Technology in Cancer Research and Treatment
ISSN 1533-0346
Volume 12, Number 4, August 2013

The Role of Regularization in Deformable Image Registration for Head and Neck Adaptive Radiotherapy

www.tcrt.org
DOI: 10.7785/tcrt.2012.500327

D. Ciardo, M.Sc.[1]*
M. Peroni, Ph.D.[2]
M. Riboldi, Ph.D.[2,3]
D. Alterio, M.D.[1]
G. Baroni, Ph.D.[2,3]
R. Orecchia, M.D.[1,4,5]

[1]Advanced Radiotherapy Center, Division of Radiotherapy, European Institute of Oncology, via Ripamonti 435, 20141 Milano, Italy
[2]Department of Bioengineering, Politecnico di Milano, via Golgi 39, 20133 Milano, Italy
[3]Bioengineering Unit, Centro Nazionale di Adroterapia Oncologica, Strada Campeggi 53, 27100 Pavia, Italy
[4]Medical Department, Centro Nazionale di Adroterapia Oncologica, Strada Campeggi 53, 27100 Pavia, Italy
[5]Università degli Studi di Milano, via Festa del Perdono 7, 20122 Milano, Italy

Deformable image registration provides a robust mathematical framework to quantify morphological changes that occur along the course of external beam radiotherapy treatments. As clinical reliability of deformable image registration is not always guaranteed, algorithm regularization is commonly introduced to prevent sharp discontinuities in the quantified deformation and achieve anatomically consistent results. In this work we analyzed the influence of regularization on two different registration methods, *i.e.* B-Splines and Log Domain Diffeomorphic Demons, implemented in an open-source platform. We retrospectively analyzed the simulation computed tomography (CTsim) and the corresponding re-planning computed tomography (CTrepl) scans in 30 head and neck cancer patients. First, we investigated the influence of regularization levels on hounsfield units (HU) information in 10 test patients for each considered method. Then, we compared the registration results of the open-source implementation at selected best performing regularization levels with a clinical commercial software on the remaining 20 patients in terms of mean volume overlap, surface and center of mass distances between manual outlines and propagated structures. The regularized B-Splines method was not statistically different from the commercial software. The tuning of the regularization parameters allowed open-source algorithms to achieve better results in deformable image registration for head and neck patients, with the additional benefit of a framework where regularization can be tuned on a patient specific basis.

Key words: Regularization; Deformable image registration; B-Splines; Demons; MIMvista; Head and neck cancer.

Introduction

Current clinical practice in head and neck (HN) cancer radiotherapy provides evidence that significant anatomical changes over the course of treatment occur, especially in terms of patient weight loss, tumor volume shrinkage and reduction of parotid volumes (1-5). Recent studies reported an average weight loss of 3.3% in HN patients (3), along with a tumor shrinkage of 70% (1) and parotid volume reduction of 49.8% (2). HN patients may therefore benefit from re-planning, as a way to minimize the effects of measured anatomical variations (4, 6). Smart and image-driven criteria for replanning individuation have recently been proposed (7), together with robust validation metrics for adapted structures evaluation (8).

Abbreviations: CTsim: Simulation Computed Tomography; CTrepl: Re-planning Computed Tomography; HN: Head and Neck; DIR: Deformable Image Registration; IMRT: Intensity Modulated Radiation Therapy; AP: Anterior-posterior; LL: Latero-lateral; SI: Superior-inferior; L-BFGS: Limited-memory Broyden-Fletcher-Goldfarb-Shanno; RMSE: Root Mean Squared Error; NMI: Normalized Mutual Information; nGTV: nodal Gross Tumor Volume; DSC: Dice Similarity Coefficient.

*Corresponding author:
D. Ciardo, M.Sc.
E-mail: delia.ciardo@ieo.it

Deformable image registration (DIR) provides the methodological framework to map deforming organs in a comprehensive adaptive treatment plan, accounting for deformations occurred along the treatment and assessing the dosimetric impact of the anatomical changes, detected using in-room image guidance (9). The task of DIR is to define the geometric transformation Φ that maps the spatial coordinates of homologous points between different imaging studies, in order to make the floating transformed image $M\,x, y, z$ similar to the reference $R(x, y, z)$, so that $R\,x, y, z = M\,\Phi(x, y, z)$. The deformation models determine the way in which the floating image can be deformed to improve the similarity with respect to the reference image.

Several DIR algorithms have been proposed (10-13). A coarse classification can be drawn to distinguish between physics-based and intensity-based algorithms (14). It has been shown that multi-resolution approaches are preferable to minimize the effects of large deformations, as those in HN adaptive treatments (15, 11). In all cases, if the optimization process is not somewhat restricted, it is possible to achieve folded configuration of the floating volumes, resulting in inconsistent topology of the deformation field and physically meaningless deformed images. One possible strategy is to control the iteration process, by means of smart convergence detection strategies (16).

When intensity-based algorithms are considered, specific strategies are needed to ensure topology preservation. This is typically accomplished by adding regularization terms in the computation of the deformation field (10, 11). Lee and colleagues suggested that regularization of the deformation field and/or the use of constraints would be beneficial to improve results in automatic intensity-based DIR (3). Recent DIR comparison studies in HN patients show that intensity-based algorithms, with embedded regularization, feature enhanced accuracy in the alignment of deformable anatomy (10). The regularization process contributes to obtain smoother deformation fields, limiting the occurrences of folding (17). Nonetheless no systematic studies have been proposed for investigating the needed amount of regularization in DIR applied to HN adaptive radiotherapy.

