

On the usefulness of gradient information in multi-objective deformable image registration using a B-spline-based dual-dynamic transformation model: comparison of three optimization algorithms

Kleopatra Pirpinia^a, Peter A. N. Bosman^b, Jan-Jakob Sonke^a, Marcel van Herk^a, Tanja Alderliesten^c

^aDept. of Radiation Oncology, The Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital (NKI-AVL), P.O. Box 90203, 1006 BE Amsterdam, The Netherlands;

^bCentrum Wiskunde & Informatica (CWI), P.O. Box 94079, 1090 GB Amsterdam, The Netherlands;

^cDept. of Radiation Oncology, Academic Medical Center (AMC), P.O. Box 22660, 1100 DD Amsterdam, The Netherlands

ABSTRACT

The use of gradient information is well-known to be highly useful in single-objective optimization-based image registration methods. However, its usefulness has not yet been investigated for deformable image registration from a multi-objective optimization perspective. To this end, within a previously introduced multi-objective optimization framework, we use a smooth B-spline-based dual-dynamic transformation model that allows us to derive gradient information analytically, while still being able to account for large deformations. Within the multi-objective framework, we previously employed a powerful evolutionary algorithm (EA) that computes and advances multiple outcomes at once, resulting in a set of solutions (a so-called Pareto front) that represents efficient trade-offs between the objectives. With the addition of the B-spline-based transformation model, we studied the usefulness of gradient information in multi-objective deformable image registration using three different optimization algorithms: the (gradient-less) EA, a gradient-only algorithm, and a hybridization of these two. We evaluated the algorithms to register highly deformed images: 2D MRI slices of the breast in prone and supine positions. Results demonstrate that gradient-based multi-objective optimization significantly speeds up optimization in the initial stages of optimization. However, allowing sufficient computational resources, better results could still be obtained with the EA. Ultimately, the hybrid EA found the best overall approximation of the optimal Pareto front, further indicating that adding gradient-based optimization for multi-objective optimization-based deformable image registration can indeed be beneficial.

Keywords: Deformable registration, multi-objective optimization, evolutionary algorithms, gradient information

1. INTRODUCTION

The goal in image registration is to find the best transformation that aligns two images. In most registration methods, one image is considered to be the source (also referred to as moving) image, which is mapped via a transformation to the target (also referred to as fixed) image.

In single-objective optimization in general, and also for image registration purposes, the gradient (the direction of greatest increase) of the function to be optimized, is frequently used to guide an optimization algorithm faster to (local) optima, especially if the gradient can be computed efficiently.¹ However, contrary to the state-of-the-art single-objective optimization registration methods, gradient information was so far not yet considered in a multi-objective optimization approach to registration.

Further author information: (Send correspondence to K.P. or T.A.)

K.P.: E-mail: k.pirpinia@nki.nl, Telephone: +31 (0)20 512 2235

T.A.: E-mail: t.alderliesten@amc.uva.nl, Telephone: +31 (0)20 566 6886

In a previously introduced multi-objective optimization framework for deformable image registration², registration is posed as a problem with multiple objectives to be optimized simultaneously. This removes the need of pre-determining a singular combination of objectives that are of interest in image registration, such as image similarity and smoothness of the deformation. For multi-objective optimization, we employed an advanced, model-based Evolutionary Algorithm (EA), known as an Estimation-of-Distribution Algorithm (EDA), which solves an optimization problem by building and sampling a probabilistic model.³ The algorithm is a population-based method, i.e., it computes and advances multiple solutions at once, resulting in a final set of solutions that represents efficient trade-offs between the objectives, called the Pareto front.

In this paper, we assess the utility of incorporating gradient information in the multi-objective optimization framework for deformable image registration. To achieve this, it is important to derive gradient information for both objectives analytically for maximum efficiency. Therefore, we used a dual-dynamic transformation model based on B-splines. Furthermore, the previously introduced dual-dynamic transformation model has the advantage of being capable of tackling large deformations as well as (dis)appearing structures.⁴ In such a model, there are two moving grids instead of one: one for the source- and one for the target image. This allows for better correlation of underlying structures in both the source- and target image. To test the impact of using gradient information, we compare three different optimization algorithms: 1) a gradient-less EA, 2) a gradient-only algorithm, and 3) a hybrid combination of these two.

