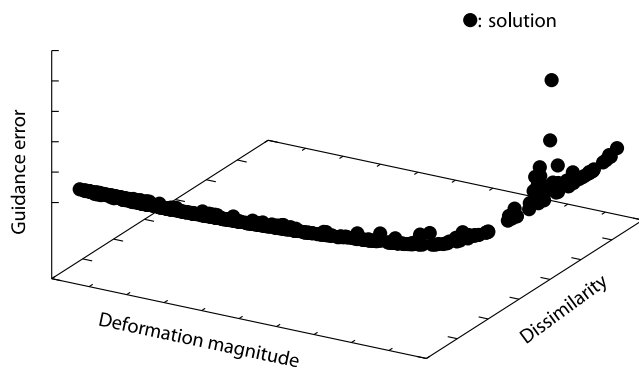


Fig. 1 3-D nondominated front for a DIR problem with three objectives.

practice in the context of multiobjective optimization for radiation treatment planning¹⁵) can as a consequence be hard and unintuitive. The key reason for this is that a straightforward use of sliders on the directly mapped data makes it virtually impossible to identify individual solutions that are very close to each other but are potentially different and potentially clinically interesting. Moreover, large parts of the slider space would map to empty space, where there are no solutions in the nondominated front. Whereas spreading the solutions in a uniform manner in the 2-D case is straightforward (by taking the minimum value (min) and maximum value (max) of one objective and redistributing uniformly the solutions in [min, max]), this is not the case for a 3-D front. Therefore, to enable easier solution exploration for the 3-D case, we mapped the nondominated front surface to the 2-D unit simplex in a way that achieved a more uniform distribution, i.e., we aimed to maximize the smallest distance between any two pairs of points. The definition of this objective can be found in Ref. 16. However, preserving to a large extent the topology of the original nondominated front (i.e., the relative pairwise distances between the points in the original front) is also desirable (exact definition of this objective can be found in Ref. 16), which conflicts with perfectly uniformly spreading the solutions across the simplex. This too poses a multiobjective optimization problem, which we solved using an EA known as the multiobjective real-valued gene-pool optimal mixing evolutionary algorithm (MO-RV-GOMEA),¹⁷ because this algorithm allowed us to quickly and reliably achieve desirable remappings. MO-RV-GOMEA for nondominated front exploration was introduced in Ref. 16, where more details about the formulation of this optimization problem and the algorithm employed to solve it can be found. MO-RV-GOMEA takes as input the nondominated front objective values normalized to [0,1]. The final solution (i.e., the configuration of the unit simplex) was chosen manually from the nondominated front of simplex configurations, to achieve a good spread that allows sensible use of sliders, while maintaining the topology of the original front sufficiently. The coordinates of the solutions in the unit simplex were used (through transformation into barycentric coordinates) as input for a set of trade-off sliders, one for each objective. The unit simplex served also as an intuitive visualization of the 3-D trade-off front. The corners of the unit simplex were the solutions that were the best in terms of each of the three objectives. An example of such a mapped nondominated front can be seen in Fig. 2.

2.4 Datasets

2.4.1 Prone-prone

Five nondominated fronts were randomly selected from a set of 10 nondominated fronts resulting from DIR (performed in a previous study, see details in Ref. 8) of breast magnetic resonance imaging (MRI) scans of breast cancer patients acquired in prone position before and after radiation treatment. Approval was obtained from the institutional review boards for the data used in the study. For this DIR problem, within elastix, two objectives were optimized (more specifically, minimized): the dissimilarity, based on the negative normalized correlation coefficient,¹⁸ and the deformation magnitude, based on the bending energy penalty,¹⁹ resulting in five 2-D fronts.

2.4.2 Prone-supine

Five nondominated fronts resulting by DIR of breast MRI scans of healthy volunteers acquired during the same scan session in prone and in supine position were used. Also here, 10 nondominated fronts were obtained in a previous study;⁸ however, as in a subset of cases it was not possible to obtain meaningful results (based on the mean TRE values), the five fronts representing the most successful DIR cases (on the basis of minimum mean TRE, the range of which for these cases was 2.9 to 5.8 mm) were selected. This DIR problem is very hard to solve, because of the large breast deformation occurring between the two positions. For this reason, an additional objective was added for minimization, which exploits guidance information present in the images, i.e., the presence of eight external MRI-visible markers attached to each breast of the volunteers. For these DIR problems, there were three objectives that needed to be minimized, dissimilarity and deformation magnitude as above, but also the guidance error, described by the mean Euclidean distance between the location of the external markers in the transformed source image and their location in the target image. Solving this optimization problem resulted in 3-D nondominated fronts.

2.5 User Interface for Solution Navigation

For the prone-prone solution navigation, the user interface consisted of two trade-off sliders, one for each objective, and in-house-developed image visualization software. The names of the objectives were formulated in a way that implied that they needed to be maximized, for a more user-friendly experience: within the optimization algorithm, the dissimilarity and the deformation magnitude were minimized, whereas in the interface the similarity and the smoothness, respectively, needed to be maximized. The observers could use the trade-off sliders to select and visualize different solutions. Moreover, the observers could choose among four visualization modes: visualizing the target image, the transformed source image, a checkerboard overlay of target and transformed source image, or a green/magenta overlay of the target and the transformed source image. Further, a deformed grid could be overlaid on each of the different visualization modes.