On the other hand, the reliability of DIR procedures is critical for the clinical implementation of effective adaptive radiotherapy protocols (18). A crucial issue is the preservation of topology in the computation of the deformation field, to maintain neighborhood relationships and integrity of anatomical structures (19). Assessing the clinical validity of DIR is a complex task (19, 20) and appropriate validation study at each anatomical site are needed (3). Comparative studies have been performed, relying on either deformable phantoms or common datasets with verification landmarks or structures segmented by experts (10-13). Visual inspection of the deformation field along with verification of topology-related properties, such as the Jacobian or inverse consistency, have been suggested as key elements for DIR validation (21).

In this work we report on DIR regularization on HN cases, showing the sensitivity of results to variable levels of regularization in the deformation field. We analyzed the influence of regularization on B-Splines (22) and on Log Domain Diffeomorphic Demons (23) DIR algorithm. Furthermore, we compared the results with a commercial implementation of DIR in the clinical package MIMvista (version 5.4.7, MIMvista Corp, Cleveland, OH). Regularization parameters for both algorithms were largely varied for a set of 10 training cases, and the best performing values were selected for a subsequent testing phase on 20 HN patients.

Materials and Methods

Image Dataset

The study was performed retrospectively on 30 HN cancer patients, who received a radiotherapy treatment at the European Institute of Oncology (IEO, Milano, Italy). All patients underwent IMRT, with a dose prescription of 60-70 Gy, delivered in 2-2.12 Gy/fraction, thus leading to a 5-7 weeks treatment length. The current protocol prescribes the acquisition of a planning contrast enhanced CT (CTsim), as well as of a re-planning CT at 40 and/or 50 Gy (CTrepl), acquired on a GE Light Speed (GE Medical System, Fairfield, CT). At each fraction, patients were immobilized by means of customized head and shoulder thermoplastic mask, bite block and headrest. The image resolution for the CT scans used for planning is approximately 1 mm in anterior-posterior (AP) and latero-lateral (LL) directions and 3 mm in superior-inferior (SI) direction. Image pre-processing included the removal of the treatment couch and a preliminary 6 degree of freedom rigid registration, in order to compensate any global misalignment that would compromise deformable registration results. The rigid registration was in any case performed on the whole image, in order to make the non-rigid registration results comparable among the different algorithms.

Algorithms for Regularized Deformable Image Registration

Mainstream intensity-based DIR algorithms such as B-splines and Demons (parametric and non-parametric, respectively) were considered in the analysis (10-13), with a benchmark based on a commercial system for DIR in clinical practice. In (24), MIMvista is described as an intensity-based algorithm with essentially limitless degrees of freedom, depending on required accuracy and execution speed. Details on the algorithm are not available and parameter tuning is not allowed to the user.

Demons Implementation

There are a number of different implementation of the original demons algorithm, as presented by Thirion (25). The Log Domain Diffeomorphic Demons is an ITK-based interesting variation formulated by Vercauteren that combines the advantage of optimizing a diffeomorphic transformation with a computationally efficient framework (19).

The algorithm aims at minimizing the global energy:

$$E(R,\ M,\ \Phi,\ s) = \sigma_i^{-2}\mathrm{Simil}(R,\ M\circ\Phi) + \sigma_c^{-2}\mathrm{Dist}(\Phi,\ s) + \sigma_T^{-2}\mathrm{Reg}(s) \quad [1]$$

where

$$\mathrm{Simil}(R,\ M\circ\Phi) = \| R - M\circ\Phi)\|^2 \quad [2]$$

is the image disparity measure (*i.e.* mean squared error),

$$\mathrm{Dist}(\Phi,\ s) = \| \Phi - s\|^2 \quad [3]$$

represents the similarity between the deformations Φ and s and Reg(s) is the degree of smoothness of the deformation. Minimizing the global energy means finding a small update u to compose with the current transformation s such that

$$\Phi = s\circ u \quad [4]$$

At each iteration, an update of the velocity field $\boldsymbol{u}^*$ is computed minimizing:

$$E_{\mathrm{diffeo}}(F,\ M,\ \Phi,\ s) = \sigma_i^{-2}\ \mathrm{Simil}(F,\ M\circ\Phi) + \sigma_c^{-2}\mathrm{Dist}(\Phi,\ s) \quad [5]$$

$$= \sigma_i^{-2}\ \| F - m(s\circ e^{\boldsymbol{u}})\|^2 + \sigma_c^{-2}\ \| e^{\boldsymbol{u}}\|^2 \quad [6]$$

Then, s is obtained as:

$$s \leftarrow K_{\mathrm{diff}} \times \exp(K_{\mathrm{fluid}} \times \boldsymbol{u}^*) \quad [7]$$

where $K_{\mathrm{fluid}} = \mathcal{G}[0, \sigma_{\mathrm{fluid}}^2 \boldsymbol{I}]$ and $K_{\mathrm{diff}} = \mathcal{G}[0, \sigma_{\mathrm{diff}}^2 \boldsymbol{I}]$ are the Gaussian kernels used to perform the fluid-like regularization of the update to the transformation and the diffusion-like regularization of the updated transformation, respectively. The amount of regularization is controlled by the amplitude of the Gaussian Kernel, thus by changing $\sigma_{\mathrm{fluid}}^2$ and σ_{diff}^2.

B-Splines Implementation

Among the parametric DIR methods, the choice of B-splines, was mainly motivated from its characteristic of local support, robustness and great flexibility. Amongst the available B-Splines algorithms for DIR, we selected the memory efficient implementation in Plastimatch (www.plastimatch.org), an open-source software for image registration that allows flexible DIR implementation as a function of a customizable parameters set (18).

In our tests, we chose as optimizer the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method, a quasi-Newton algorithm known for its superior performance when dealing with high-dimensionality problems (25, 27). Plastimatch B-Splines registration algorithm features a second derivative additive regularization, whose contribution is regulated by a weight parameter λ (17).

