Stopping Criteria for Log-Domain Diffeomorphic Demons Registration

Citation: Peroni, M. et al. "Stopping Criteria for Log-Domain Diffeomorphic Demons Registration." Technology in Cancer Research & Treatment 15, 1 (February 2016): 77–90 © 2014 The Authors

As Published: http://dx.doi.org/10.7785/tcrtexpress.2013.600269

Publisher: SAGE Publications

Persistent URL: http://hdl.handle.net/1721.1/110986

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Technology in Cancer Research and Treatment
ISSN 1533-0346
Volume 15 Number 1 February 2016
2013 August 31. Epub ahead of print.

Stopping Criteria for Log-Domain Diffeomorphic Demons Registration: An Experimental Survey for Radiotherapy Application

M. Peroni, Ph.D.[1]*
P. Golland, Ph.D.[2]
G. C. Sharp, Ph.D.[3,4]
G. Baroni, Ph.D.[1,5]

www.tcrt.org
DOI: 10.7785/tcrtexpress.2013.600269

A crucial issue in deformable image registration is achieving a robust registration algorithm at a reasonable computational cost. Given the iterative nature of the optimization procedure an algorithm must automatically detect convergence, and stop the iterative process when most appropriate. This paper ranks the performances of three stopping criteria and six stopping value computation strategies for a Log-Domain Demons Deformable registration method simulating both a coarse and a fine registration. The analyzed stopping criteria are: (a) velocity field update magnitude, (b) mean squared error, and (c) harmonic energy. Each stopping condition is formulated so that the user defines a threshold ε, which quantifies the residual error that is acceptable for the particular problem and calculation strategy. In this work, we did not aim at assigning a value to ε, but to give insights in how to evaluate and to set the threshold on a given exit strategy in a very popular registration scheme. Experiments on phantom and patient data demonstrate that comparing the optimization metric minimum over the most recent three iterations with the minimum over the fourth to sixth most recent iterations can be an appropriate algorithm stopping strategy. The harmonic energy was found to provide best trade-off between robustness and speed of convergence for the analyzed registration method at coarse registration, but was outperformed by mean squared error when all the original pixel information is used. This suggests the need of developing mathematically sound new convergence criteria in which both image and vector field information could be used to detect the actual convergence, which could be especially useful when considering multi-resolution registrations. Further work should be also dedicated to study same strategies performances in other deformable registration methods and body districts.

Key words: Optimization; Deformable registration; Demons; Stopping criteria.

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italia
[2]Computer Science and Artificial Intelligence Laboratory, Electrical Engineering & Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[3]Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA
[4]Harvard Medical School, Boston, MA
[5]Bioengineering Unit, Centro Nazionale di Adroterapia Oncologica, Pavia, Italy

Introduction

In recent years, image guidance has gained popularity in radiation therapy for treatment geometry verification (1, 2). From setup uncertainties, the focus of research has moved to the compensation of anatomical deformations induced in patients both by physiological motion (*e.g.* bladder filling, respiratory motion) and by therapy response (*e.g.* tumor growth or regression), aimed at implementing adaptive strategies in radiation oncology (3). In addition, inter-subject probabilistic segmentation of relevant anatomical structures has become more important (4). Several algorithms for deformable image registration (DIR)

Abbreviation: CT: Computed Tomography.

*Corresponding author:
Marta Peroni, Ph.D.
E-mail: marta.peroni@psi.ch

have been proposed to address these issues, as iterative optimization process that recovers non-rigid deformations between a fixed or reference image I_f and a moving or target image I_m (5). Amongst these, one of the most popular is the demons algorithm, introduced by Thirion (6). The main limitations to the clinical use of DIR are the computational cost and the robustness of the parameter set. Several efforts have been made towards algorithm optimization with remarkable results in reducing the time needed for registration (7, 8), also thanks to GPU implementation (9, 10). A major advance towards robust and time convenient DIR in clinical applications would be brought in by optimizing the stopping condition, preventing unnecessary extra iterations in the optimization. Schmidt-Richtberg *et al*. (11) used the image similarity metric directly as an escape condition, without analyzing the vector field update (12). In the original log-diffeomorphic demons paper, Vercauteren *et al.* (13) did not provide iteration related stopping criteria, but assessed algorithm performance in terms of harmonic energy and number of vector field voxels with negative Jacobian elements. In this work, experimental evaluation was carried out using five synthetic non-rigid deformations applied to clinical quality CT image volumes of a dosimetry phantom and three CT scans of head and neck patients. Overall, six different strategies of convergence curve description were assessed. We ranked each stopping rule in terms of speed, convergence and deformation recover capability in synthetic controlled cases and trace guidelines for their deployment in a log-diffeomorphic demons deformable image registration.

Materials and Methods

Description of the Demons Algorithm

Since Thirion's paper (6), the so-called demons registration has gained popularity for intra-modality image registration and several variations have been proposed. Vercauteren *et al.* (13) described a Log-Domain Diffeomorphic Demons algorithm that features the advantage of optimizing a diffeomorphic transformation within a computationally efficient framework. The optimization of the cost function occurs in two different steps. The algorithm minimizes the cost function:

$$E(s,c;I_f,I_m)=\frac{1}{\sigma_i^2}\,\mathrm{simil}(I_f,I_m\circ c)+\frac{1}{\sigma_x^2}\,\mathrm{dist}(s,c)+\frac{1}{\sigma_T^2}\,\mathrm{reg}(s) \quad [1]$$

Where s and c are the transformation calculated to bring I_m and I_f into alignment and a hidden variable for point correspondences (15) respectively. The term $\mathrm{simil}(I_f,I_m\circ c)=\frac{1}{2\left|\Omega P\right|}\sum_{p\in\Omega P}\left|I_f-I_m\circ c\right|^2$ is the image disparity measure, Ω_P defines the overlap region between I_f and I_m, $\mathrm{dist}(s,c)=\left\|s-c\right\|^2$ quantifies the similarity between the deformations s and c, reg(s) is the degree of smoothness of the deformation, and σ_i^2, σ_x^2, σ_T^2 balance the contribution of the three terms of the cost function. The last term penalizes large gradients, and models prior knowledge about the transformation to recover. The transformation to be recovered is represented by a smooth and stationary velocity field v, according to the equation $s=\exp(v)$ which is a diffeomorphic deformation. At each iteration a field u is than computed for an update in the form $s\circ\exp(u)$ or $\exp(v)\circ\exp(u)$. We essentially look for a velocity field $V(v,\varepsilon u)$, such that

$$\exp(V(v,\varepsilon u))\approx\exp(v)\circ\exp(\varepsilon u) \quad [2]$$

where ε only underlines that we look for an a small εu (*i.e.* the update) but an arbitrary v (*i.e.* the full transform). As firstly shown in (16), a valid approximation of $V(v,\varepsilon u)$ can be formulated using the Backer-Campbell-Hausdorff formula:

$$V(v,\varepsilon u)=v+\varepsilon u+\frac{1}{2}[v,\varepsilon u]+\frac{1}{12}[v,[v,\varepsilon u]]+0\left(\left\|\varepsilon u\right\|^2\right) \quad [3]$$

where $[v,u](p)=\mathrm{Jac}(v)(p)\cdot u(p)+Jac(u)(p)\cdot v(p)$ at each point p.