2. MATERIALS AND METHODS

2.1 B-spline-based dual-dynamic transformation model

Free-form deformation models based on B-splines^{5,6} are widely used in image registration because they provide flexibility, transformation smoothness, and computational efficiency. In such models, the source image is deformed by manipulating an underlying grid of control points that govern, via interpolation, the deformation and thereby the transformation of the image.

We build upon a previously introduced concept of a dual-dynamic transformation model that has shown to be capable of handling large deformations.⁴ In this model, both images are considered moving. Here, we propose a B-spline-based dual-dynamic transformation model. We consider two meshes $n_x \times m_y$ of control points ϕ_s, ϕ_t associated with the source- and target image respectively, and a "virtual" fixed grid of dimensions $N \times M$. A mapping between the images is defined indirectly by mapping the points of the virtual grid to both the source- and the target image through two B-spline transformations $T_s, T_t: R^2 \rightarrow R^2$ which are defined as follows:

$$T_s(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 B_i(u) B_j(v) \phi_{s_{i+k, j+l}}, \quad T_t(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 B_i(u) B_j(v) \phi_{t_{i+k, j+l}},$$

where $u = \frac{x}{n_x} - \left\lfloor \frac{x}{n_x} \right\rfloor$,

$$v = \frac{y}{m_y} - \left\lfloor \frac{y}{m_y} \right\rfloor,$$

$$k = \left\lfloor \frac{x}{n_x} \right\rfloor - 1,$$

$$l = \left\lfloor \frac{y}{m_y} \right\rfloor - 1,$$

$$0 \leq x < N,$$

$$0 \leq y < M,$$

and B_i, B_j are cubic B-spline basis functions. The sets of control points ϕ_s, ϕ_t are the parameters to be optimized.

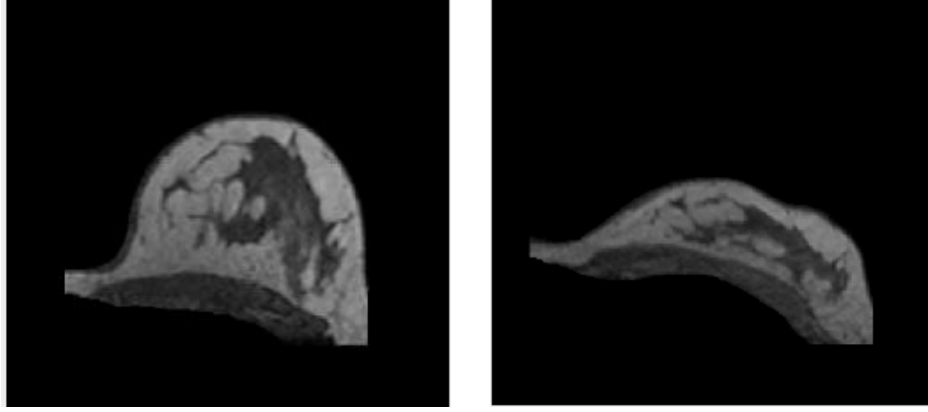


Figure 1. Axial slices from breast MRI scans acquired from a healthy volunteer. *Left*: prone breast MRI (source image). *Right*: supine breast MRI (target image).

2.2 Multi-objective deformable image registration

We formulate deformable image registration as a multi-objective optimization problem. Two objectives are defined: one related to the quality of fit, i.e., the degree of similarity between the images, and one related to the smoothness of the transformations which favors smooth transformations over strong deformations. These objectives are conflicting; one (related to similarity) needs to be maximized, the other (related to smoothness) needs to be minimized. As a consequence, a unique optimal solution does not exist, but a set of equally good ones do, each one better in one objective than the other solutions but worse in the other objective. EAs are known to deal excellently with multi-objective problems⁷, being able to advance multiple solutions (the so-called population) simultaneously, resulting in a final set of solutions (i.e., efficient trade-offs between the objectives) called the Pareto front. Such an optimization method can find possibly existing Pareto-optimal solutions that cannot be found by repeatedly using single-objective optimization methods with different linearly weighted combinations of the objectives (i.e., if the Pareto front is concave).⁷

2.3 Objectives

For the similarity measure we use the sum of squared differences in grey value between the target- and the source image. This measure must be minimized. To calculate this sum, we perform uniform sampling on the virtual grid, acquiring locations in both the source- and target image. We then derive the grey values at these locations using pre-processed B-spline coefficients^{8,9}, and calculate the sum of their squared differences. Note that the use of this particular similarity measure is not a necessity. Other measures could be used here as well.