For a vector field $\boldsymbol{u} = (u_x, u_y, u_z)$ the regularization term is given as:

$$c_{\mathrm{Reg}} = \int c_{\mathrm{Reg},x} + c_{\mathrm{Reg},y} + c_{\mathrm{Reg},z} \quad [8]$$

where

$$c_{\mathrm{Reg},x} = \left(\frac{\partial u_x^2}{\partial x^2}\right)^2 + \left(\frac{\partial u_x^2}{\partial y^2}\right)^2 + \left(\frac{\partial u_x^2}{\partial z^2}\right)^2 + \left(\frac{\partial u_x^2}{\partial x\,\partial y}\right)^2 + \left(\frac{\partial u_x^2}{\partial x\,\partial z}\right)^2 + \left(\frac{\partial u_x^2}{\partial y\,\partial z}\right)^2 \quad [9]$$

The integration is performed over the domain of the fixed image. This gives the general form of the cost function to be minimized as follows:

$$c = c_{\mathrm{Imm}} + \lambda c_{\mathrm{Reg}} \quad [10]$$

$$= c_{\mathrm{Imm}} + \lambda \int\int\int_{-\infty}^{+\infty} \left[\left(\frac{\partial^2 \boldsymbol{u}}{\partial x^2}\right)^2 + \left(\frac{\partial^2 \boldsymbol{u}}{\partial y^2}\right)^2 + \left(\frac{\partial^2 \boldsymbol{u}}{\partial z^2}\right)^2 + 2\left(\frac{\partial^2 \boldsymbol{u}}{\partial x \partial y}\right)^2 + 2\left(\frac{\partial^2 \boldsymbol{u}}{\partial y \partial z}\right)^2 + 2\left(\frac{\partial^2 \boldsymbol{u}}{\partial z \partial x}\right)^2\right] dx\,dy\,dz \quad [11]$$

where c_{Imm} accounts for the similarity of the images and c_{Reg} is the proper regularization metric term.

Experiments

In order to compare objectively the performance of the two algorithms and the commercial software package, we selected 10 non-contoured duplets of CT scans (CTsim and CTrepl) as blinded dataset. We run the DIR procedures between CTsim and CTrepl chosen as reference and floating image, respectively, varying regularization parameters values from 10^{-5} to 1 (with one magnitude order increase step) for λ and from 0 to 3 (with step of half a voxel, 0.5 mm in our dataset) for σ_{fluid} and σ_{diff}. Registrations results were evaluated in terms of residual root mean square error (RMSE) and normalized mutual information (NMI) between the CTrepl and corresponding output images. These first results let us define the best performing values of σ_{fluid}, σ_{diff} and λ, that were used for the subsequent accuracy verification phase. In order to generalize the verification results, we considered an independent dataset composed of CT studies in 20 patients, on which we performed deformable registration between each pair of CTsim and CTrepl using

regularized B-Splines, Demons algorithm and MIMvista. On this set of images, mandibles, parotid glands and nodal gross tumor volumes (nGTVs) were retrospectively drawn by the same expert radiation oncologist. Registrations results were evaluated in terms of RMSE, NMI, Dice similarity coefficient (DSC) (28), 3D residual center of mass distance and surface distance between the manually segmented volumes of CTsim and the deformed structures, which were obtained by warping the original ones drawn on CTrepl according to the vector fields resulting from the DIR process. Post registration NMI was calculated between the warped CT and the CTrepl and scaled for the NMI value of CTrepl. The Surface Distance was computed relying on the Hausdorff Distance implementation provided by the Insight Toolkit (ITK, www.itk.org), as the average of the bi-directional Hausdorff distances obtained over the binary contours of interest using the Danielsson distance maps (29).

All the registrations realized by the open-source software were computed on a 2.4 GHz Linux machine (single core, RAM 16 GB).

We evaluated the differences among algorithms outcomes by using a Friedman test for a paired-wise comparison at 0.01 level of significance and post-hoc comparison based on group ranks.

Results

Variability as a Function of the Amount of Regularization

In Figure 1, we report the median and inter-quartile distribution of RMSE (Figure 1A and 1B, respectively) and NMI (Figure 1C and 1D) for all analyzed Log Domain Diffeomorphic Demons registrations performed for training, as a function of σ_{fluid} and σ_{diff}. In Figure 1B and 1D, the higher variability corresponds to the case in which no regularization is set (σ_{fluid} and σ_{diff} both equal to zero). It also is evident that larger values of regularization parameters do not necessarily imply better outcomes, but rather higher RMSE and lower NMI. The optimal configuration is found as trade-off between the two regularization parameters in terms of RMSE minimization and NMI maximization. The dominant parameter seems to be σ_{diff}, which led to worse results when being increased. Conversely, no significant changes were caused by varying the value of σ_{fluid}, which simply implied an increase of the computational cost (>1.5 hours).

Therefore, during the subsequent verification phase, we set σ_{fluid} to zero (*i.e.* no fluid-like regularization) and we tested values of σ_{diff} of 0.5 and 1.5, *i.e.* at the best and at the borderline scores. Calculation time ranges between 15 and 30 minutes, approximately. Imposing a σ_{diff} of 0.5 voxels means to have a very low amount of regularization, whereas 1.0 or 1.5 voxel kernel radius are reported by Vercauteren and colleagues as their preferred values (23, 30).

The evaluation of the performance of the B-Splines deformable image registration (Figure 2) points out that, apart from few cases, the amount of regularization does not have a significantly different impact on the final RMSE and NMI values. Nonetheless, the absence of regularization or a too large λ is reflected by an worsening in the chosen evaluation metrics. In order to highlight this influence, λ was fixed to 10^{-3} and 5×10^{-2}, in the attempt to assess the relationship between the amount of regularization and the final evaluation scores (DSC, 3D center of mass distance and surface distance). Average calculation time for B-Splines registrations is inferior to 10 minutes.