For the prone-supine solution navigation, the user interface consisted of a set of three trade-off sliders, the in-house-developed image visualization software, and the unit simplex. Similarly to the other two objectives, within the optimization the guidance error was to be minimized, whereas in the interface the marker match was to be maximized. The observers could use the trade-off sliders to inspect solutions, while at the same time

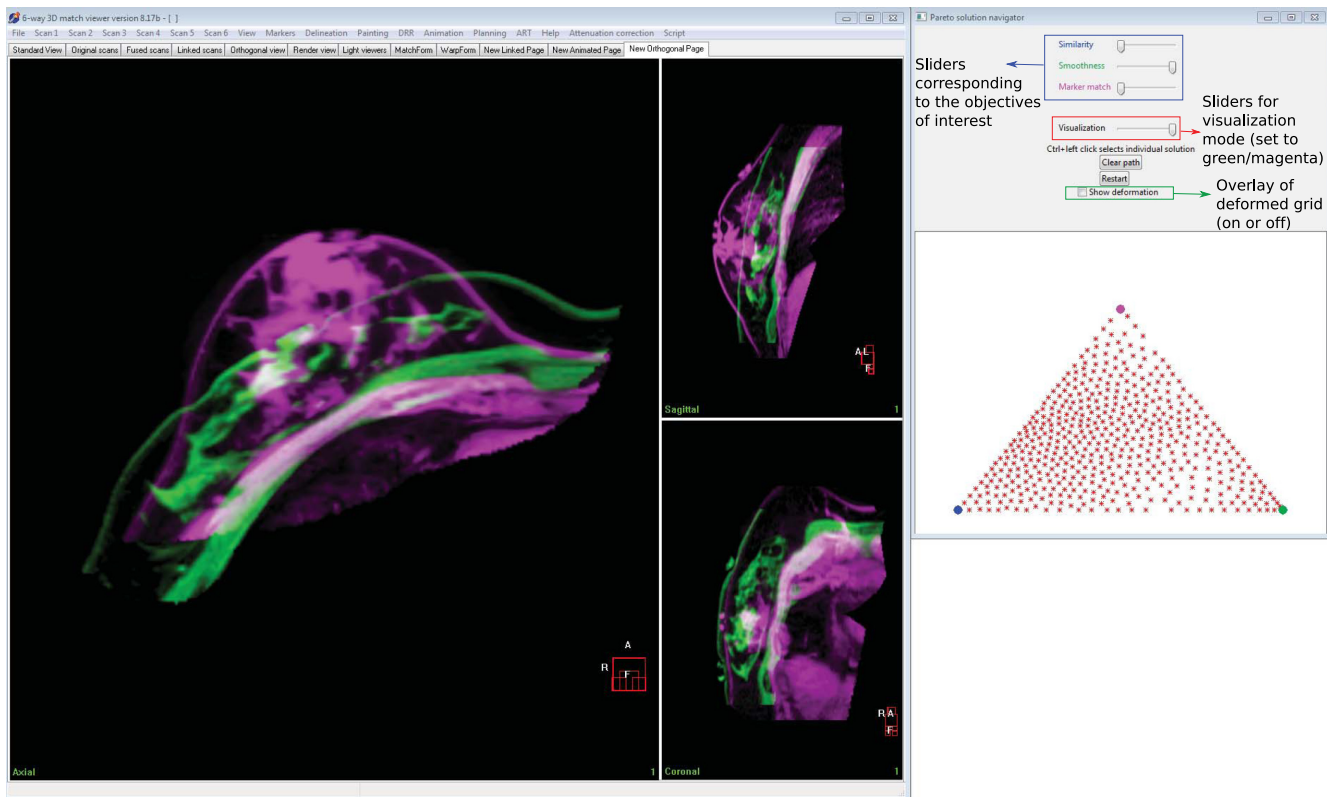


Fig. 2 The user interface for prone-supine solution navigation. Target image is shown in green, transformed source image is shown in magenta. The overlay looks gray wherever the two images have similar intensity.

observing the location of these solutions in the unit simplex. Another option available to the observers was the possibility to select individual solutions by clicking on them in the unit simplex. The observers could visualize the path of the already inspected solutions, and clear it whenever desired.

The starting point of the solution navigation for each case was a solution with a very low amount of deformation. The user interface for the prone-supine solution navigation can be seen in Fig. 2. Every solution that was selected for inspection by the observers, either by sliding or clicking on the unit simplex, was saved.

2.6 Observers

Five radiation oncologists/physician assistants specialized in breast cancer (group 1) and five experienced breast radiologists (group 2) were asked to navigate the nondominated fronts and select their preferred DIR outcome, basing their selection on the alignment of structures within the breast. Group 1 was familiar with the in-house developed software for image visualization. Prior to the start of the solution navigation session, each observer was given a short tutorial on the study and the user interface, followed by testing the software on a prone-supine DIR case not included in the study. The session was also audio recorded and timed. At the end of the session, the observers were asked to complete two questionnaires.