To address the ill-posedness of the deformable image registration problem, the second objective is introduced to measure smoothness, using second order derivatives of the deformation field.^{5,10,11} We use a smoothing term that involves the second order derivatives of the two deformation fields. Specifically, the smoothing term is the sum of squared second order derivatives of the B-spline transformations and is denoted as ΔT . We formulate the smoothness objective as the sum of the squared differences of the smoothness terms evaluated at pairs of points in the source- and target image.

2.4 Three optimization algorithms

2.4.1 Gradient-less EA

We employ a specific type of EA, known as EDA, that aims to exploit features of the problem's structure via probabilistic modeling. The specific EDA we use is called iMAMaLGAM (incremental Multi-objective Adapted Maximum Likelihood Gaussian Model mixture), in which the probabilistic model is a Gaussian mixture distribution. In related work, iMAMaLGaM was shown to perform excellently, finding sufficient approximations of the optimal Pareto front for various well-known benchmark problems.¹²

2.4.2 Gradient-only algorithm

As a result of being constructed using B-spline transformations as well as using B-spline pixel interpolation, both the similarity- and the smoothness objective are smooth functions. Therefore, we are able to derive analytic expressions of the gradient for both objectives, which makes the use of gradient information much more computationally efficient.

For the similarity objective, calculating the analytic expression for the gradient comes down to calculating the following term:

$$\nabla \text{Similarity} = \frac{\partial \text{Similarity}}{\partial I_i} \frac{\partial I_i}{\partial T_i} \frac{\partial T_i}{\partial \phi_i},$$

where $I_i = I_s, I_t$ are the B-spline pixel-interpolation functions associated with the source- and the target image, $T_i = T_s, T_t$ are the B-spline transformations, and $\phi_i = \phi_s, \phi_t$ are the control points. More details about the derivation of the analytic gradient can be found in the Appendix. For the gradient of the smoothness objective we essentially have to calculate the following term:

$$\nabla \text{Smoothness} = \frac{\partial(\Delta T_i)}{\partial \phi_i}.$$

In single-objective optimization, the application of the gradient is quite straightforward; it is used to indicate the direction in which the function to be optimized improves the most. Many powerful single-objective gradient-based optimization algorithms exist, such as the conjugate gradients method.¹³ However, in a multi-objective framework the notion of improvement is different and much less is known about the use and possible benefit of using gradient information in such a setting. We use a multi-objective gradient-based algorithm from literature that exploits gradient information in three different ways.¹⁴

ROCG The first exploitation method is called Random Objective Conjugate Gradients (ROCG); here, every time the method is called, one objective is chosen *randomly* and the conjugate gradients technique is performed for that objective only.

AORL The second exploitation method is called Alternating Objectives Repeated Line search (AORL). AORL performs a line search in the direction of the negative gradient of an alternatingly chosen objective - this reduces the chance that the algorithm converges to a local optimum for one objective, while making the other worse.

CORL The third method is called Combined Objectives Repeated Line search (CORL). CORL exploits gradient information in a purely multi-objective manner. The set of all improving directions for both objectives is calculated (which can be done on the basis of the gradients of the individual objectives), one direction is chosen randomly, and a line search is performed in this direction, improving both objectives simultaneously.

For maximum efficiency, the multi-objective gradient-based algorithm adaptively decides each time an exploitation method is called which method (i.e., ROCG, AORL, or CORL) to use, based on the number of improved solutions found by that method compared to the number of evaluations it has used.

2.4.3 Hybrid EA

Whether the use of gradient information in real-valued multi-objective optimization is beneficial or not, depends highly on the problem at hand. High dimensionality and existence of local optima are problematic for pure gradient-based algorithms. A combination, however, of the EA with gradient techniques used in an adaptive manner has been shown to be capable of overcoming these difficulties and to find improvements on the Pareto front compared to using solely the EA. Therefore, we consider also this hybridization here. In this hybrid algorithm, the gradient method is applied to solutions that the EA has generated, i.e., at the end of every evolutionary generation cycle.