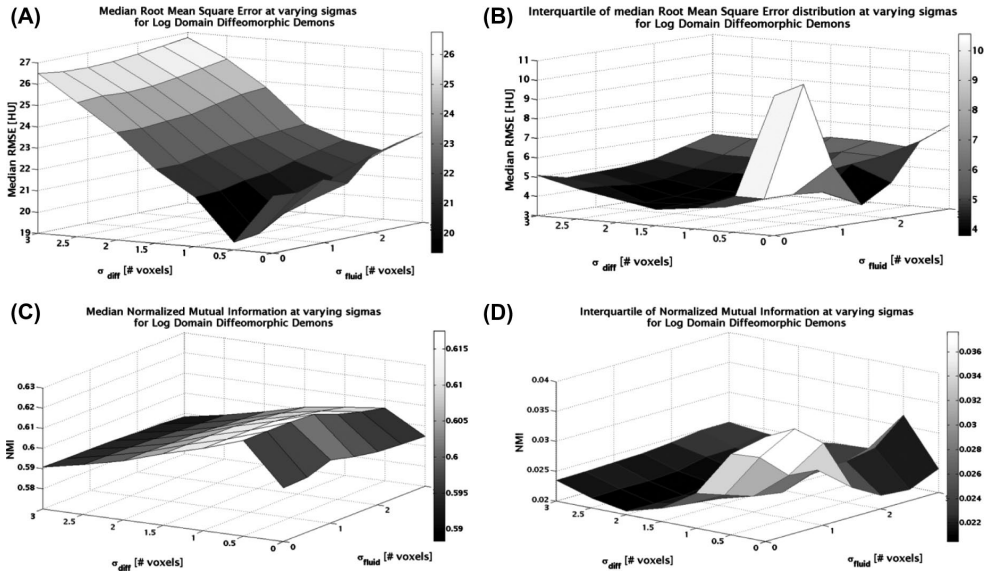

Figure 1: Median and inter-quartile RMSE (panels **A** and **B**, respectively) and NMI (panels **C** and **D**, respectively) at varying σ_{diff} and σ_{fluid} from 0 to 3 for Log Domain Diffeomorphic Demons. The reported colorbars reflect the different scale of values of median and inter-quartile RMSE and NMI.

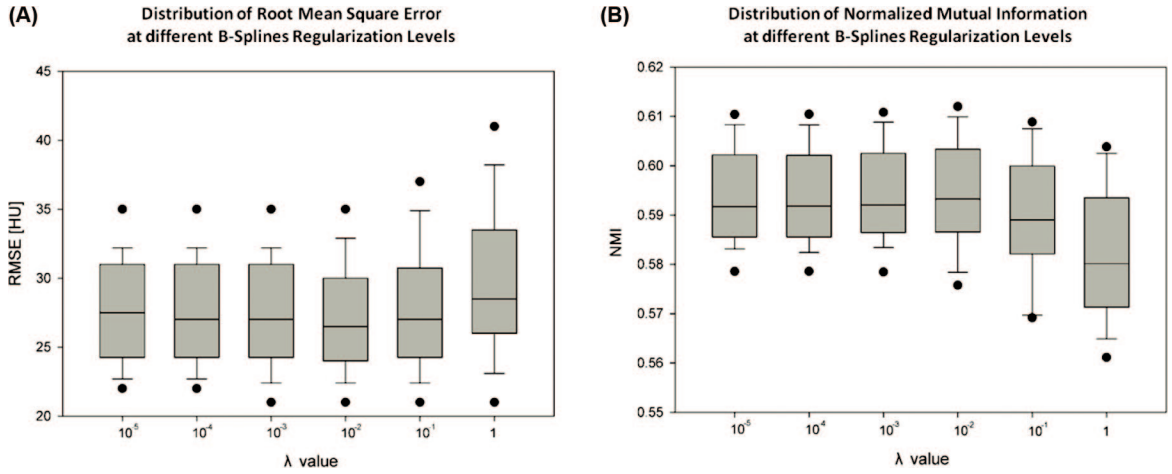

Figure 2: Median and inter-quartile RMSE (panel **A**) and NMI (panel **B**) at varying λ from 10^{-5} to 1 for B-Splines. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the 90th and 10th percentiles, and the dots represent the extreme outliers.

Accuracy Evaluation

As previously described, the best performing regularization parameters for both parametric and non-parametric registration approaches, were applied on a testing dataset. We calculated the median and inter-quartile RMSE values at chosen λ, σ_{diff} and σ_{fluid} and we checked their agreement with the results obtained in the previous phase for a wider variation of regularization parameters. In Log Domain Demons implementation, median RMSE (25th-75th percentile) reached 22 HU (19-25) for $\sigma_{diff} = 0.5$ and