2.7 Usability Evaluation

To assess the observers' perceived usability of the solution navigation tool, we used the SUS,¹⁰ which consists of 10 questions,

each with five response options on a linear scale, ranging from 1 = strongly disagree to 5 = strongly agree. The maximum obtainable score for this questionnaire, which would indicate the perfect system, is 100. An SUS score above 68 would be considered above average. To gain further insight into the SUS scores, SUS scores were mapped to adjective ratings, according to which a mean SUS score above 70 indicates an acceptable or good system, whereas a mean SUS score of 85 or above indicates an excellent system.²⁰

Further, a second set of questions was designed according to guidelines^{21,22} in order to extract more software-specific information about the observers' experience. It consisted of six questions with five response options on a linear scale similar to the SUS, and three questions that varied in format: the observers were asked whether they preferred to use only the trade-off sliders, only the unit simplex, or if they preferred to use both. Further, they were asked whether in their opinion the DIR outcomes that were provided were too many, too few, or of the appropriate amount. Lastly, they were asked an open-ended question on whether they had any suggestions that could improve the navigation tool.

2.8 Selected Solution Evaluation

To quantify the quality of the registration outcomes, and also to investigate the variation in selected outcome quality between the observers, we calculated the mean TRE for each solution as follows. The Euclidean distance between the locations of each one of eight expert-defined internal anatomical landmarks in the transformed source image and in the target image was calculated, and then the average distance was calculated. We consider

a solution with a low mean TRE to be a solution of high quality. In previous work,⁸ the interobserver variability (based on two observers) was shown to be ~ 1.0 mm. To test if there were significant differences in performance between the two groups of observers, based on the quality of solutions that they selected, we used multiple Wilcoxon signed-rank tests, with the Bonferroni correction to account for multiple comparisons ($n = 10$). Following this, the significance level becomes $p = 0.05/10 = 0.005$.

3 Results

The overall usability of the software was rated highly, with a mean SUS score of 87 over all 10 observers, and therefore in the “excellent” range (see Table 1). Multiple observers remarked that the multiobjective solution navigation system was very easy to learn and intuitive to use, which is reflected in questions 3 and

Table 1 SUS scores for the two observer groups.

	1	2	3	4	5	Mean
Group 1	92.5	87.5	92.5	92.5	82.5	89.5
Group 2	77.5	82.5	95.0	100.0	67.5	84.5
All observers						87.0

Table 2 Mean (standard deviation) of responses to questionnaire 1. Response options range from 1 = strongly disagree, 2 = disagree, 3 = neither agree or disagree, 4 = agree, and 5 = strongly agree.

Questionnaire 1 (SUS)	Group 1	Group 2	All
1. I think I would like to use this system frequently	4.4 (0.5)	4.2 (1.1)	4.3 (0.8)
2. I found the system unnecessarily complex	1.4 (0.5)	1.4 (0.5)	1.4 (0.5)
3. I thought the system was easy to use	5.0 (0.0)	4.6 (0.5)	4.8 (0.3)
4. I think that I would need the support of a technical person to be able to use this system	2.8 (1.5)	2.2 (1.6)	2.5 (1.6)
5. I found the various functions in this system were well integrated	4.4 (0.5)	4.4 (0.9)	4.4 (0.7)
6. I thought there was too much inconsistency in this system	1.0 (0.0)	1.8 (0.8)	1.4 (0.4)
7. I would imagine that most people would learn to use this system very quickly	4.8 (0.4)	4.8 (0.4)	4.8 (0.4)
8. I found the system very cumbersome to use	1.2 (0.4)	1.2 (0.4)	1.2 (0.4)
9. I felt very confident using the system	4.4 (0.9)	4.2 (0.8)	4.3 (0.9)
10. I needed to learn a lot of things before I could get going with this system	1.0 (0.0)	1.4 (0.5)	1.2 (0.3)

7 of the SUS (Table 2). Further, they felt confident using the system (Table 2, question 9). They became more comfortable using the software during the study (Table 3, question 2). The observers were also quite satisfied with their registration outcomes, slightly more for the prone-prone than for the prone-supine DIR (questions 5 and 6 of Table 3). Five observers found that they would rather use both the sliders as well as the simplex for navigation, whereas the remaining five would rather use only the sliders for the prone-supine DIR cases. Three observers used exclusively the sliders to select solutions, whereas observer 4 of group 1 used almost exclusively the unit simplex to locate their preferred solution for test cases 3, 4, and 5. The observer that gave the lowest grade on the solution navigation tool as described by the SUS score (67.5) gave a 2 and 3, respectively, to questions 5 and 6 of questionnaire 2, the lowest scores observed for these questions. Regarding the prone-prone DIR cases, this observer felt that the outcomes, given the easier

Table 3 Mean (standard deviation) of responses to subset of questions of questionnaire 2. Response options range from 1 = strongly disagree, 2 = disagree, 3 = neither agree or disagree, 4 = agree, and 5 = strongly agree. In question number 7, the percentage of every group that gave each response is reported.