We used the efficient second-order symmetric optimization scheme proposed by Vercauteren *et al.* (17), which further parameterizes the deformation in the log-domain, with inverse $s^{-1}(x)=\exp(-v)(x)$. In the symmetric scheme proposed by Vercautern *et al.* (17)), a direct and inverse energy are calculated as:

$$E_{\mathrm{forward}}(I_f,I_m,s,u_{\mathrm{forward}})=\mathrm{simil}\left(I_f,I_m\circ s\circ\exp(u_{\mathrm{forward}})\right)+\left\|u_{\mathrm{forward}}\right\|^2 \quad [4]$$

$$E_{\mathrm{backward}}(I_m,I_f,s^{-1},u_{\mathrm{backward}})=\mathrm{simil}\left(I_m,I_f\circ s^{-1}\circ\exp(u_{\mathrm{backward}})\right)+\left\|u_{\mathrm{backward}}\right\|^2$$

u_{forward} and u_{backward} are further defined as the corresponding demons forces updates. The optimal s is than found as

$$s_{\mathrm{opt}}=\arg\min_s\left(E_{\mathrm{forward}}+E_{\mathrm{backward}}\right) \quad [5]$$

where E_{backward} simply denotes the inverse of the transformation s. Therefore, the algorithm proceeds as follows:

1. Calculate u_{forward} and u_{backward} that minimize $E_{\text{forward}}(I_f, I_m, s, u_{\text{forward}})$ and $E_{\text{backward}}(I_m, I_f, s^{-1}, u_{\text{backward}})$ respectively.

2. Smooth with a Gaussian kernel K so that $u = \frac{1}{2} K \times (u_{\text{forward}} - u_{\text{backward}})$.

3. Update v (Backer-Campbell-Hausdorff formula)

Stopping Criteria

Comparable to any other iterative optimization, the result of any deformable registration greatly depends on the selected stopping criteria. The most basic strategy is to terminate the algorithm after a predetermined number of iterations. Although very simple, this is only weakly related to the actual convergence. Ideally, we would like to design conditions based on the terms in equations 1 to 3. One first criteria is based on image intensities disparity between I_f and I_m at the current iteration t with respect to a previous time-point. Since our experiments were using mono-modal images, we used Mean Squared Error (MSE). A sufficiently small threshold ε_{MSE}, quantifying the error reduction between iterations, can be defined, such that if

$$\frac{\text{MSE}_{\text{current}} - \text{MSE}_{\text{previous}}}{\text{MSE}_{\text{previous}}} > \varepsilon_{\text{MSE}} \qquad [6]$$

the registration algorithm is terminated. Looking at the left side of the equation, values close to zero would indicate plateau or convergence, whereas positive ratio would indicate divergence. In this case, the registration can be allowed to continue provided that non-convergence stays below the user defined threshold ε_{MSE}.

An alternative convergence criterion is based on the Harmonic Energy (HE). HE is defined as the average over all voxels of the squared Frobenius norm of the vector field Jacobian and is correlated to the smoothness of the field. If the registration is converging, $\text{HE}_{\text{current}}$ would than be greater than $\text{HE}_{\text{previous}}$ and as such the escape condition becomes

$$\frac{\text{HE}_{\text{current}} - \text{HE}_{\text{previous}}}{\text{HE}_{\text{previous}}} < \varepsilon_{\text{HE}} \qquad [7]$$

The last condition we examined is based on Quantity of Update (QU) between iterations. QU is defined by Yang *et al.* (8) as:

$$\text{QU}_m = \frac{\sum |dr_m|}{\sum |r_{m-1}|} \qquad [8]$$

where m is the iteration identifier, r_{m-1} is the field at a previous iteration and dr_m is the update field at current iteration. As QU_m decreases with convergence, we can compute:

$$\left| \frac{\text{QU}_m - \text{QU}_{m-1}}{\text{QU}_{m-1}} \right| < \varepsilon_{\text{INCR}} \qquad [9]$$

We expect that at convergence, $\text{QU}_m \to 0$ thus decreasing the ratio value below the threshold. As before, we extend this quantity to the velocity field, enforced to be stationary in log-domain demons (13). For each of the considered convergence metrics, the stopping criteria requires the comparison with a previous stopping condition value t, *i.e.* chosen metric at a previous iteration. Table I summarizes the calculation strategies, which were evaluated in terms of convergence speed.

Table I

Current, previous and calculated stopping condition value (t) for MSE, HE and quantity of update. The current and previous t were tested in this work in terms of convergence speed. These formulations are used in the final computation of the chosen stopping criteria. For conditions from D to F and stopping criteria HE, the minimum shall be substituted with a maximum, given HE definition.

	Current value	Previous value	Computed value
A	t	$t-1$	$\frac{t-(t-1)}{t-1}$
B	t	$t-3$	$\frac{t-(t-3)}{t-3}$
C	t	$t-5$	$\frac{t-(t-5)}{t-5}$
D	t	Min (last 6 iterations)	t-min (last 6 iterations)
E	Min (t, $t-1$, $t-2$)	Min ($t-1$, $t-2$, $t-3$)	Min (t, $t-1$, $t-2$) – min ($t-1$, $t-2$, $t-3$)
F	Min (t, $t-1$, $t-2$)	Min ($t-3$, $t-4$, $t-5$, $t-6$)	Min (t, $t-1$, $t-2$) – min ($t-3$, $t-4$, $t-5$, $t-6$)

Experiments

First experimental tests were performed on images taken on a RANDO® (The Phantom Laboratory, Salem, NY) dosimetry phantom, by means of clinical CT scanner (GE Medical System Light Speed, Fairfield, CT), using supine setup and clinical acquisition protocols. The acquired volume was 512 × 512 × 123 voxels, with [0.94, 0.94, 3] mm element spacing. Registrations were carried out on data coming from three head and neck patients acquired with the same CT scanner and image resolution of the phantom. We chose one patient with a moderate degree of image artifacts (patient 1), one with very few reconstruction artifacts (patient 3) and one, where severe image artifacts due to dental fillings were present (patient 2). Cross-section images for each of the selected patients are reported in Figure 1.

When running the Log-Domain Diffeomorphic Demons registration, we used symmetric computation of the gradient and Backer-Campbell-Hausdorff (BCH) 2-terms expansion.

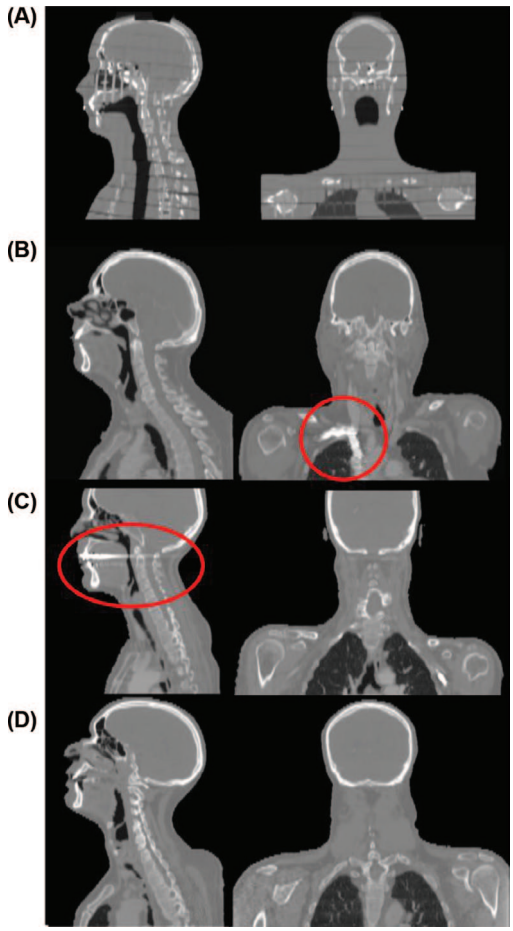

Figure 1: An example slice per dataset used in this work. Panel **A** shows one slice of RANDO®, in which we note the several layers it is composed of. In the red circle, chemotherapy tube for patient 1 (Panel **B**). Note the metal artifacts in patient 2 (red circle, Panel **C**).

Table II

Chosen gaussian parameters. Parameters of the gaussian distribution used to generate five artificial non-rigid 128 × 128 × 128 vector fields. Each vector field is the sum of three gaussians with different centers, standard deviation, and relative weight.