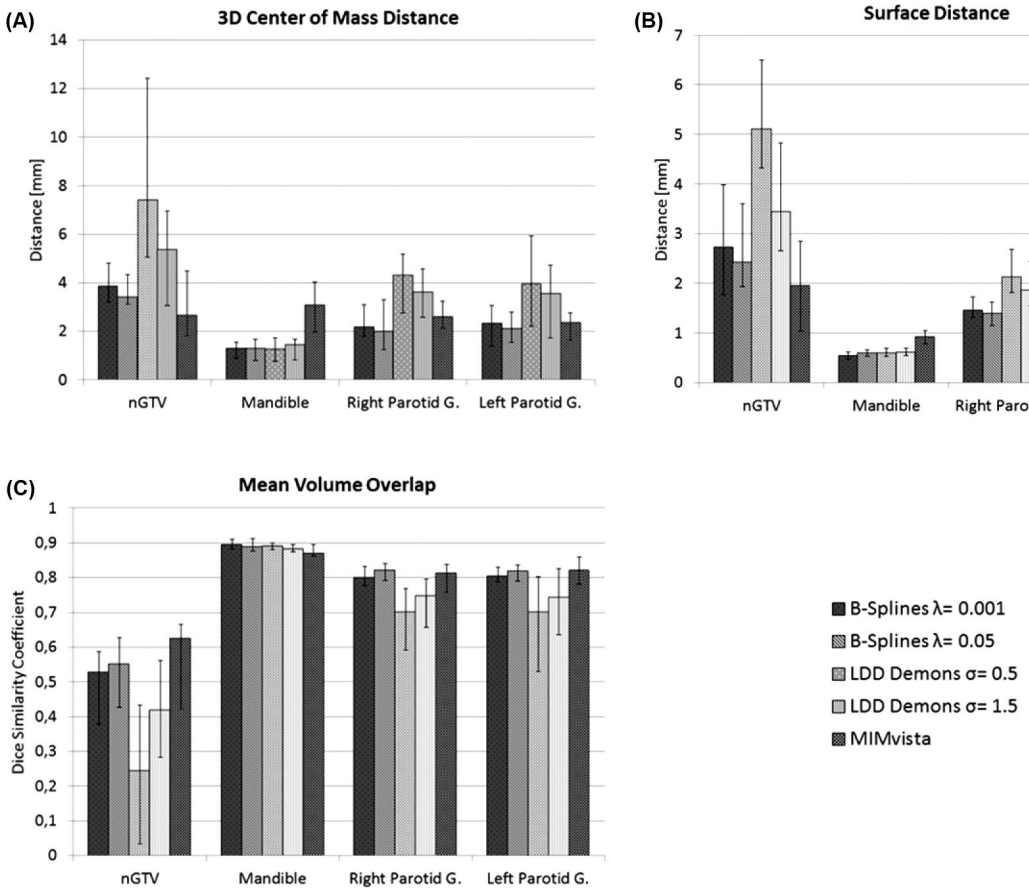

Figure 3: The median, 25th and 75th percentiles distributions of 3D center of mass distance (panel **A**), surface distance (panel **B**) and Dice similarity index (panel **C**) for n-GTV, mandible, right and left parotid glands are shown (from left to right: B-Splines with $\lambda = 10^{-3}$, B-Splines with $\lambda = 0.05$, Log Domain Demons with $\sigma_{diff} = 0.5$ and $\sigma_{fluid} = 0.0$, Log Domain Demons with $\sigma_{diff} = 1.5$ and $\sigma_{fluid} = 0.0$, MIMvista).

23 HU (22-26) for $\sigma_{diff} = 1.5$, whereas, for B-Splines implementation it results in 28 HU (26.00-31.00) for $\lambda = 10^{-3}$ and 28 HU (26-31) for $\lambda = 0.05$. Similarly, the median NMI (25^{th}-75^{th} percentile) was 0.60 (0.59-0.60) for $\sigma_{diff} = 0.5$ and 0.61 (0.60-0.62) for $\sigma_{diff} = 1.5$ for Log Domain Demons implementation, and 0.59 (0.58-0.60) for $\lambda = 10^{-3}$ and for $\lambda = 10^{-3}$ in B-Splines implementation.

The performance assessments of the algorithms were obtained by comparing the anatomical structures, which were manually drawn on the original planning CT and propagated according to the estimated deformation field, with those outlined on the re-planning CT. In Figure 3, the distribution (median and inter-quartile range) of 3D center of mass distance, DSC and surface distance for the 20 patients included in this study are reported. We computed the scores for both open-source implementations and MIMvista with a separate ad-hoc designed tool (www.plastimatch.org).

No statistical difference exists between the distributions of the metrics for the two chosen λ values, for any of the considered evaluation metrics. Log Domain Diffeomorphic Demons with $\sigma_{diff} = 0.5$ was the worst performing combination for all structures except for the mandible, which was warped accurately by the algorithm. In Figure 4 an example of warped parotid glands is shown for the different algorithms and levels of regularization that were tested. For the Demons algorithm the warping of anatomical structures suffered from a significant lack of smoothing (panels A and B), whereas B-Splines resulted in more regularly shaped contours (panels C and D), thus confirming the numerical results in Figure 3.

For the mandible, no statistical difference was found among the algorithms in terms of DSC values. MIMvista was significantly different from both B-Splines and Demons propagated contours and post-hoc comparison revealed that it was worse performing in terms of 3D center of mass and median surface distance.

For the right parotid gland, no significant difference was found between the Log Domain Demons implementations at different σ_{diff} for all metrics. Conversely, both regularization levels were significantly different from Plastimatch regularization levels and MIMvista for all metrics. By looking at ranks, Log Domain Diffeomorphic Demons implementation exhibited the poorest performance amongst the three algorithms. Regularized B-Splines did not show statistically significant differences of all metrics with respect to MIMvista for both parotids.

For left parotid, Log Domain Diffeomorphic Demons with $\sigma_{diff} = 0.5$ performed differently from B-Splines with $\lambda = 0.05$ and MIMvista in terms of both DSC and mean surface distance; no difference was found among the methods in 3D center of mass distance.