Questionnaire 2	Group 1	Group 2	All
1. I fully understood how to use the software prior to the start of the study	4.0 (1.0)	4.4 (0.5)	4.2 (0.8)
2. During the study, I became more comfortable using the software	5.0 (0.0)	4.8 (0.4)	4.9 (0.2)
3. The user interface for prone-prone solution navigation was easy to use	4.8 (0.4)	4.8 (0.4)	4.8 (0.4)
4. The user interface for prone-supine solution navigation was easy to use	4.6 (0.9)	4.4 (0.9)	4.5 (0.9)
5. I am satisfied with the clinical quality of my selected prone-prone DIR outcomes	4.6 (0.5)	4.2 (1.3)	4.4 (0.9)
6. I am satisfied with the clinical quality of my selected prone-supine DIR outcomes	4.2 (0.8)	3.8 (0.8)	4.0 (0.8)
7. I would rather use			
a. both the sliders and the simplex	60%	40%	50%
b. only the sliders	40%	60%	50%
c. only the simplex	0%	0%	0%

Table 4 Mean (standard deviation) length of the solution navigation session in minutes.

	Prone-prone	Prone-supine
Group 1	13.8 (4.7)	22.1 (10.6)
Group 2	10.5 (0.5)	18.5 (7.2)

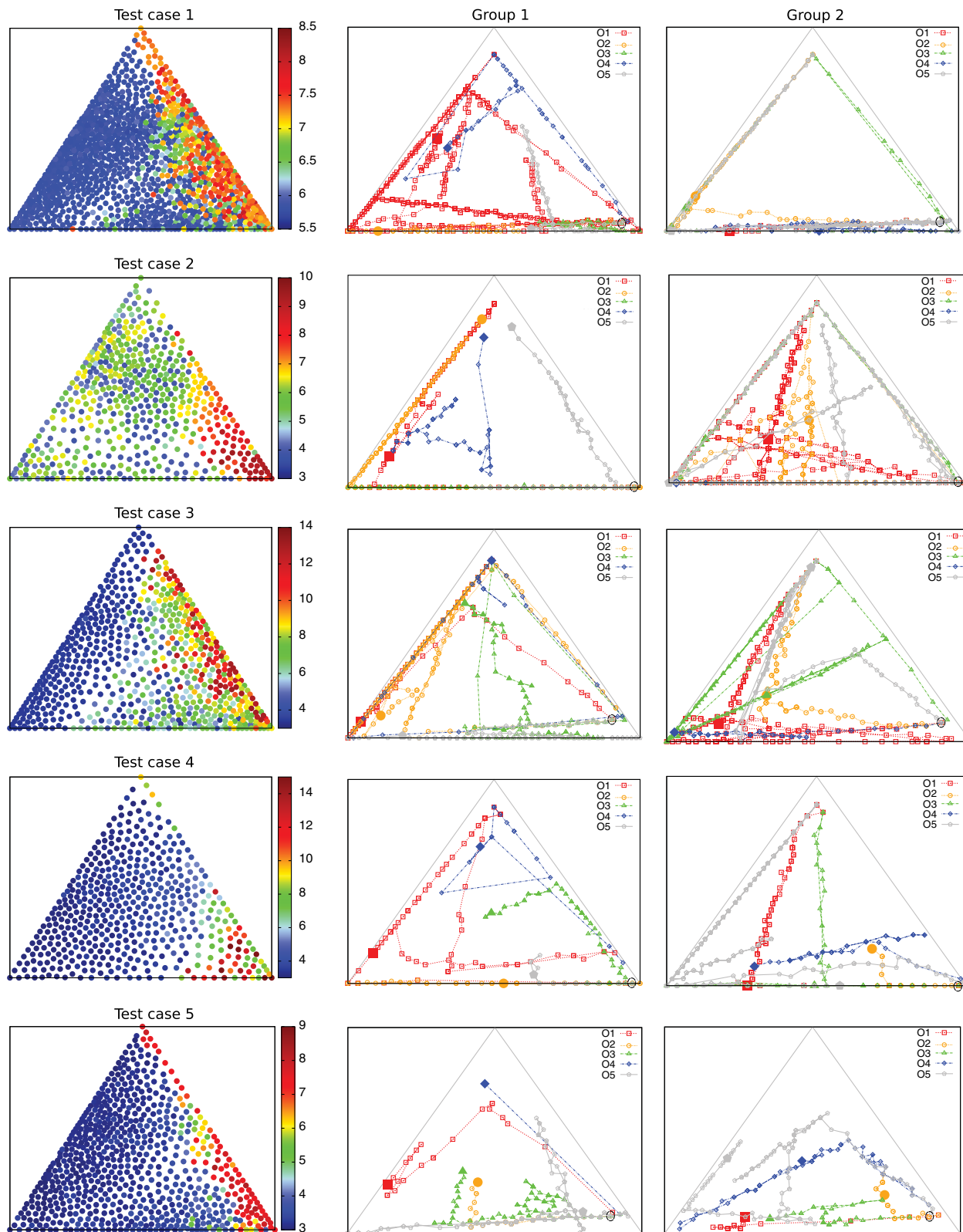


Fig. 3 Solution navigation of the prone-supine test cases for the two observer groups. The unit simplex is color-coded with the mean TRE in mm. As the mean TRE distribution can vary between cases, the color-bar scales are different to better illustrate the complexity of each DIR case. The filled points represent the final selected solution by each observer. Round nonfilled point represents starting point for each case. The corner points of the simplex are the solutions that score best in each of the three objectives (similarity in bottom left, smoothness of deformation in bottom right, and marker match in top corner), and their values are not the same for all cases, neither in weights nor objective values.