Test number		Application voxel	Radius [# voxels]	Relative weight	Standard deviation [# voxels]
1	*x*	[30 30 30]	15	5	12
	y	[50 50 50]	30	5	18
	z	[80 80 80]	10	100	36
2	*x*	[60 60 60]	25	1	12
	y	[50 50 60]	30	1	18
	z	[60 60 55]	45	100	36
3	*x*	[60 60 60]	25	10	12
	y	[50 50 60]	30	10	18
	z	[60 60 55]	45	1000	36
4	*x*	[60 60 30]	40	0.15	10
	y	[30 30 60]	20	0.15	15
	z	[60 60 90]	30	0.15	22
5	*x*	[60 60 90]	30	0.1	22
	y	[60 60 90]	30	0.1	22
	z	[60 60 90]	30	0.1	22

We only smoothed the velocity field, with a Gaussian filtering with σ equal to 1.5 pixels. The maximum update step length was 2 pixels and no histogram matching was performed.

Ground truth deformations were generated from five artificial non-rigid vector fields, simulating the distribution of left-right, anterior-posterior and inferior-superior components with a smooth, localized Gaussian deformation. The chosen Gaussians were centered in different voxels in a matrix of size 128 × 128 × 128 and featured variable standard deviation and amplitude, as reported in Table II. We re-sampled the Gaussian deformations on the image grid, thus generating a continuous and smooth vector field which represented the reference for the testing of Log-Domain Diffeomorphic Demons image registration algorithm. With this vector field, we obtained five warped phantoms (I_{fph}) (Figure 2) and five warped patient CTs (I_{fpa}). Jacobian minimum and maximum values are shown in Table III, while their overall distribution with respect to phantom image is shown in Figure 3.

Table III

Minimum and maximum Jacobian values for each of the synthetic deformation field. Note in particular that no folding (*i.e.* negative Jacobian) is reported.

Deformation field	Minimum Jacobian	Maximum Jacobian
1	0.83	1.13
2	0.96	1.05
3	0.81	1.27
4	0.98	1.02
5	0.95	1.06

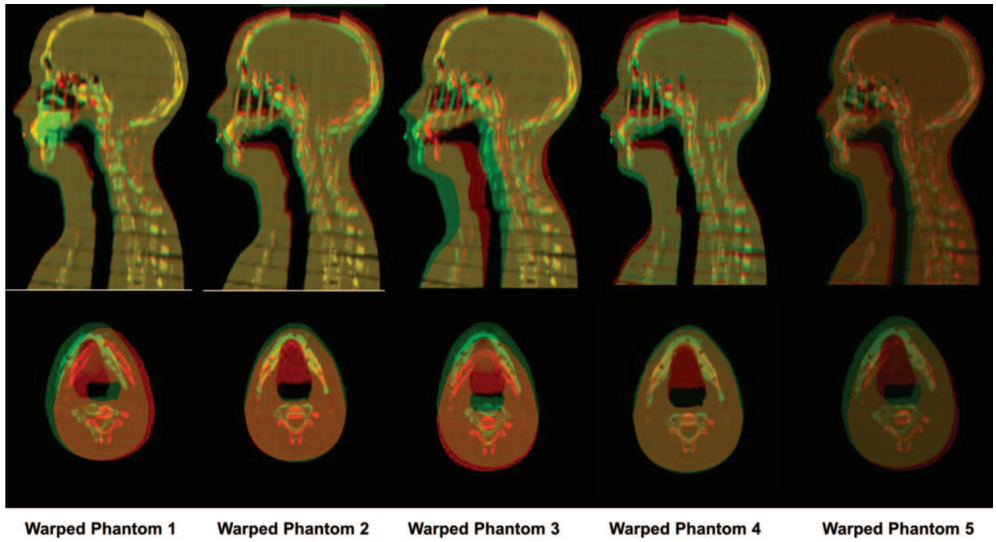

Figure 2: Overlay of warped (green) and original (red) phantom. The same deformations were used to warp patient CT scans.

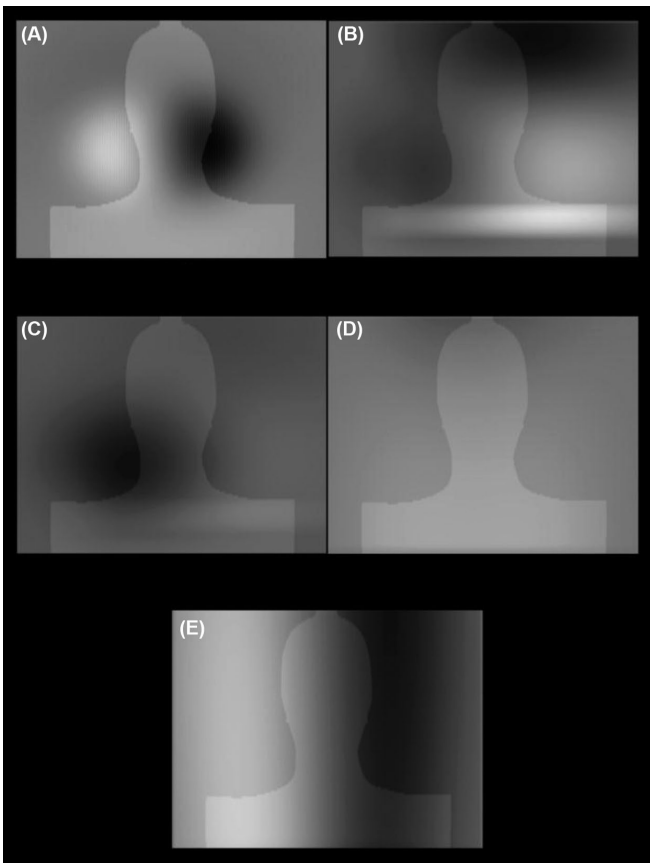

Figure 3: Coronal view of the Jacobian of the synthetic deformation fields. Black indicates minimum Jacobian value, whereas larger values tend to white colour. Overlayed is the mask of the phantom image, for a better understanding of the location of minimum and maximum Jacobian.

We registered the original image volume (I_{mph}) to each of the five warped phantoms (I_{fph}). We applied the same strategy also for the patients data, registering the original image (I_{mpa}) to the corresponding five warped volumes (I_{fpa}). For each image pair, we ran 500 iterations of Log-Domain Diffeomorphic Demons algorithm at one-eighth of the image resolution (*i.e.* a coarse registration stage) and 300 iterations at full image resolution (*i.e.* a fine registration stage) separately. No multi-resolution registrations were anyhow performed. Initialization of the registration was done aligning the center of the image volumes, but a rigid registration was not needed for our purely deformable synthetic experiments.

All experiments were conducted on a quad-core workstation (3.2 GHz each) equipped with 16GB RAM and 1.5GB dedicated graphical memory. All software was implemented in C++ and the statistical analysis is conducted in Matlab 7.11.

When computing the convergence values, we excluded background (air) voxels from the computation. At each iteration m, we computed the stopping condition t according to the strategies reported in Table I.

Absolute registration quality was calculated as the vector length of the residual deformation at each iteration, *i.e.* subtracting the current deformation field from the corresponding ground-truth displacements. Given the image resolution, we defined an acceptance threshold on the vector length residual error equal half a voxel both at one-eight resolution and at full resolution. This corresponds to 1.875 mm for the coarse and to 0.469 mm for the fine registration respectively. For each data set, we assessed the number of iterations required by the algorithm to reach this level of accuracy (v_{required}). We call t_{critical_i} the value of a stopping criteria at v_{required}, with $i = [1;N]$ ($N = 5$), number of analyzed deformation. Ideally, t_{critical_i} should be the same for

all registration cases (phantom and patients data), so that the algorithm always stops exactly at the error acceptance threshold. For phantoms and patients separately and for each stopping criteria, we further defined a $t_{optimal}$ amongst the i different $t_{critical}$ as the one closest to zero. We ranked the performance of the stopping criteria in terms of number of extra iterations needed to reach the expected $t_{optimal}$ for each escape condition separately. In addition, convergence speed and stability were considered in the evaluation of the tested stopping conditions.