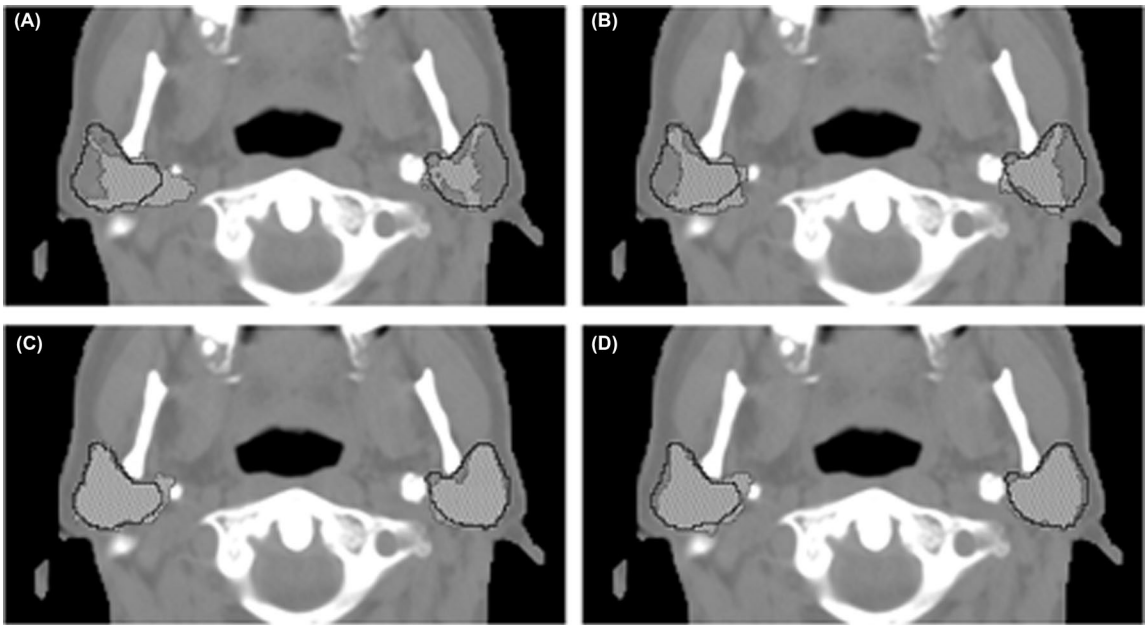

Figure 4: An example of warped parotid glands contours (dashed gray regions) versus original parotid glands contours (solid black) for a Log Domain Demons with $\sigma_{diff} = 0.5$ and $\sigma_{fluid} = 0.0$ (panel **A**), Log Domain Demons with $\sigma_{diff} = 1.5$ and $\sigma_{fluid} = 0.0$ (panel **B**), B-Splines with $\lambda = 10^{-3}$ (panel **C**), B-Splines with $\lambda = 0.05$ (panel **D**).

For nGTV, all methods but MIMvista exceeded 3 mm residual 3D center of mass distance; the B-Splines algorithm remained anyhow below the maximum voxel size (3 mm) in terms of surface distance, as depicted in Figure 3. Log Domain Diffeomorphic Demons with $\sigma_{diff} = 0.5$ provided significantly different results with respect to MIMvista, whereas all other methods were equivalent. Log Domain Diffeomorphic Demons residual errors as well as DSC mean volume overlap were not satisfactory, whereas both B-Splines regularizations and MIMvista DSC were consistently above 0.5.

Discussion and Conclusion

In this work, we analyzed two different approaches, *i.e.* parametric and non-parametric, to deformable image registration to assess the need of regularization in HN patients. The related performance was analyzed on 10 HN patients at different regularization levels in terms of image similarity. The selected regularization levels were then evaluated on 20 HN patients in terms of nGTV, mandible and parotid glands similarity, quantified by means of DSC, 3D center of mass distance and mean surface distance. We also compared the open-source parametric and non-parametric DIR implementations (B-Splines and Log Domain Diffeomorphic Demons) with a commercial software for clinical data analysis (MIMvista).

Figures 1A and 1C, in which the trend of RMSE and NMI median values are presented as a function of varying regularization parameters, show that for Log Domain Demons, no significant influence of fluid-like regularization (σ_{fluid}) exists on these statistical measures. The only effect was an increment of the computational costs (number of iterations) leading to hardly acceptable registration time for clinical application. The parameter σ_{diff} demonstrated instead to have a potential impact on final registration results. In particular, the absence of any sort of regularization caused results instability, as depicted by the higher inter-quartile range in Figure 1B and 1D, for $\sigma_{diff} < 0.5$. However, too high regularization causes an increase in median RMSE and a decrease in median NMI values ($\sigma_{diff} > 2.0$), thus suggesting un upper limit for σ_{diff}.

It is important to underline that the application of Log Domain Diffeomorphic Demons with $\sigma_{diff} = 0.5$ corresponds to a very small regularization, which may produce very similar results to non-regularized procedures. The contours suffer from the lower smoothing (Figure 4) much more than gray levels warping and their quantitative evaluation reflects the irregularities (Figure 3). Results demonstrated this hypothesis, underlining the significant difference between $\sigma_{diff} = 0.5$ and $\sigma_{diff} = 1.5$, despite the absence of a relevant difference after visual inspection of the images. As a general rule, when choosing the regularization parameter, the best trade-off between required accuracy and number of iterations should be investigated. However, for the Log Domain Diffeomorphic Demons, regularization parameters values $\sigma_{diff} = 1.5$ and $\sigma_{fluid} = 0.0$ provided the ideal results. Results obtained for B-Splines registration (Figure 2) suggest that a very low level of regularization is enough to improve the registration outcomes. Too large λ values, instead, hinder the understanding of small variations in patient anatomy. At increasing amount of regularization, in fact, the registration emphasizes a plastic behavior and finally tends to be rigid. The opportunity to reach comparable results in terms of RMSE and NMI for low level of regularization contribution, suggests that the increasing computational cost at increasing regularization levels should be taken into account when choosing the optimal value.

For B-Splines implementation, the absence of regularization brings along less variability than imposing a too large λ, because the latter may actually be a cause of divergence of the optimization algorithm. In fact, at each iteration, the update gets smaller due to the smoothing, causing the cost function to vary without reaching a satisfactory estimation of the deformation. On the other hand, the absence of smoothing can result into sharp discontinuities and/or small folding, which may cause a higher residual RMSE difference. According to our analysis, the best performing trade-off between amount of regularization and computational cost was obtained by setting λ equal to 0.001, leading to results comparable with those obtained by the commercial software.