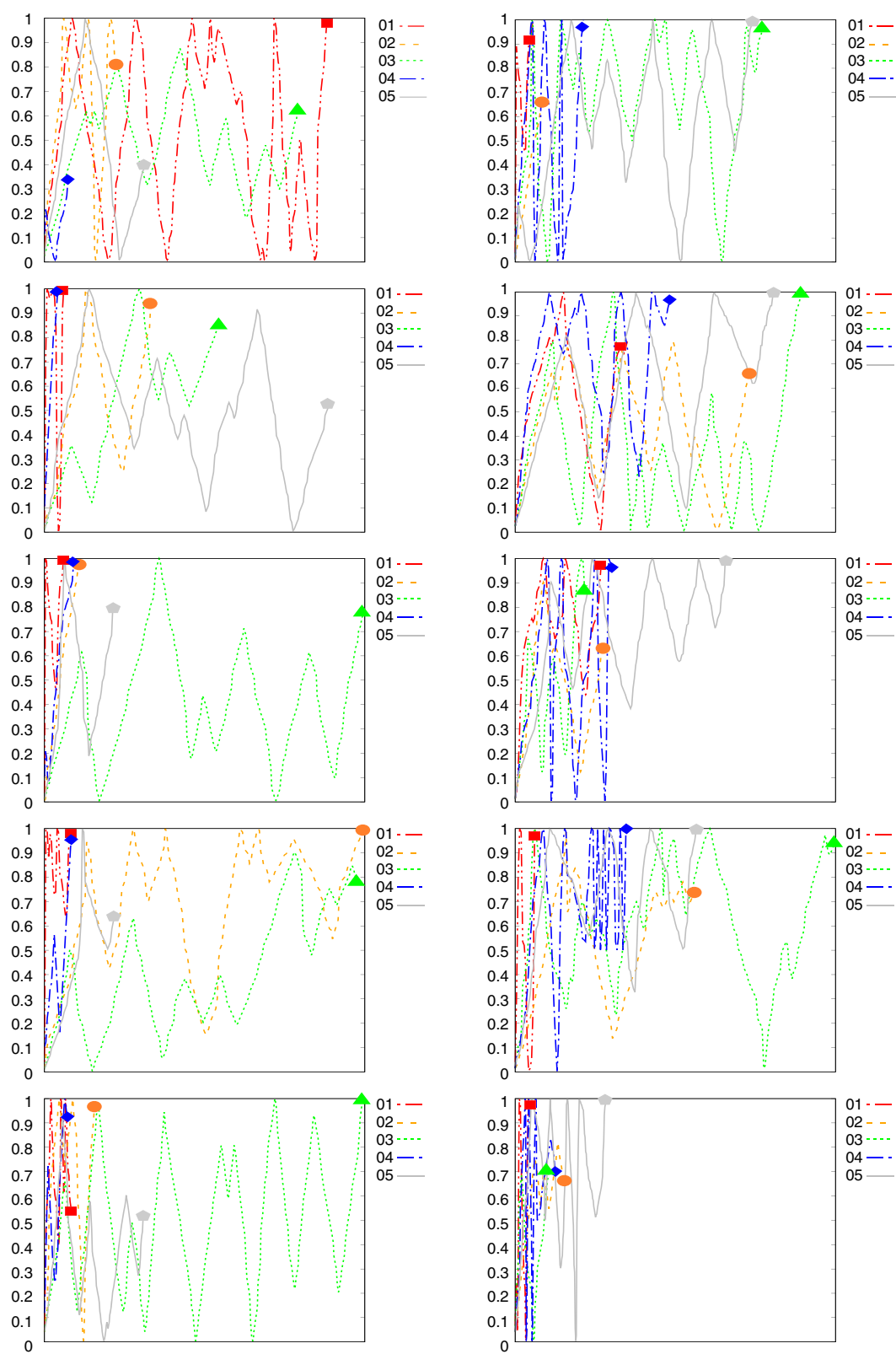


Fig. 4 From left to right: solution navigation for prone-prone cases for group 1 and group 2. From top to bottom: prone-prone cases 1, 2, 3, 4, and 5. The x-axis represents the length of the solution navigation session, and it has been normalized per case, in order to better compare the two groups. The y-axis is the position of the slider describing the similarity. The filled points are the final selected solutions of each observer.

nature of the DIR task, were not good enough. The dissatisfaction of this observer with the DIR outcomes may be related to the low SUS score. Observer 5 of group 1 indicated in the SUS that they did not feel very confident using the software, and in retrospect this observer appeared to have selected outcomes with the largest mean TRE on average compared with the other observers. This observer did not use at all the unit simplex. An observer suggested incorporating a function that allows saving temporarily a DIR outcome, to which they could go back. Total navigation time for the full session that included all 10 test cases varied from 19 to 56 minutes. Navigation time was shorter for group 2 for both prone-prone and prone-supine cases compared with group 1. In both cases, prone-supine navigation time was longer (see Table 4). The navigation process of each observer for the prone-supine test cases can be seen in Fig. 3. Test case 2 along with test case 3 were considered to be the most challenging, as reflected in the solution navigation process, where a large number of solutions were selected for inspection by the observers before making their final selection (Fig. 3). The navigation process for the prone-prone cases can be seen in Fig. 4. It can be seen that for the prone-prone cases,