Results

Convergence Properties

Figure 4 illustrates the median among the five phantoms data of Mean Squared Error, Harmonic Energy and Quantity of Update percent values calculated at each iteration m according to the strategies listed in Table I and as defined in paragraph II.B in the coarse registration case. No threshold ε was defined for this analysis. For example, 0% variation would suggest no difference between current and previous stopping condition value, whereas 100% indicates that current is the double of the previous one. All strategies identified convergence for the assigned registration problem. However some exhibited a lower computational efficiency with respect to others. For example, strategies B, C and F require more iterations than E. Moreover, B and D require a longer initialization (respectively 3 and 5 iterations) to achieve a reasonable metric value, thus possibly increasing iterations unnecessarily. Strategy A is susceptible to early termination, as only the immediately previous iteration is considered. Strategy D is a good candidate for Mean Squared Error (Figure 4A), and Quantity of Update (Figure 4C) but can be rather fluctuating for Harmonic Energy (Figure 4B). Conversely, strategy E seems to capture the best tradeoff between speed of convergence and stability, in particular if we consider Harmonic Energy. If we now look at the behavior of the different calculation strategies at full resolution, we note very similar results (Figure 5), being E the most favorable calculation strategy for all the analyzed metrics. Note the trend of D for Harmonic Energy (Figure 5B) in the iterations 10 to about 25, in which the disagreement with calculation E is rather evident. The behavior of the graphs is very similar for patient study and therefore we do not report it. In the following experiments, we used strategy E, based on the results of this analysis.

Deformation Recovery Capability

Figures 6 and 7 show the plot of the median, 25th and 75th percentile for the residual of the displacement field of the five phantom and three patient cases, respectively, for coarse registration. It is evident that more iterations do not necessarily always imply better convergence, and this reinforces the concept that a stopping condition is needed. Differences in the speed of deformation field recovery were found between phantom and patient cases (Figures 6 and 7). Comparing phantom and patient 1 data for the fifth displacement field (see Table II), we observe that residual error reaches 2 mm in 35 iterations and 15 iterations, respectively. Figure 8 reports the trends of the deformation field recovery (median residual deformation) for phantom (panel A) and patient 1 (panel B). In panel C, we show the plots of the median residuals for patient 2 and 3, separately for each used synthetic transform. Note that, for patient 1, we were not able to recover the deformation with the desired level of accuracy only once, whereas in patient 2 and 3, eight of the registrations did not actually converge.

For phantom case, the results on the metric performances are reported in Tables IV and V. Table IV illustrate the different stopping condition values at $v_{required}$ (1.875 mm and 0.469 mm) for coarse and fine registration respectively. The chosen critical value ($t_{optimal}$) is highlighted in bold. The number of extra iterations required to reach the stopping condition value at $v_{required}$ are listed in Table V. At coarse resolution, the stopping criteria that required fewer iteration were Harmonic Energy and Mean Squared Error. Conversely, Quantity of Update required more iterations for the given threshold value. At full image resolution, instead, MSE was the clear winner exit strategy followed by Quantity of Update and Harmonic Energy.

Out of fifteen tested coarse registrations on the three patient data, the median residual deformation was not smaller than 1.875 mm in two cases, which were therefore excluded from the study. The overall results are presented in Table VI. Harmonic Energy outperformed Mean Squared Error in terms of capability to detect the iteration in which the optimization should stop. However, Quantity of Update was not robust enough against the lower image quality of patient cases. In patient 2 data (most severe artifacts), Mean Squared Error and Harmonic Energy performed much better than Quantity of Update.

At high resolution instead, we had a larger number of cases not reaching accuracy level (9/15), as already reported in Figure 8. The patient with larger number of successful registrations was patient 1, whereas 300 iterations at full resolution were able to recover only deformation 5 (*i.e.* one of the largest) for patients 2 and 3 (see Figure 8C). In Table VII, we report the metric values at $v_{required}$ together with the $t_{optimal}$ (in bold) chosen for this patient. In the same table, we also show the number of extra iterations needed to reach the chosen $t_{optimal}$. Also for this case, Mean Squared Error demonstrated to be the most suitable stopping condition for registrations at full resolution.

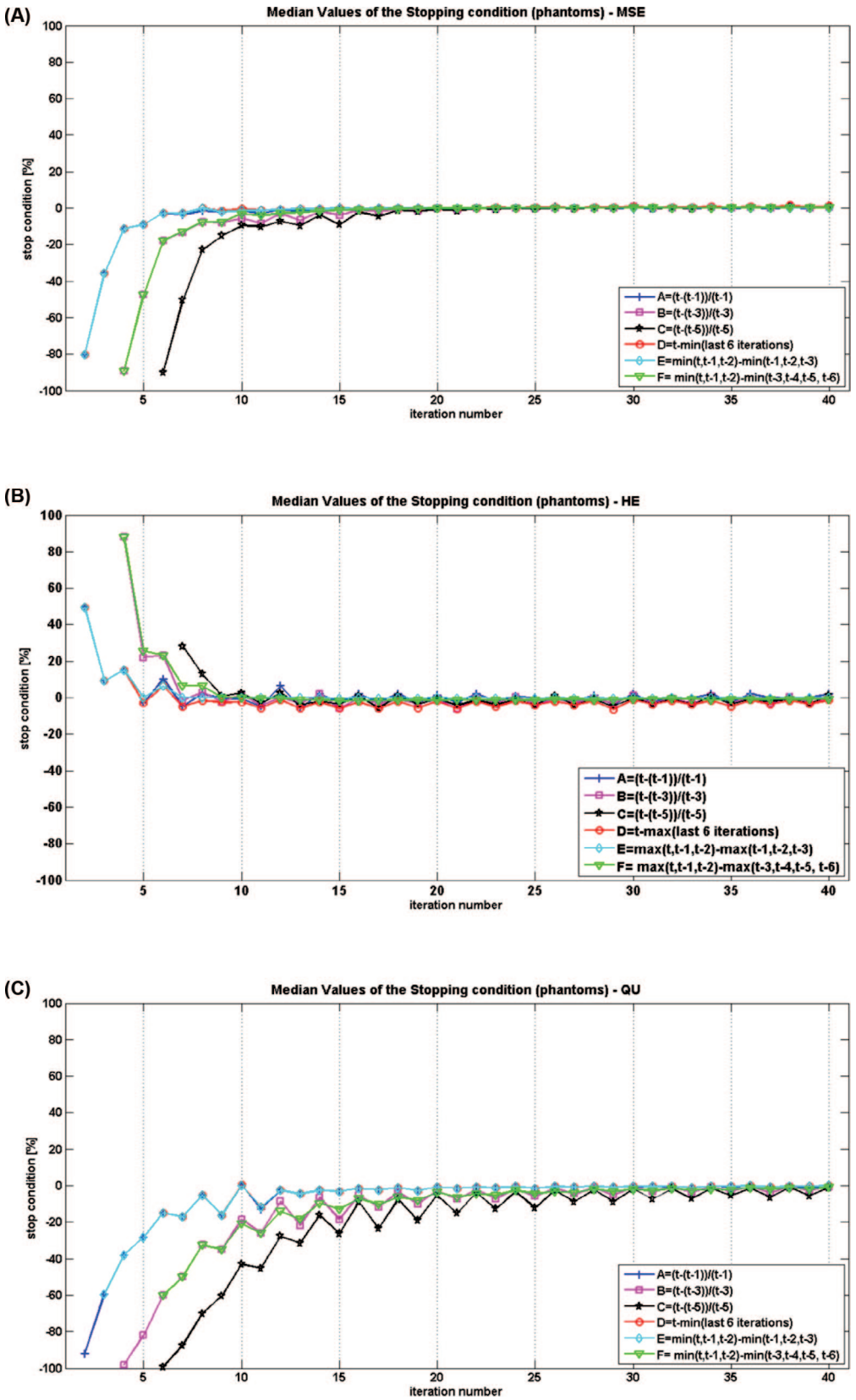

Figure 4: Median stopping condition value at each iteration for the phantom study over the five deformation for MSE, HE and QU for coarse registration. Better performances of a stopping criteria are identified by fewer iterations used to reach convergence.