2.5 Experiments

We considered a large-deformation registration case; prone-supine breast MRI registration (Figure 1). MRI scans were acquired from a healthy volunteer. First, the two MRI scans were rigidly registered on the bony anatomy. Subsequently, one central pair of 2D slices was selected. We performed registration using two 8×8 grids of control points, which resulted in 256 parameters to be optimized. To compare the algorithms' performance and to ensure robustness in the results (as all algorithms include a stochastic component), we ran each algorithm (the gradient-less EA, the gradient-only

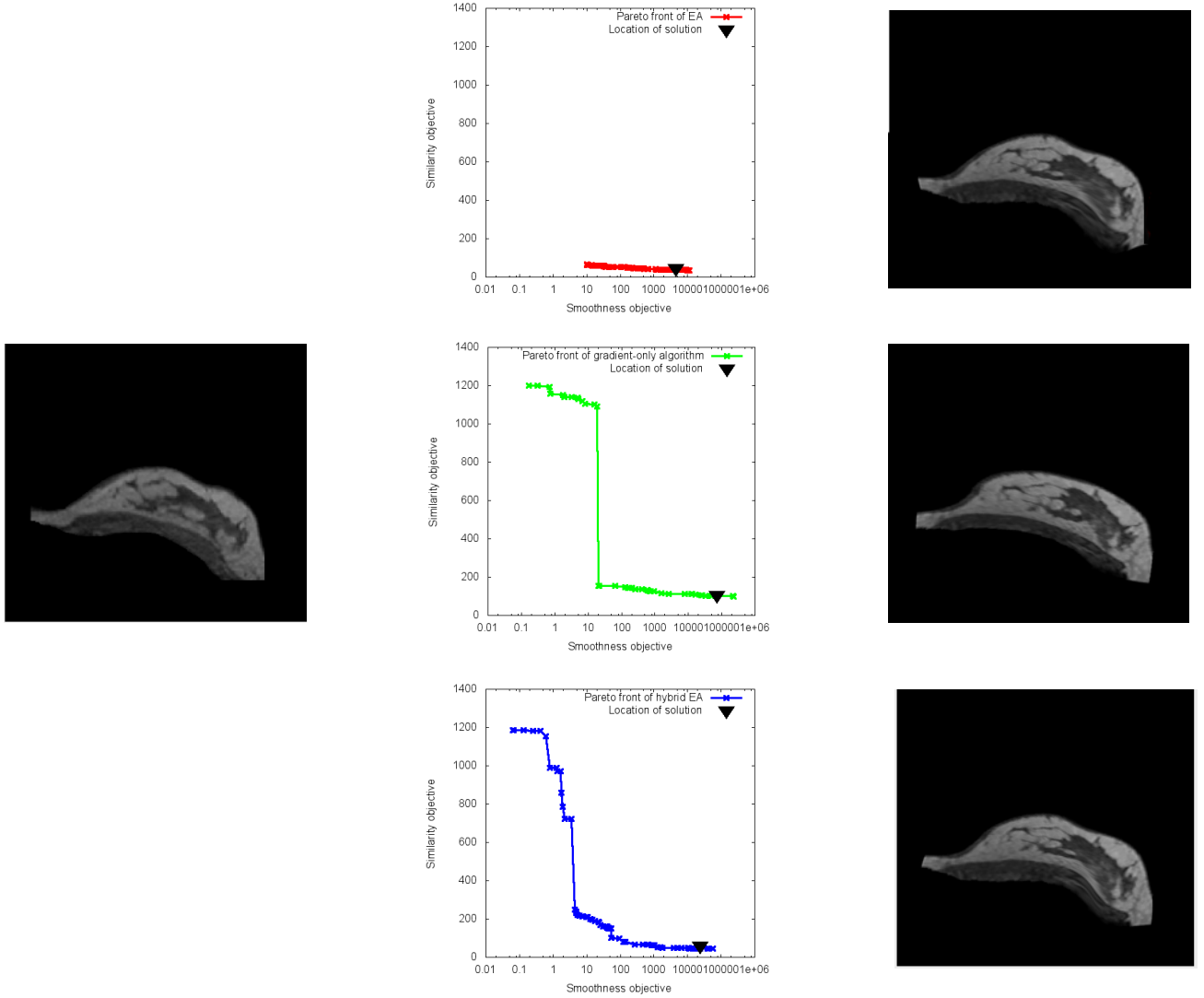


Figure 2. Results obtained with a budget of 2,500,000 evaluations. Result of *from top to bottom*: the gradient-less EA, the gradient-only algorithm, the hybrid EA. *From left to right*: target image, resulting Pareto fronts, and transformed source images associated with the selected solutions (black triangle) on the Pareto front.