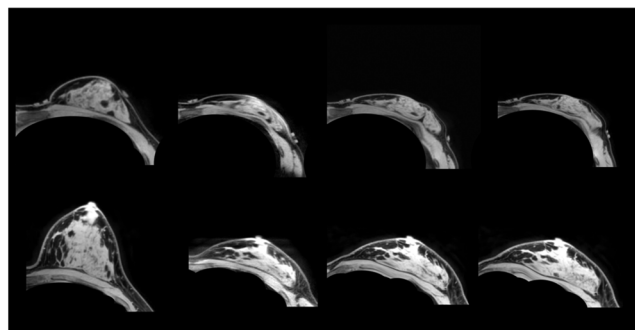


Fig. 6 From left to right: source image, target image, selected DIR outcome with lowest mean TRE, selected DIR outcome with largest mean TRE for prone-supine test case 2 (upper row), and prone-supine test case 3 (lower row).

multiple observers explored the entire set of solutions before making their final selection, and often (as opposed to the prone-supine cases) they selected a solution with, or very close to, the maximum similarity. The largest variation in quality of selected solutions was found for test case 2 of the prone-supine

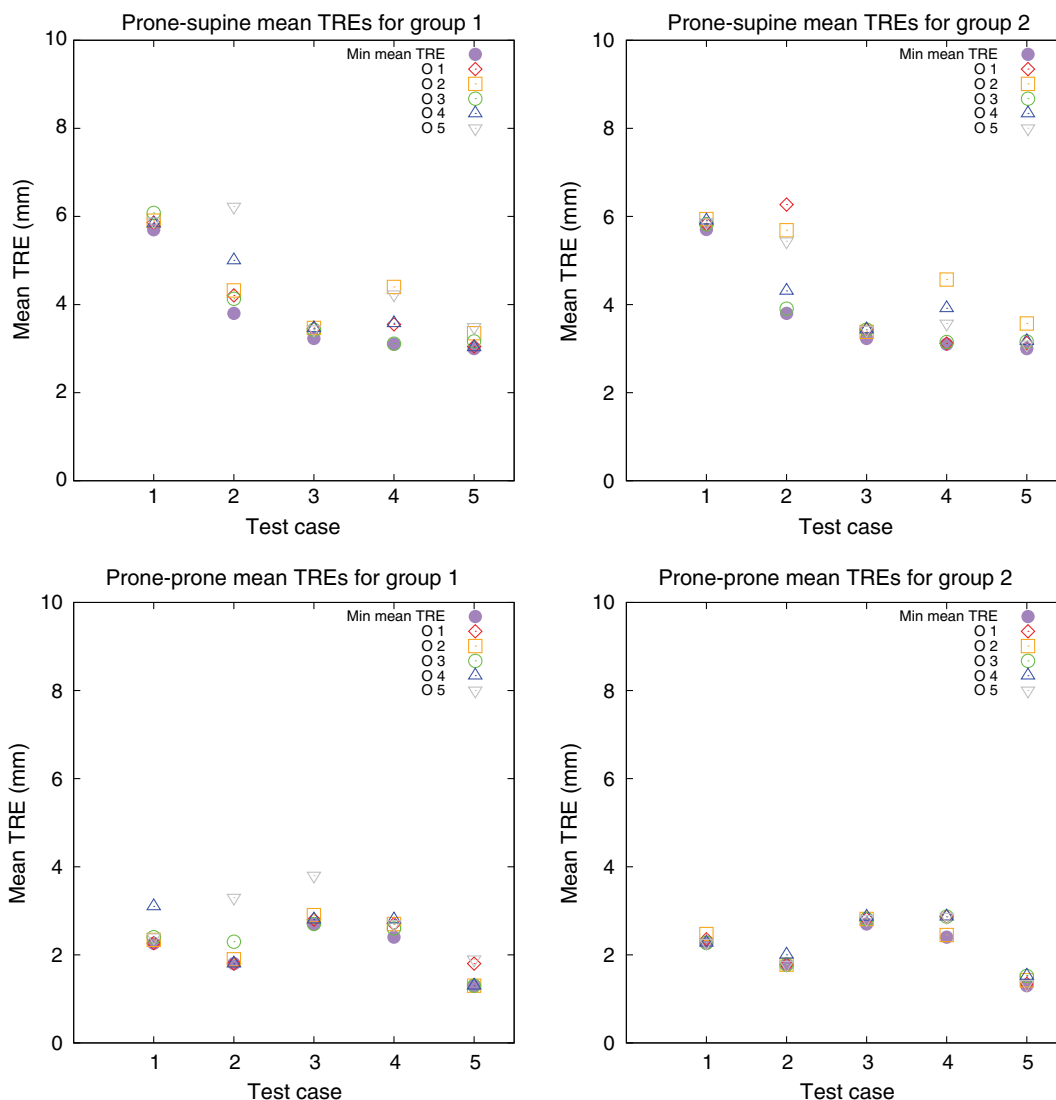


Fig. 5 Mean TRE of preferred solutions of observers along with solution with minimum mean TRE for prone-supine DIR (upper row) and prone-prone DIR (lower row) cases for group 1 (left) and group 2 (right).