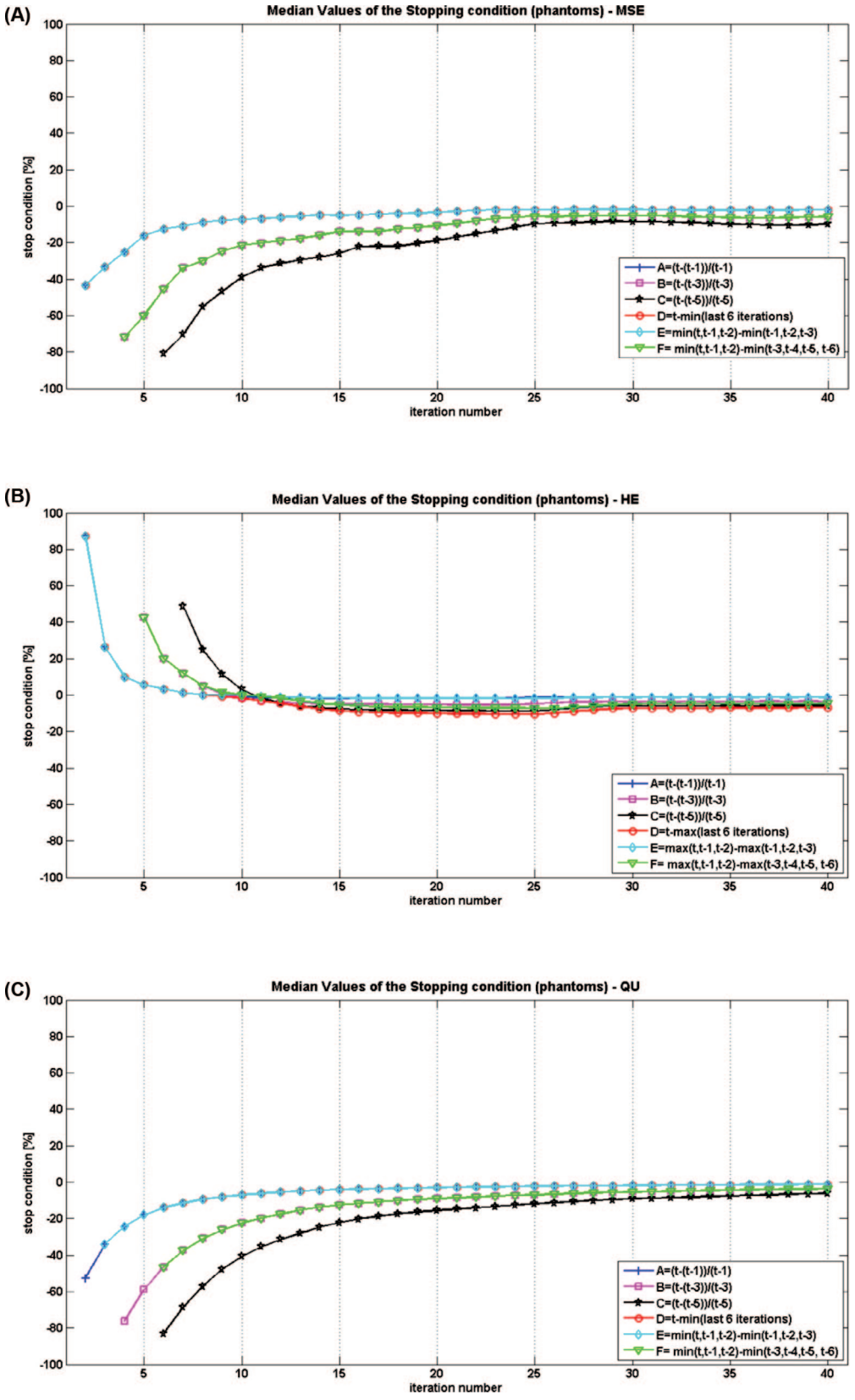

Figure 5: Median stopping condition value at each iteration for the phantom study over the five deformation for MSE, HE and QU in the fine registration case. Better performances of a stopping criteria are identified by fewer iterations used to reach convergence.

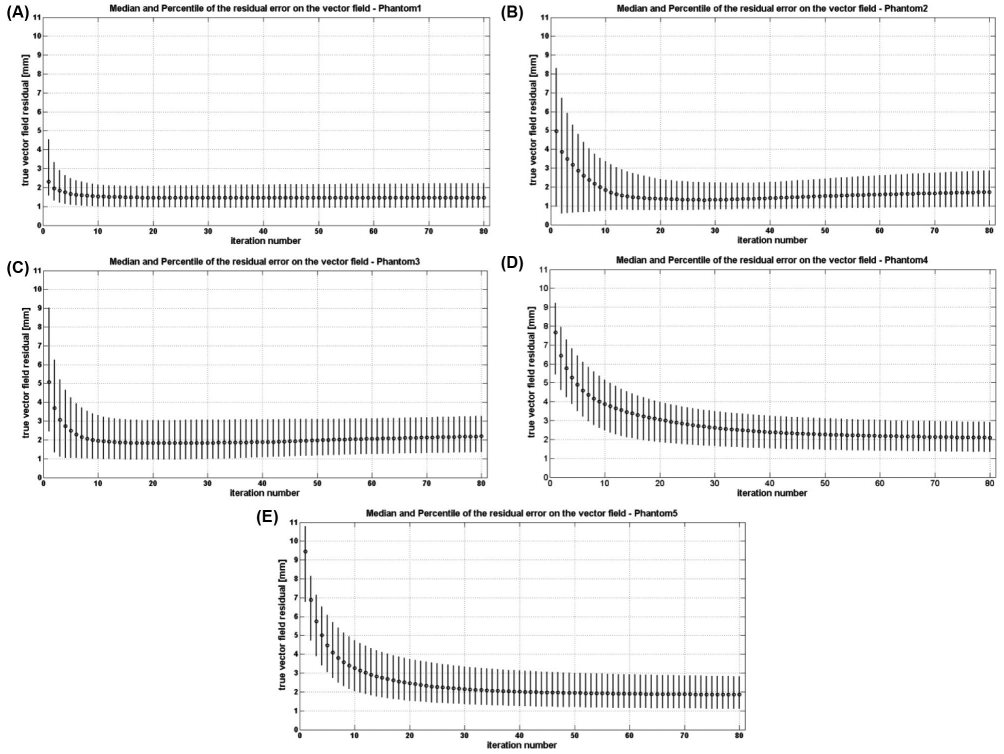

Figure 6: Residual errors on the vector field for phantom study at coarse resolution. Graphs show median, 25th and 75th percentile for each analyzed deformation (from **A** to **E**). The ideal stopping condition can robustly identify the minimum plateau (*i.e.* convergence) in the fewest iterations.