Median and inter-quartile RMSE and NMI values calculated for the verification dataset are in agreement with those obtained during the testing phase at the best performing regularization levels, as their median and variability are in the same range of the first 10 patients, thus confirming that the regularization levels were reasonably tuned in the testing phase. Given that nGTV volumes are much smaller than all other structures, mispositioning is more likely to occur than for larger structures. Mandible is considered a very good control structure, as its high contrast is generally a good landmark for intensity based registration. In this case, both B-Splines and Log Domain Demons registration algorithms outperformed MIMvista, whose parameters are very likely established as a trade-off between bony anatomy accuracy and soft tissue match. Such a behavior for MIMvista is compatible with a highly regularized DIR algorithm, where the matching of high contrast structures can be penalized in favor of a regular vector field.

Our results show that open source DIR algorithms, with appropriate selection of regularization parameters, are able to achieve optimal results in HN patients, if compared to a commercial software package approved for clinical use. No appreciable differences were highlighted for soft tissue structures, whereas the tuning of regularization parameters provides adequate means to balance between bony anatomy

and soft tissue matching. As for Demons implementation, we proved that insufficient regularization has a significant impact on the overall DIR accuracy. Patient-specific regularization parameters optimization, as a function of the specific structures which are more relevant to be monitored at each treatment fraction, represents an essential feature for the finest optimization of DIR procedure in a clinical workflow. The price to be paid consists of an initial fine tuning of the regularization parameters, which may be performed during the first five weeks of treatment, which were identified as the period without significant changes over baseline in HN cancer treatment (31). The fine optimization process could benefit from the generation and application of synthetic deformation fields, simulating the most likely changes of the specific patient along the course of the treatment. Such a strategy may represent a promising way to realize adaptive radiotherapy in a computationally efficient, customizable and versatile approach.

Conflict of Interest

We certify that regarding this paper, no actual or potential conflicts of interests exist; the work is original, has not been accepted for publication nor is concurrently under consideration elsewhere, and will not be published elsewhere without the permission of the Editor and that all the authors have contributed directly to the planning, execution or analysis of the work reported or to the writing of the paper.

References

1. Barker, J. L. Jr, Garden, A. S., Ang, K. K., O'Daniel, J. C., Wang, H., Court, L. E., Morrison, W. H., Rosenthal, D. I., Chao, K. S., Tucker, S. L., Mohan, R., Dong, L. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. *Int J Radiat Oncol Biol Phys 59*, 960-970 (2004). DOI: 10.1016/j.ijrobp.2003.12.024
2. Vakilha, M., Hwang, D., Breen, S. L., Dawson, L. A., Ringash, J., Bayley, A., Kim, J., Cummings, B., O'Sullivan, B., Waldron, J. Changes in position and size of parotid glands assessed with daily cone-beam CT during image-guided IMRT for head and neck cancer: implications for dose received. *Int J Radia Oncol Biol Phys 69*, S438-439 (2007). DOI: 10.1016/j.ijrobp.2007.07.1600
3. Lee, C., Langen, K. M., Lu, W., Haimerl, J., Schnarr, E., Ruchala, K. J., Olivera, G. H., Meeks, S. L., Kupelian, P. A., Shellenberger, T. D., Mañon, R. R. Evaluation of geometric changes of parotid glands during head and neck cancer radiotherapy using daily MVCT and automatic deformable registration. *Radiother Oncol 89*, 81-88 (2008). DOI: 10.1016/j.radonc.2008.07.006
4. Jensen, A. D., Nill, S., Huber, P. E., Bendl, R., Debus, J., Münter, M. W. A clinical concept for interfractional adaptive radiation therapy in the treatment of head and neck cancer. *Int J Radiat Oncol Biol Phys 82*, 590-596 (2012). DOI:10.1016/j.ijrobp.2010.10.072
5. Schwartz, D. L., Dong, L. Adaptive radiation therapy for head and neck cancer – can an old goal evolve into a new standard?. *J Oncol 2011* (2001). DOI:10.1155/2011/690595
6. Peroni, M., Ciardo, D., Spadea, M. F., Riboldi, M., Comi, S., Alterio, D., Baroni, G., Orecchia, R. Automatic segmentation and online virtualCT in head-and-neck adaptive radiation therapy. *Int J Radiat Oncol Biol Phys 84,* e427-e433 (2012). DOI: 10.1016/j.ijrobp.2012.04.003
7. Paganelli, C., Peroni, M., Baroni, G., Riboldi, M. Validation of deformable registration in adaptive radiation therapy with scale invariant feature transform. *9th IEEE International Symposium on Biomedical Imaging 2012*, 680-683 (2012). DOI: 10.1109/ISBI.2012.6235639
8. Peroni, M., Spadea, M. F., Riboldi, M., Baroni, G., Chen, G. T. Y., Sharp, G. C. Validation of an automatic contour propagation method for lung cancer 4D adaptive radiation therapy. *Proceedings – 2009 IEEE International Symposium on Biomedical Imaging, From Nano to Macro, ISBI 2009, Art. no. 5193241,* 1071-1074 (2009). DOI: 10.1109/ISBI.2009.5193241
9. Elstrøm, U. V., Wysocka, B. A., Muren, L. P., Petersen, J. B., Grau, C. Daily kV cone-beam CT and deformable image registration as a method for studying dosimetric consequences of anatomic changes in adaptive IMRT of head and neck cancer. *Acta Oncol 49*, 1101-1108 (2010). DOI: 10.3109/0284186X.2010.500304
10. Castadot, P., Lee, J. A., Parraga, A., Geets, X., Macq, B., Grégoire, V. Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors. *Radiother Oncol 89*, 1-12 (2008). DOI: 10.1016/j.radonc.2008.04.010
11. Kashani, R., Hub, M., Balter, J. M., Kessler, M. L., Dong, L., Zhang, L., Xing, L., Xie, Y., Hawkes, D., Schnabel, J. A., McClelland, J., Joshi, S., Chen, Q., Lu, W. Objective assessment of deformable image registration in radiotherapy: a multi-institution study. *Med Phys 35*, 5944-5953 (2008). DOI: 10.1118/1.3013563
12. Brock, K. K. Results of a multi-institution deformable registration accuracy study (MIDRAS*). Int J Radiat Oncol Biol Phys 76*, 583-596 (2010). DOI: 10.1016/j.ijrobp.2009.06.031
13. Murphy, K., van Ginneken, B., Reinhardt, J. M., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G. E., Garcia, V., Vercauteren, T., Ayache, N., Commowick, O., Malandain, G., Glocker, B., Paragios, N., Navab, N., Gorbunova, V., Sporring, J., de Bruijne, M., Han, X., Heinrich, M. P., Schnabel, J. A., Jenkinson, M., Lorenz, C., Modat, M., McClelland, J. R., Ourselin, S., Muenzing, S. E., Viergever, M. A., De Nigris, D., Collins, D. L., Arbel, T., Peroni, M., Li, R., Sharp, G. C., Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Smeets, D., Loeckx, D., Song, G., Tustison, N., Avants, B., Gee, J. C., Staring, M., Klein, S., Stoel, B.C., Urschler, M., Werlberger, M., Vandemeulebroucke, J., Rit, S., Sarrut, D., Pluim, J. P. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE Trans Med Imaging 30*, 1901-1920 (2011). DOI: 10.1109/TMI.2011.2158349
14. Al-Mayah, A., Moseley, J., Hunter, S., Velec, M., Chau, L., Breen, S., Brock, K. Biomechanical-based image registration for head and neck radiation treatment. *Phys Med Biol 55*, 6491-6500 (2010). DOI: 10.1088/0031-9155/55/21/010
15. Kostelec, P. J., Weaver, J. B., Healy, D. M. Multiresolution elastic image registration. *Med Phys 25*, 1593-1604 (1998). DOI: S0094-2405(98)01209-7
16. Peroni, M., Golland, P., Sharp, G. C., Baroni G., Ranking of stopping criteria for log domain diffeomorphic demons application in clinical radiation therapy. *Conf Proc IEEE Eng Med Biol Soc 2011*, 4884-4887 (2011). DOI: 10.1109/IEMBS.2011.6091210
17. Shusharina, N., Sharp, G. Analytic Regularization for Landmark-based Image Registration. *Phys Med Biol 57*, 1477-1498 (2012). DOI: 10.1088/0031-9155/57/6/1477
18. Xing, L., Siebers, J., Keall, P. Computational challenges for image-guided radiation therapy: framework and current research. *Semin Radiat Oncol 17*, 245-57 (2007). DOI: 10.1016/j.semradonc.2007.07.004
19. Noblet, V., Heinrich, C., Heitz, F., Armspach, J. P. Retrospective evaluation of a topology preserving non-rigid registration method. *Med Image Anal 10*, 366-384 (2006). DOI: 10.1016/j.media.2006.01.001
20. Fallone, B. G., Rivest, D. R., Riauka, T. A., Murtha, A. D. Assessment of a commercially available automatic deformable registration system. *J Appl Clin Med Phys 11*, (2010). DOI: 10.1120/jacmp.v11i3.3175