Table 5 p-values testing the difference in the performance in terms of mean TRE between groups 1 and 2 for cases 1–5 for prone-prone DIR and prone-supine DIR.

Case	Prone-supine	Prone-prone
1	0.625	1.000
2	0.312	0.812
3	1.000	0.125
4	0.437	0.812
5	0.750	0.812

dataset (Fig. 5). For this test case, misalignment can be observed in the DIR outcomes at the outer side of the breast in supine position, due to image intensity inhomogeneities in the supine image (see upper row of Fig. 6). For the rest of the prone-supine cases, as well as the prone-prone cases, the observers selected solutions with a mean TRE close to the minimum mean TRE (Fig. 5). There were no significant differences in the performance of the two groups (see Table 5). The selected solutions with the lowest mean TRE as well as those with the largest mean TRE for test cases 2 and 3 can be seen in Fig. 6.

4 Discussion

In this work, we presented a simplex-based navigation tool for a posteriori selection of the preferred DIR outcome from a set of trade-off solutions, with an application to breast MRI. It is the first time, to the authors' knowledge, that a multiobjective optimization framework developed for DIR is evaluated using a specially designed user interface. The results indicate that the combination of this framework with the simplex-based navigation tool can be used in clinical practice to find the preferred registration outcome for multiobjective DIR, as the majority of the observers were able to select DIR outcomes with which they were satisfied, and had a positive perception of its usability.

The assessability of a DIR outcome based on visual inspection alone can be considered inadequate in some cases, as good alignment may have been achieved with an incorrect deformation. In this study, however, visual assessment is complemented with the knowledge of the interplay between objectives. In particular, using the sliders and seeing the differences between DIR outcomes while navigating gives insight into the amount of deformation occurring with respect to the image similarity, making DIR outcome selection much more insightful.

In this study, we investigated the feasibility of DIR solution navigation with two and three objectives, but the framework can accommodate more objectives. However, in those cases, only the slider feature can be used, as the simplex visualization for more than four dimensions would become complex or impossible.

One of the limitations of this study is the lack of features in the images that can be of clinical interest (e.g., tumor presence in the case of the prone-supine DIR problem), which would have made the solution selection criteria more specific, and thereby possibly reduced the variation between the selected solutions, in case where the DIR outcome was not perfect, such as in prone-supine cases 2 and 3, respectively. Further, the limited use of DIR in everyday clinical practice made the selection of the

appropriate observer group challenging. For this reason, two observer groups were selected: the radiologists, since they are experts on breast MRI, and the radiation oncologists, since they are familiar with (mostly rigid) registration approaches.

A limitation of the method is that there is yet no automatic way to select a simplex configuration from a nondominated front of such configurations as this was done manually for this study by a multiobjective optimization expert. Based on this pilot study, however, we observed that all the selected simplex configurations have similar objective value for one of the two objectives (and more specifically the objective related to the uniformity of the spread of the solutions in the simplex), and are close to the knee of the nondominated front. It may, therefore, be well possible that we could automatically derive a solution using this information.

Multiple observers remarked that although they did select a final solution, an entire region of the simplex close to that solution contained acceptable registration outcomes. This is valuable information, as identification of clinically interesting regions of the nondominated front could be used as *a priori* information for the optimization algorithm that approximates the Pareto front of the DIR problem, improving its efficiency and performance. Further, it may be possible to derive a range of parameter configurations that yield solutions in the clinically interesting region of the nondominated front for any image pair, making the use of DIR more efficient. We also noted that, although group 1 was more familiar with the in-house developed software for image visualization, there were no significant differences in their performance compared with the radiologists' group, as both groups found high-quality DIR outcomes in terms of mean TRE and rated the software highly. This indicates that the tool is learnable by people with different backgrounds and training. Moreover, the short solution selection time (on average 3 min per test case) allows for possible incorporation of the process in clinical practice, although in the case of a larger number of objectives this time is expected to increase. The high refresh rate of the software allows almost real-time visualization of a high number of solutions in this short time frame. The meta-optimization procedure is the most computationally expensive, because of the high number of DIRs to be performed, but this can be mitigated using DIR software that can run on GPUs.²³ The EAs are easy to parallelize, and MO-RV-GOMEA in particular has already been implemented on a GPU.²⁴ The EAs as well as elastix are open-source.

The use of single-objective DIR with manually determined parameters was shown to not be sufficiently robust for complex DIR problems such as prone-supine DIR.⁸ Therefore, a patient-specific multiobjective approach may be more appropriate as it ensures that the selected outcome is clinically acceptable, without requiring cumbersome parameter adaptations and/or rerunning DIR software, but rather using only a transparent solution navigation and selection tool.

Further, care should be taken when evaluating the quality of a DIR outcome solely based on the mean TRE. The approach illustrates the high subjectivity in the assessment of the quality of the DIR outcome. Even for the less-challenging cases of prone-prone DIR, where the mean TRE remained relatively low, there were highly variable responses with regard to the satisfaction with the clinical quality of the DIR outcome.