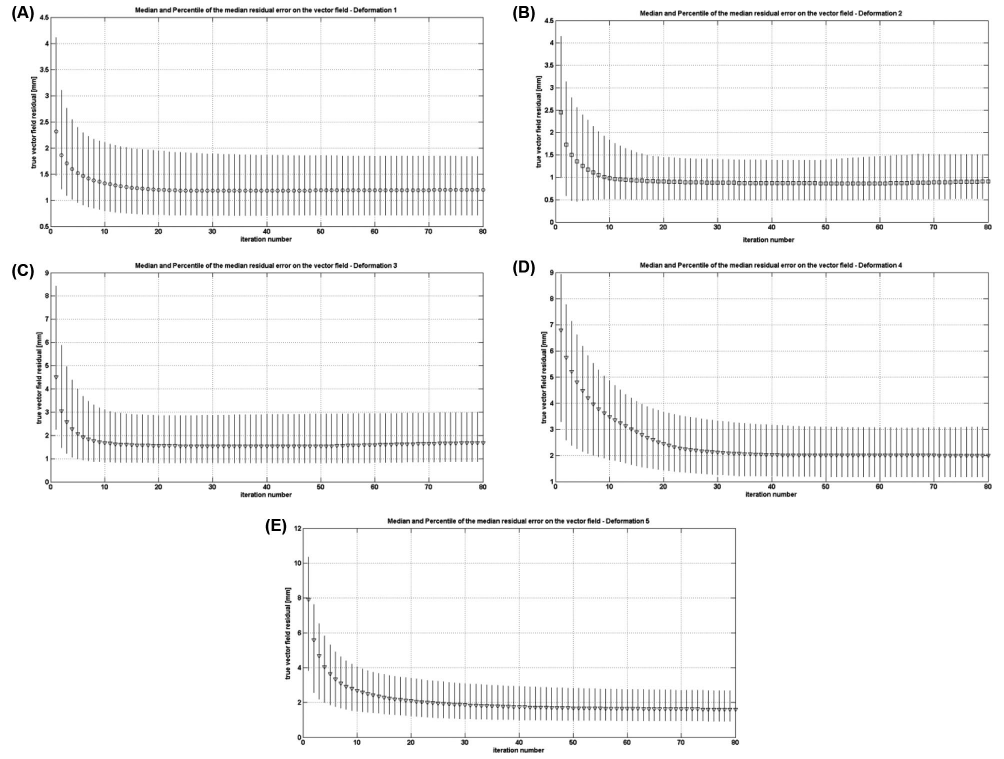

Figure 7: Median, 25th and 75th percentile of the residual error distribution over the three patients for each simulated deformation at coarse resolution. Trend is similar to phantom, but slower in convergence.

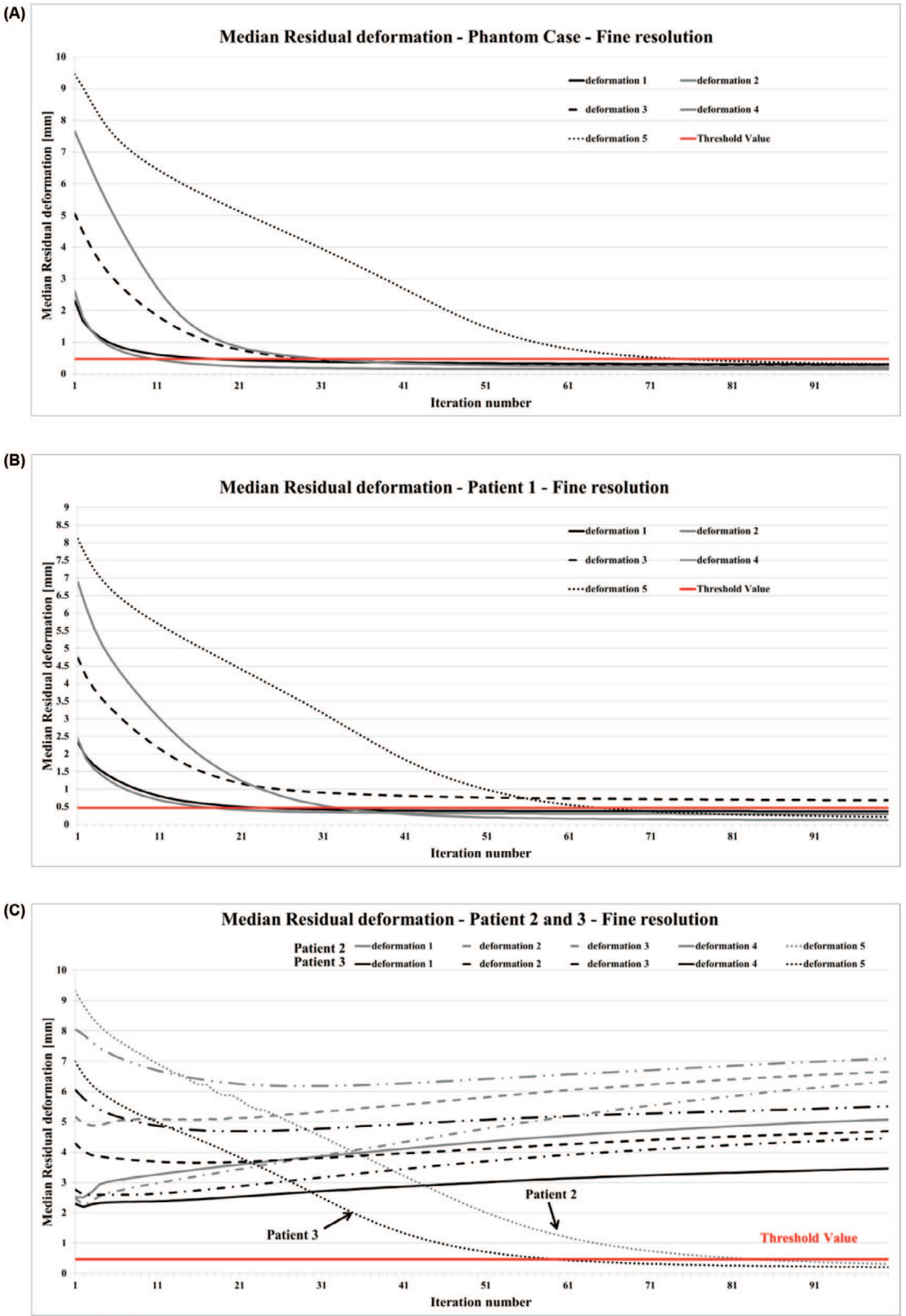

Figure 8: Summary of phantom and patient convergence at full resolution. In panel **A**, we report the median residual deformation in the five phantom cases (in red, the chosen accuracy value), whereas panel **B** and **C** refer to patient cases. In particular, panel **B** shows convergence curves and accuracy level for patient 1 at fine registration. Note how deformation 3 never reaches the desired threshold. In panel **C**, we show the same results for patient 2 (gray) and 3 (black). In this case, only deformation 5 was successfully recovered.

Table IV

Metric performances for phantom study in coarse and fine registrations. $t_{critical}$ at the iteration corresponding to the chosen threshold (1.875 mm and 0.469 for coarse and full resolution respectively).

	Coarse registration				Fine registration			
Test number	$v_{required}$ (1.875 mm)	MSE	HE	QU	$v_{required}$ (0.469 mm)	MSE	HE	QU
1	3	−0.373	0.027	−0.243	18	−0.026	−0.011	−0.022
2	10	−0.033	0	−0.022	11	−0.087	−0.089	−0.062
3	14	−0.019	0	−0.014	31	−0.026	−0.007	−0.017
4	295	0	−0.002	$-2.38e^{-5}$	32	−0.052	−0.035	−0.031
5	71	0	−0.003	−0.001	75	−0.016	−0.020	−0.011

Table V

Extra iterations for phantom study in coarse and fine registrations. N umber of extra iteration the algorithm needed to reach the minimum of the $t_{critical}$ (*i.e.* $t_{optimal}$).

	Coarse registration			Fine registration		
Test number	MSE	HE	QU	MSE	HE	QU
1	19	2	19	8	103	17
2	11	0	13	14	27	14
3	8	0	13	17	0	15
4	0	2	0	17	40	19
5	0	1	0	0	50	0
Total	38	**5**	45	56	220	65

Table VI

Extra iterations in the patient study for both coarse and fine registrations. Number of extra iteration the algorithm needed to reach the minimum of the $t_{critical}$ (*i.e.* $t_{optimal}$). Valid registrations are defined as those cases for which the desired $v_{required}$ was reached. Patient 2 and 3 were not evaluated at full resolution, because only one registration did actually converge to our accuracy level.