algorithm, and the hybrid EA) 10 times, obtaining in total 30 Pareto fronts. Since the optimal Pareto front cannot be known beforehand, the set of best solutions out of these 30 Pareto fronts was taken as the surrogate optimal Pareto front. To measure each algorithms' rate of convergence to the surrogate optimal Pareto front, we used a well-known indicator in multi-objective optimization, called Inverted Generational Distance (IGD).¹⁵ IGD measures how far a result (i.e., an approximation front) is from the (surrogate) optimal Pareto front as well as how well the solutions in an approximation front are spread along the (surrogate) optimal Pareto front. A value of 0 for the IGD corresponds to optimality, i.e., all solutions on the Pareto front are covered by at least one solution in the approximation front. Therefore, the lower the IGD value, the better.

3. RESULTS

The gradient-less EA as well as the hybrid EA, in combination with the use of the B-spline-based dual-dynamic transformation model, are capable of finding good solutions, as opposed to the gradient-only algorithm, that yields poorer results (Figure 2). When comparing the solutions found by all runs of the algorithms in Figure 3, we see that the

gradient-only algorithm is less powerful than the gradient-less EA, since it finds inferior approximation fronts compared to the gradient-less EA and the hybrid EA. Moreover, the hybrid EA is capable of finding a better-spread approximation of the surrogate optimal Pareto front than the gradient-less EA. Even if the solutions that the hybrid EA finds in the interesting part of the Pareto front (for this case: high value for the smoothness objective, combined with a low value for the similarity objective) are not better than those that the gradient-less EA finds, visually they are comparably good, as can be seen in Figure 2. Moreover, the *average* IGD convergence graphs in Figure 5 indicate that the hybrid EA outperforms both the gradient-less EA and the gradient-only algorithm in terms of the number of required evaluations to approximate the surrogate optimal Pareto front with a certain quality level which can be interpreted as the convergence rate. It finds interesting solutions already after 500,000 evaluations, as shown in Figure 4. The gradient-less EA starts with a well-spread Pareto front, but in the end it finds only solutions that are confined to one region (Figure 4). This ultimately results in deteriorating IGD values (Figure 5).

4. DISCUSSION AND CONCLUSION

For the first time in multi-objective deformable image registration, analytical gradient information is derived based on a B-spline-based dual-dynamic transformation model. The possible benefit of exploiting this information in multi-objective image registration is assessed, with promising results. The assessment is based on the comparison of three different optimization algorithms. Moreover, although B-spline deformation models are widely used, they have not been yet studied in a multi-objective framework.

We assessed the utility of multi-objective gradient information for the task of deformable image registration, within a previously introduced multi-objective optimization framework, combined with a B-spline-based dual-dynamic transformation model to account for large deformations. Results from the comparison of three optimization algorithms show that, even given smooth objectives, the use of common multi-objective gradient-only algorithms is not powerful enough, however, exploitation of the gradient in combination with an evolutionary algorithm does have the potential to improve registration results and achieve faster overall convergence.

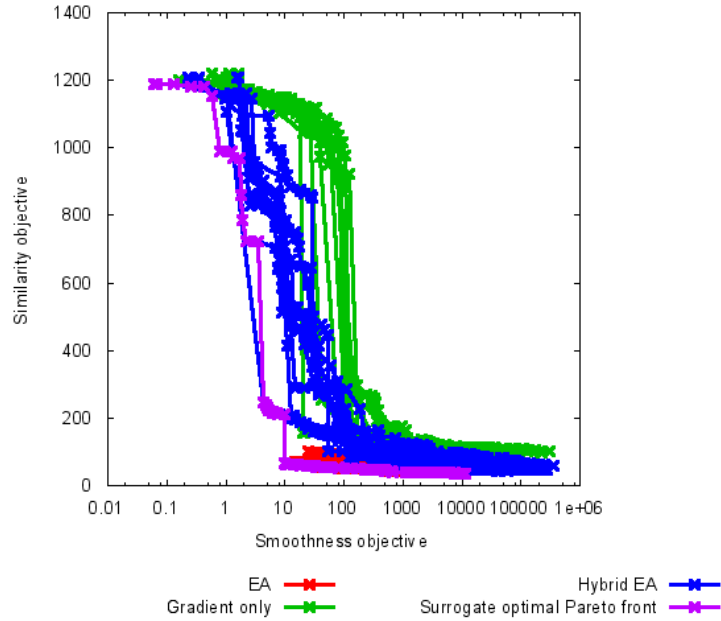


Figure 3. All 30 Pareto fronts of the three optimization algorithms together with the surrogate optimal Pareto front.