This work further illustrates that some DIR cases are inherently hard and sometimes very challenging to be solved to

clinical satisfaction, regardless of the multiobjective automated tuning approach used in this study, which ultimately remains dependent on the underlying single-objective DIR software (albeit getting the most out of it). This is part of the reason why some observers were not 100% satisfied. A purely multi-objective DIR algorithm or improvements to the existing single-objective DIR software could overcome this. Moreover, the tool presented in this article could in that case still be used directly, and the results are expected only to improve.

Disclosures

The authors have no conflicts of interest to disclose.

Acknowledgments

The authors would like to thank P. A. Bouter from Centrum Wiskunde & Informatica for the unit simplex configurations, N. N. Y. Janssen for recommending the SUS, as well as A. N. Scholten, A. O. J. Vreeswijk, C. Veenstra, P. H. M. Elkhuisen, P. K. de Koekoek-Doll, C. A. H. Lange, G. A. O. Winter-Warnars and E. G. Klompenhouwer from the Netherlands Cancer Institute for their participation in the study. This project is funded by the Dutch Cancer Society (KWF; Grant No. KWF 2012-5716).

References

1. A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: a survey," *IEEE Trans. Med. Imaging* **32**, 1153–1190 (2013).
2. H. Wang et al., "Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.* **61**, 725–735 (2005).
3. T. Vercauteren et al., "Diffeomorphic demons: efficient non-parametric image registration," *NeuroImage* **45**(1), S61–S72 (2009).
4. S. Klein et al., "Elastix: a toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging* **29**, 196–205 (2010).
5. G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinematics," *IEEE Trans. Image Process.* **5**(10), 1435–1447 (1996).
6. L. Han et al., "Development of patient-specific biomechanical models for predicting large breast deformation," *Phys. Med. Biol.* **57**(2), 455–472 (2012).
7. T. Alderliesten, J.-J. Sonke, and P. A. N. Bosman, "Deformable image registration by multi-objective optimization using a dual-dynamic transformation model to account for large anatomical differences," *Proc. SPIE* **8669**, 866910 (2013).
8. K. Pirpinia et al., "The feasibility of manual parameter tuning for deformable breast MR image registration from a multi-objective optimization perspective," *Phys. Med. Biol.* **62**, 5723–5743 (2017).
9. J. Fitzpatrick and J. West, "The distribution of target registration error in rigid-body point-based registration," *IEEE Trans. Med. Imaging* **20**, 917–927 (2001).
10. J. Brooke et al., "SUS-a quick and dirty usability scale," *Usability Eval. Ind.* **189**(194), 4–7 (1996).
11. J. Branke, K. Deb, and K. Miettinen, *Multiobjective Optimization: Interactive and Evolutionary Approaches*, vol. **5252**, Springer Science & Business Media, Berlin, Heidelberg (2008).
12. K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, New York (2001).
13. S. Rodrigues, P. Bauer, and P. A. N. Bosman, "A novel population-based multi-objective CMA-ES and the impact of different constraint handling techniques," in *Proc. of the Annual Conf. on Genetic and Evolutionary Computation GECCO '14*, ACM, New York, pp. 991–998 (2014).
14. M. Pelikan, K. Sastry, and D. E. Goldberg, "Multiobjective hBOA, clustering, and scalability," in *Proc. of the 7th Annual Conf. on Genetic and Evolutionary Computation, GECCO '05*, ACM, New York, pp. 663–670 (2005).
15. D. L. Craft et al., "Improved planning time and plan quality through multicriteria optimization for intensity-modulated radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.* **82**, e83–e90 (2012).
16. A. Bouter et al., "Spatial redistribution of irregularly-spaced pareto fronts for more intuitive navigation and solution selection," in *Proc. Annual Conf. on Genetic and Evolutionary Computation, GECCO '17*, ACM Press, New York, pp. 1697–1704 (2017).
17. A. Bouter et al., "The multi-objective real-valued gene-pool optimal mixing evolutionary algorithm," in *Proc. of Annual Conf. on Genetic and Evolutionary Computation, GECCO '17*, ACM Press, New York, pp. 537–544 (2017).
18. K. Briechele and U. D. Hanebeck, "Template matching using fast normalized cross correlation," *Proc. SPIE* **4387**, 95–102 (2001).
19. G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia (1990).
20. A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: adding an adjective rating scale," *J. Usability Stud.* **4**(3), 114–123 (2009).
21. T. F. Burgess, "Guide to the design of questionnaires: a general introduction to the design of questionnaires for survey research," *Information System Services, University of Leeds* 2001, <https://nats-www.informatik.uni-hamburg.de/pub/User/InterculturalCommunication/top2.pdf>
22. W. C. Leung, "How to design a questionnaire," *Student BMJ* **9**, 187–189 (2001).
23. P. Bhosale et al., "GPU-based stochastic-gradient optimization for non-rigid medical image registration in time-critical applications," *Proc. SPIE* **10574**, 105740R (2018).
24. A. Bouter et al., "Large-Scale parallelization of partial evaluations in evolutionary algorithms for real-world problems (to appear)," in *Proc. of Annual Conf. on Genetic and Evolutionary Computation, GECCO '18*, ACM Press, New York (2018).

Biographies for the authors are not available.