	Coarse registration				Fine registration			
Patient number	# of valid registrations	MSE	HE	QU	# of valid registrations	MSE	HE	QU
1	5	171	135	288	4	27	167	39
2	4	25	9	93	1	–	–	–
3	4	65	30	68	1	–	–	–
Total	13	261	179	449	6	27	167	39

Table VII

Metric Values and number of extra iterations required by patient 1 at full resolution. Note that case 3 is not reported, because the final median residual of the deformation did not reach the desired accuracy level of 0.469 mm (see also Figure 8).

	Fine registration – metric values at $v_{required}$ (0.469 mm) – Patient 1				Fine registration – number of extra iterations – Patient 1		
Test number	$v_{required}$	MSE	HE	QU	MSE	HE	QU
1	24	−0.015	−0.008	−0.008	0	0	12
2	18	−0.025	−0.048	−0.033	8	24	10
4	33	−0.024	−0.038	−0.034	14	33	17
5	65	−0.020	−0.018	−0.014	5	110	0
				Total	27	167	39

Discussion

In this work, we compared four different stopping conditions for a deformable registration algorithm. The aim was to give insights in how to evaluate and to set the threshold on a given exit while at the same time ensure convergence. The most popular strategy is to use a predefined maximum number of iterations as escape condition. This is not robust to differences in image quality, as well as to variations in pre-processing or initialization. Not only is it hard for non-expert users to choose is the right amount, running a pre-defined number of iterations can lead to sub-optimal results.

In addition, we believe that the definition of error-correlated thresholds is more intuitive than the setting of an abstract quantity such as the number of iterations.

We restricted our analysis to mono-modal registration and non-parametric registration. Warping original images with synthetic continuous deformation fields allows not only to compare our registration results with a ground-truth deformation, but also to analyze the non-rigid transformation, without any need for rigid pre-alignment. All results were evaluated in the patient area only, to avoid including edge effects in the analysis. Given that most of the deformation is recovered at coarse resolution, we tested 500 iterations at one eighth of the original image resolution. To analyze a more clinical scenario, we also separately run 300 iterations at full resolution. No multi-resolution registration was tested, as concatenating different stages each with un-optimized number of iterations can actually lead to over- or under-estimated performance of the chosen escaping criteria. We chose to run registration at the coarsest stage only to prevent biases in the evaluation of the different stopping conditions, as we are mostly interested in the evaluation of the ability of detecting when to stop the registration given a certain level of accuracy (in our case half the dimension of the resampled pixel). At the same time, the registration at full image resolution resembles more closely a real clinical scenario, in which the full anatomical details want to be recovered. In our work, quite a few registrations at full resolution did not reach the desired accuracy level and started diverging. The final warped image is also not accurately matching the reference image. The reason for these failures has to be found in the absence of an initialization/coarse registration before the fine anatomical details recovering, thanks to whom the shape of the cost function and hence the registration problem could be better posed. The image quality and the type of deformation to be recovered definitely have a larger influence in these registration cases, where all the pixel information is deployed. Furthermore, full-resolution registration without recovery of the largest bulk deformations might hamper the performances of smoothness-based indices (*e.g.* Harmonic Energy), as demonstrated by our results. Although the results obtained in this work should be further correlated with the parameter variations, this study represents a first systematic effort of evaluating possible stopping conditions based on the evolution of the registration itself.

Head and neck images were analyzed, simulating asymmetrical non-rigid deformations concentrated around neck area. Appropriate synthetic warping should be designed when studying the performances of the stopping criteria in other body district, but testing of different registration models or optimization schemes do not require significant changes in our experimental setup. Multi-modal registration could also be simulated, provided that registration metrics based stopping criteria are coherently updated. We designed our work as a validation study providing results in terms of discrepancy from a synthetic non-rigid field after registration between original image and its warped version. The same experiments shall be repeated for unknown ground-truth and different images (*e.g.* simulation and replanning CT) of the same patient and to other body district in order to extend the validity of our study. A preliminary idea on the impact of the different image quality can be derived by the comparison between phantom and patients results.

We tested six different stopping condition calculation strategies (Table I) designed based upon equation 1 and 2. We evaluated the resulting deformable registration algorithm performance in terms of speed, convergence and stability. We formulated each stopping condition so that the final user needs to define a threshold ε, which quantifies the residual error that is acceptable for the particular problem and escape criteria calculation strategy. In this work, we did not aim at assigning a value to ε, but to study and compare the most common stopping condition strategies in a very popular registration scheme.

In our dataset, both for coarse and fine resolution registrations, the best performance was obtained by comparing the minimum of the chosen metric evaluated at the last three iterations including the current one with the minimum value at the three iterations before the last. The computation of the minimum is replaced by the maximum for the Harmonic Energy case, which, by definition, increases during registration. For all tested stopping criteria, at both resolution levels, the worst performing strategies in convergence speed were those relying on the subtraction of cost function values at current and some previous iteration. This approach slows down significantly the iterative procedure especially in the finer stages of a multi-resolution registration strategy, leading to longer computational times. Conversely, reducing the number of iterations in which the convergence metrics is verified (see for example calculation strategy A) may cause the algorithm to stop in local minima or converge into a plateau, thus leading to registration errors.

When at convergence, the cost function evaluation might start jumping between two opposite small deformation (and small discontinuities) in the attempt of detecting the global minimum, which are reflected also in our analysis. When comparing results on phantom and patients data at coarser resolution, we noticed that the number of iterations for convergence slightly increased for patients, particularly when the image quality was severely compromised by artifacts. In addition, the synthetic deformation field was not recovered with the desired accuracy (residual $<$1.875mm or 0.469 respectively) in some cases. The larger number of non-valid registrations (*i.e.* registration that did not actually reach the desired v_{required}) was found in the full resolution registration cases. However, this does not mean that the chosen registration algorithm is necessarily unstable or inaccurate, but that

a multi-resolution registration would be desirable for these registrations cases. The fact that the same registration converges at coarser resolution reinforces our hypothesis that multi-resolution would be needed, to overcome the overflow of information that cause a sort of over fitting of the deformation in a suboptimal field.

The assessed stopping conditions (Mean Squared Error, Harmonic Energy, Quantity of Update) differ in the way in which they capture different aspects of a non-rigid registration task.

Mean Squared Error monitors the evolution of the difference between intensities, according to the assumption of intensity conservation. As we restrict our analysis to a mono-modal intensity-based, looking at Mean Squared Error as stopping condition equates to analyze the convergence curve of the first term directly. The results presented for Mean Squared Error can be easily generalized to other popular metrics such as mutual information or correlation coefficient. However, choosing Mean Squared Error as stopping condition implies not monitoring the deformation field throughout the computation and therefore not detect discontinuities. Nonetheless, it might take advantage of the sharp image differences due to artifacts, as testified from the few iterations needed from this metric in the full image resolution examples.

Quantity of Update is computed on the velocity field and is directly connected to $||u||^2$, which drives the deformation in a demons based strategy. If the relative update in the pulling force goes to zero, the algorithm shall move to a finer registration stage or stop. Quantity of Update was found to be sensitive to image artifacts, which might induce abrupt changes in the update field and compromise convergence. This underlines the crucial influence of image quality in deformable registration.