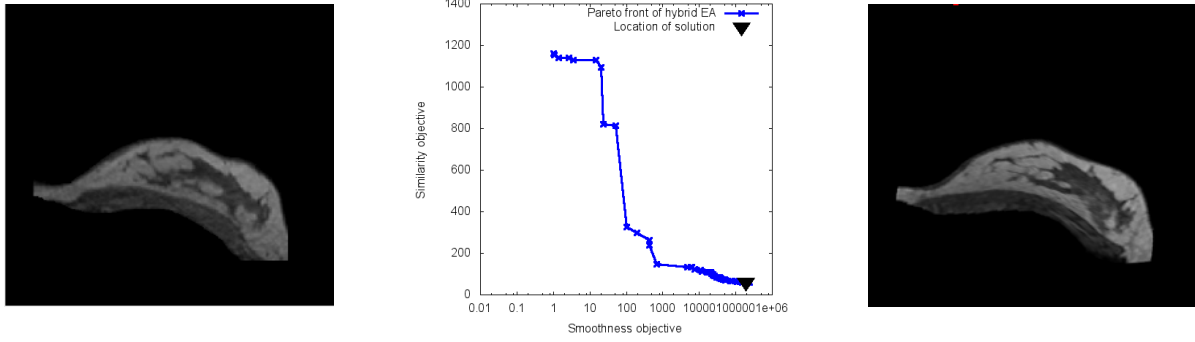


Figure 4. Example of the performance of the hybrid EA with a budget of 500,000 evaluations. *Left*: target image. *Middle*: resulting Pareto front. *Right*: transformed source image associated with the selected solution (black triangle) on the Pareto front.

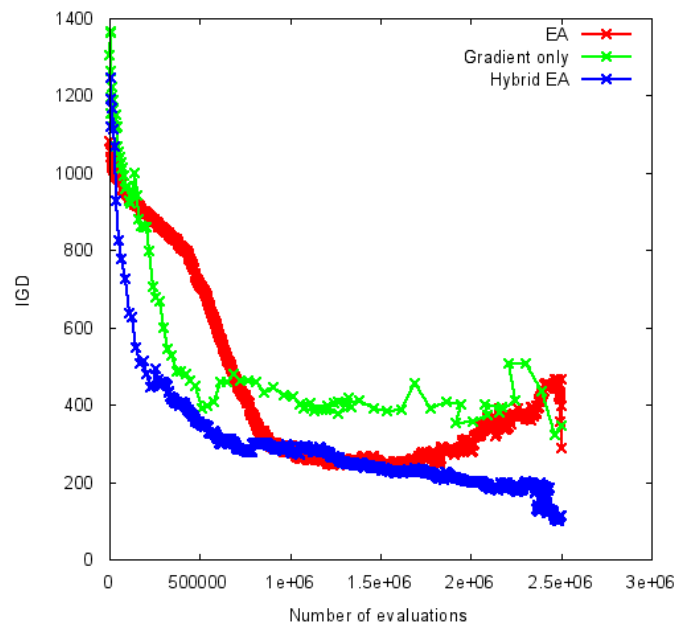


Figure 5. Average convergence graphs with IGD.

ACKNOWLEDGEMENTS

Financial support of this work was provided by the Dutch Cancer Society (Grant No. KWF 2012-5716).