21. Yang, D., Brame, S., El Naqa, I., Aditya, A., Wu, Y., Goddu, S. M., Mutic, S., Deasy, J. O., Low, D. A. Technical note: DIRART–A software suite for deformable image registration and adaptive radiotherapy research. *Med Phys 38*, 67-77 (2011). DOI: 10.1007/978-3-642-03474-9_236
22. Shackleford, J. A., Kandasamy, N., Sharp, G. C. On developing B-spline registration algorithms for multi-core processors. *Phys Med Biol 55*, 6329-6351 (2010). DOI: 10.1088/0031-9155/55/21/001
23. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage 45*, S61-S72 (2009). DOI: 10.1016/j.neuroimage.2008.10.040
24. Piper, J. W. Evaluation of an intensity-based free-form deformable registration algorithm. *Med Phys 34*, 2353-2354 (2007). DOI: 10.1118/1.2760445
25. Thirion, J.-P. Non-rigid matching using demons. *Computer Vision and Pattern Recognition, CVPR'96*, 245-251 (1996). DOI: 10.1016/S1361-8415(98)80022-4
26. Liu, D. C., Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math Program 45*, 503-528 (1989). DOI: 10.1007/BF01589116
27. Schreibmann, E., Chen, G. T., Xing, L. Image interpolation in 4D CT using a BSpline deformable registration model. *Int J Radiat Oncol Biol Phys 64*, 1537-1550 (2006). DOI: 10.1016/j.ijrobp.2005.11.018
28. Dice, L. R. Measures of the amount of ecologic association between species. *Ecology 26*, 297-302 (1945). DOI: 10.2307/1932409
29. Danielsson, P.-E. Euclidean distance mapping. *Computer Graphics and Image Processing 14*, 227-248 (1980). DOI: 10.1016/0146-664-X(80)90054-4
30. Dru, F., Fillard, P., Vercauteren, T. An ITK implementation of the symmetric log-domain diffeomorphic demons algorithm. *Insight Journal – 2009 January-June 2009* (2010).
31. Ricchetti, F., Wu, B., McNutt, T., Wong, J., Forastiere, A., Marur, S., Starmer, H., Sanguineti, G. Volumetric change of selected organs at risk during IMRT for oropharyngeal cancer. *Int J Radiat Oncol Biol Phys 80*, 161-8 (2011). DOI: 10.1016/j.ijrobp.2010.01.071

Received: August 1, 2012; Revised: November 9, 2012; Accepted: December 21, 2012