In both datasets at coarse registration, Harmonic Energy emerged as the winning metric, which outperformed both Quantity of Update and Mean Squared Error. Harmonic Energy is sensitive to the smoothness of the velocity field, *i.e.* its regularity, which is correlated both to the physical meaningfulness of the transform and to the degree of convergence. This quantity might deteriorate on a less smooth deformation field. Most of the deformable registration algorithms are regularized or inverse consistent, so that the continuity and smoothness of the deformation field is guaranteed. Depending upon where the regularization occurs (at each iteration or at convergence), this stopping criteria should be used carefully. In Log-Domain Diffeomorphic Demons algorithm, the update field and the current velocity field smoothness is controlled by the dimension of smoothing Gaussian kernel. In the present work, the amount of regularization was chosen based on our previous work (18). The regularity of the update as well as of the whole deformation shall be guaranteed to compute vector field based metrics. Harmonic energy definition was defined based on the velocity field, instead of the deformation itself, as the first is assumed stationary in our algorithm. Harmonic Energy was however outperformed by Mean Squared Error in full resolution cases, where the latter takes advantage of higher contrasts and has more information to monitor the actual convergence. Mean Squared Error in fact suffers from the blurring of the image that is typical of low resolution, resampled image (either because the analyzed dataset has lower resolution or because we are looking at a multi-resolution algorithm), whereas the accuracy of the information used in Harmonic Energy calculation is not influenced by resampling. This difference in the performances of the analyzed stopping conditions suggests that further work is needed, not only towards the identification of a stopping criteria that could be efficiently used on for all body districts, but also towards the development of new monitoring quantities looking at both image similarity and deformation field updates, especially in multi-resolution registration case.

Conclusions

In this work, we compared three different stopping criteria amongst the most popular ones in literature (Mean Squared Error, Harmonic Energy, and Quantity of Update) for a deformable image registration algorithm. Different strategies were tested on a dataset consisting of CT images of an anthropomorphic phantom and three patients, which were warped by five synthetic non rigid continuous deformations. Regardless of the specific stopping criteria applied and at both coarse and fine resolution, the best performing strategy was the one comparing the minimum (maximum for Quantity of Update) of the last three stopping condition values with the minimum (maximum for the Quantity of Update) of the three iterations before the current one. The criteria were than ranked in terms of speed, convergence and stability. Both for patient and phantom in a coarse registration, the most efficient condition is based on Harmonic Energy. Quantity of Update ranked second in the phantom study, but demonstrated to be sensitive to image quality. Mean Squared Error might be helped by strong anchor points in the images (such as metal artifacts) and is in fact the winning strategy for full resolution registrations.

The concept used in this paper can be easily extended to any transformation type, but according to the employed model, it might need extension and modification. The chosen dataset includes head and neck cases only, but could be easily be repeated for other anatomical districts, to verify the generalization capability of the analyzed stopping conditions. Furthermore, evaluation of escaping criteria in cases where the ground truth deformation is either unknown or derived from a non-image based deformable registration algorithm, such as for example point-based registration, should also be analyzed.

Acknowledgments

This work was supported in part by the Rocca Fundation and the National Alliance for Medical Image Computing (NAMIC), NIH Grant 2-U54EB005149-06. The authors would like to thank the Advanced Radiotherapy Center, European Institute of Oncology, Milan, Italy for providing dataset used in this project.

References

1. Bujold A, Craig T, Jaffray D & Dawson LA. Image-guided radiotherapy: has it influenced patient outcomes? *SeminRadiatOncol 22*(1), 50-61 (2012). DOI: 10.1016/j.semradonc.2011.09.001
2. Dawson LA & Menard C. Imaging in radiation Oncology: a perspective. *Oncologist 15*(4), 338-49 (2010). PMID:2041339
3. Schwartz D & Dong L. Adaptive radiation therapy for head and neck cancer – can an old goal evolve into a new standard? *Journal of Oncology 2011*, 1-16 (2011). DOI: 10.1155/2011/690595
4. Chen A, Niermann KJ, Deeley MA & Dawant BM. Evaluation of multiple-atlas-based strategies for segmentation of the thyroid gland in head and neck CT images for IMRT. *Phys Med Biol 57*(1), 93-111 (2012). DOI: 10.1088/0031-9155/57/1/93
5. Murphy K, van Ginneken B, Reinhardt JM, Kabus S, Ding K, Deng X, Cao K, Du K, Christensen GE, 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 MP, Schnabel JA, Jenkinson M, Lorenz C, Modat M, McClelland JR, Ourselin S, Muenzing SE, Viergever MA, De Nigris D, Collins DL, Arbel T, Peroni M, Li R, Sharp GC, Schmidt-Richberg A, Ehrhardt J, Werner R, Smeets D, Loeckx D, Song G, Tustison N, Avants B, Gee JC, Staring M, Klein S, Stoel BC, Urschler M, Werlberger M, Vandemeulebroucke J, Rit S, Sarrut D & Pluim JP. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE Trans Med Imaging 30*(11), 1901-1920 (2011). DOI: 10.1109/TMI.2011.2158349
6. Thirion JP. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis 2*(3), 243-260 (1998). PMID: 9873902
7. Holden M. A review of geometric transformations for nonrigid body registration. *IEEE Trans Med Imaging 27*, 111-28 (2008). DOI: 10.1109/TMI.2007.904691
8. Yang D, Li H, Low D, Deasy JO & El Naqa I. A fast inverse consistent deformable image registration method based on symmetric optical flow computation. *Phys Med Biol 53*(21), 6143-6165 (2008). DOI: 10.1088/0031-9155/53/21/017
9. Gu X, Choi D, Men C, Pan H, Majumdar A & Jiang S. Implementation and evaluation of various demons deformable image registration algorithms on a GPU. *Phys Med Biol 55*(1), 207-19 (2010). DOI: 10.1088/0031-9155/55/1/012
10. Sharp GC, Kandasamy N, Singh H & Folkert M. GPU-based streaming architectures for fast cone-beam CT image reconstruction and demons deformable registration. *Phys Med Biol 52*, 5771-83 (2007). PMID: 17881799
11. Schmidt-Richtberg A, Ehrhardt J, Werner R & Handels H. Diffeomorphic Diffusion Registration of Lung CT Images. In: van Ginneken B, Murphy K, Heimann T, Pekar V, Deng X (Eds.), Medical Image analysis for the clinic: A grand Challenge, Workshop Proceedings of the 13th International Conference MICCAI, pp. 55-62 (2010).
12. Nithiananthan S, Brock KK, Daly MJ, Chan H, Irish JC & Siewerdsen JH. Demons deformable registration for CBCT-guided procedures in the head and neck: convergence and accuracy. *Med Phys 36*(10), 4755-4764 (2009). PMID: 19928106
13. Vercauteren T, Pennec X, Perchant A & Ayache N. Diffeomorphic Demons: Efficient Non-parametric Image Registration. *NeuroImage45*(Suppl1),S61-72(2009).DOI:10.1016/j.neuroimage.2008.10.040
14. Peroni M, Golland P, Sharp GC & 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
15. Cachier P, Badinet E, Dormont D, Pennec X & Ayache N. Iconic feature based nonrigid registration: the PASHA algorithm. *Comput Vis Image Underst 89*(2-3), 272-298 (2003). DOI: 10.1016/S1077-3142(03)00002-X
16. Bossa M, Hernandez M & Olmos S. Contributions to 3D Diffeomorphic Atlas Estimation: Application to Brain Images, In: Proceedings of the 10th International Conference on Medical Imaging Computing and Computer Assisted Intervention (MICCAI 2007). 4791 LNCS (PART 1), pp. 667-674 (2007).
17. Vercauteren T, Pennec X, Perchant A & Ayache N. Symmetric Log-Domain Diffeomorphic Registration: A Demons-based Approach, In: Proceedings of the 11th International Conference on Medical Imaging Computing and Computer Assisted Intervention (MICCAI 2008), 5241 LNCS (PART 1), pp. 754-761 (2008).
18. Ciardo D, Peroni M, Riboldi M, Alterio D, Baroni G & Orecchia R. The role of regularization in Deformable Image Registration for Head and Neck Adaptive Radiotherapy. *Technol Cancer Res Treat*, In Press (2013). DOI: 10.7785/tcrt.2012.500327

Received: February 5, 2013; Revised: June 11, 2013; Accepted: June 26, 2013