REFERENCES

- [1] Klein, S., Staring, M. and Pluim, J. P. W., "Comparison of gradient approximation techniques for optimisation of mutual information in nonrigid registration," *Proc. SPIE* 5747, 192-203 (2005).
- [2] Alderliesten, T., Sonke, J.-J., Bosman, P. A. N., "Multi-objective optimization for deformable image registration: proof of concept," *Proc. SPIE* 8314, 831420 (2012).
- [3] Lozano, J. A., Larrañaga, P., Inza, I. and Bengoetxea, E., [Towards a new evolutionary computation. *Advances in estimation of distribution algorithms*], Springer-Verlag, Berlin (2006).
- [4] Alderliesten, T., Sonke, J.-J. and Bosman, P. A. N., "Deformable image registration by multi-objective optimization using a dual-dynamic transformation model to account for large anatomical differences," *Proc. SPIE* 8669, 866910, doi:10.1117/12.2006783 (2013).
- [5] Rueckert, D., Sonoda, L. I., "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Trans Med Imaging* 18(8), 712-721 (1999).
- [6] Lee, S., Wolberg, G. and Shin, S. Y., "Scattered data interpolation with multilevel B-splines," *IEEE Trans Vis Comput Graph*, 3(3), 228-244 (1997).
- [7] Deb, K., [Multi-objective optimization using evolutionary algorithms], John Wiley & Sons Inc., New York (2001).
- [8] Thévenaz, P., Blu, T. and Unser, M., "Interpolation revisited," *IEEE Trans Med Imaging* 19(7), 739-758 (2000).
- [9] Thévenaz, P., Blu, T. and Unser, M., "Image interpolation and resampling," [Handbook of medical imaging], Academic Press Inc, Orlando, 393-420 (2000).
- [10] Grimson, W. E. L., "An implementation of a computational theory of visual surface interpolation," *Comput Vis Graph Image Process* 22(1), 39-49 (1983).
- [11] Fischer, B. and Modersitzki, J., "A unified approach to fast image registration and a new curvature based registration technique." *Linear Algebra Appl.* 380, 107-124 (2004).
- [12] Bosman, P. A. N. and Alderliesten, T., "Incremental Gaussian model-building in multi-objective EDAs with an application to deformable image registration," *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*, 241-248 (2012).
- [13] Hestenes, M. R. and Stiefel, E., [Methods of conjugate gradients for solving linear systems], National Bureau of Standards Washington, DC, Washington (1952).
- [14] Bosman, P. A. N., "On gradients and hybrid evolutionary algorithms for real-valued multiobjective optimization." *IEEE Trans Evol Comput* 16(1), 51-69 (2012).
- [15] Knowles, J. and Corne, D., "On metrics for comparing nondominated sets." *Proc. 2002 Congress on Evolutionary Computation* 1, 711-716 (2002).

APPENDIX

For the transformation model we have two mappings $T_s, T_t : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. We have furthermore two sets $n_x \times m_y$ of control points $\phi_s = (\phi_{s_x}, \phi_{s_y})$, $\phi_t = (\phi_{t_x}, \phi_{t_y})$ associated with the source- and target image. The mappings T_s, T_t can be rewritten as follows:

$$\begin{aligned} T_{s_x}(x, y) &= \sum_{i=0}^3 \sum_{j=0}^3 B_i \left(\frac{x}{n_x} - \left\lfloor \frac{x}{n_x} \right\rfloor \right) B_j \left(\frac{y}{m_y} - \left\lfloor \frac{y}{m_y} \right\rfloor \right) \phi_{s_{x_i+k, j+l}} \\ T_{s_y}(x, y) &= \sum_{i=0}^3 \sum_{j=0}^3 B_i \left(\frac{x}{n_x} - \left\lfloor \frac{x}{n_x} \right\rfloor \right) B_j \left(\frac{y}{m_y} - \left\lfloor \frac{y}{m_y} \right\rfloor \right) \phi_{s_{y_i+k, j+l}} \\ T_{t_x}(x, y) &= \sum_{i=0}^3 \sum_{j=0}^3 B_i \left(\frac{x}{n_x} - \left\lfloor \frac{x}{n_x} \right\rfloor \right) B_j \left(\frac{y}{m_y} - \left\lfloor \frac{y}{m_y} \right\rfloor \right) \phi_{t_{x_i+k, j+l}} \\ T_{t_y}(x, y) &= \sum_{i=0}^3 \sum_{j=0}^3 B_i \left(\frac{x}{n_x} - \left\lfloor \frac{x}{n_x} \right\rfloor \right) B_j \left(\frac{y}{m_y} - \left\lfloor \frac{y}{m_y} \right\rfloor \right) \phi_{t_{y_i+k, j+l}} \end{aligned}$$

where $k = \left\lfloor \frac{x}{n_x} \right\rfloor - 1$, $l = \left\lfloor \frac{y}{m_y} \right\rfloor - 1$.

The similarity measure we use is the sum of squared intensity differences. For a dual-dynamic transformation model it is defined as follows:

$$Similarity = \sum_{x, y} [I_s(T_{s_x}(x, y), T_{s_y}(x, y)) - I_t(T_{t_x}(x, y), T_{t_y}(x, y))]^2$$

where I_s and I_t are interpolated intensity functions for the source- and target image defined also in terms of cubic B-splines:

$$\begin{aligned} I_s(x, y) &= \sum_{p_1 \in \mathbb{Z}} \sum_{p_2 \in \mathbb{Z}} c_{p_1, p_2} B(x - p_1) B(y - p_2), \\ I_t(x, y) &= \sum_{n_1 \in \mathbb{Z}} \sum_{n_2 \in \mathbb{Z}} d_{n_1, n_2} B(x - n_1) B(y - n_2), \end{aligned}$$

where $p_i, n_i, i = 1, 2$ loop over the dimensions of the images, and $c_{p_1, p_2}, d_{n_1, n_2}$ are the pre-processed image B-spline coefficients.

To calculate the gradient of the similarity measure with respect to the control points, we need to calculate the partial derivative

$$\frac{\partial Similarity}{\partial \phi_i} \quad \forall \phi_i, \quad i = s, t.$$

For example, if we want to calculate the partial derivative for the coefficient $\phi_{s_{x_1,1}}$, we proceed as follows:

$$\begin{aligned}
\frac{\partial \text{Similarity}}{\partial \phi_{s_{x_1,1}}} &= \frac{\partial \sum_{x,y} [I_s(T_{s_x}, T_{s_y}) - I_t(T_{t_x}, T_{t_y})]^2}{\partial \phi_{s_{x_1,1}}} \\
&= \sum_{x,y} 2(I_s(T_{s_x}, T_{s_y}) - I_t(T_{t_x}, T_{t_y})) \left(\frac{\partial I_s(T_{s_x}, T_{s_y})}{\partial \phi_{s_{x_1,1}}} - \frac{\partial I_t(T_{t_x}, T_{t_y})}{\partial \phi_{s_{x_1,1}}} \right) \\
&= \sum_{x,y} 2(I_s(T_{s_x}, T_{s_y}) - I_t(T_{t_x}, T_{t_y})) \cdot \\
&\quad \left[\sum_{p_1} \sum_{p_2} c_{p_1, p_2} \frac{\partial B(T_{s_x} - p_1) B(T_{s_x} - p_2)}{\partial \phi_{s_{x_1,1}}} - \sum_{n_1} \sum_{n_2} d_{n_1, n_2} \frac{\partial B(T_{t_x} - n_1) B(T_{t_x} - n_2)}{\partial \phi_{s_{x_1,1}}} \right]
\end{aligned}$$

If we look at the right term of the product, since $\phi_{s_{x_1,1}}$ is associated with the source image, the right sum will be equal to 0. Now, for one pair (x, y) we can calculate the left term:

$$\frac{\partial B(T_{s_x} - p_1) B(T_{s_x} - p_2)}{\partial \phi_{s_{x_1,1}}} = \frac{\partial B(T_{s_x} - p_1)}{\partial \phi_{s_{x_1,1}}} B(T_{s_x} - p_2) + B(T_{s_x} - p_1) \frac{\partial B(T_{s_x} - p_2)}{\partial \phi_{s_{x_1,1}}}$$

Then since $\phi_{s_{x_1,1}}$ could be a part only of the T_{s_x} mapping, the right term of the above equation is again equal to 0, and therefore the derivative comes down to calculating the following term:

$$\frac{\partial B(T_{s_x} - p_1)}{\partial \phi_{s_{x_1,1}}} = \frac{\partial B(T_{s_x} - p_1)}{\partial T_{s_x}} \frac{\partial T_{s_x}}{\partial \phi_{s_{x_1,1}}}$$

If $\phi_{s_{x_1,1}}$ is not in the neighborhood of (x, y) , then the above term will be equal to 0. Otherwise, the left term will be a polynomial of second degree and the right term will consist of the product of the two B-spline functions associated with $\phi_{s_{x_1,1}}$.