Evaluation and Comparison of Deformable Image Registration Algorithms for 4D CT Images

Xiaokun Hu¹, Guangpu Shao¹, Jimin Yang¹ and Juan Yang¹*

¹School of Physics and Electronics, Shandong Normal University, Jinan, Shandong, China

*Corresponding author’s e-mail: juan.yang@sdnu.edu.cn

Abstract. Deformable image registration (DIR) is crucial in adaptive radiation therapy. However, the validation is a challenging work due to the lack of gold-standard. This study proposed an evaluation framework by using point-to-point displacement vector field (DVF). Three DIR algorithms including original flow, active Demons and symmetric force Demons were validated for ten lung 4D CT images with landmarks. DVFs derived from DIR algorithms (dDVF) and manually measured according to landmarks (mDVF) were analyzed and compared. Their target registration errors (TRE) and the relationship of lung motion and three-dimensional TRE were explored. The distance discordance metric (DDM) values were calculated. For all cases, the active Demons algorithm had the smallest TRE value and DDM values and it showed the poor correlation between 3D TRE and lung motion, which indicated that this algorithm outperformed other DIR algorithms enrolled in this study.

Preliminary results demonstrated that the proposed evaluation framework had a potential ability of providing clinical guidance for the selection of appropriate algorithm in radiation therapy.

1. Introduction

The accuracy of DIR algorithm appears to be significant to enable safe and accurate implementation of the ART technology. Currently, a growing number of publications have evaluated the DIR accuracy for various anatomical sites, imaging modalities and modified DIR algorithms using various metrics. For example, Schlachter et al. [1] proposed a visualization framework to explore and assess DIR accuracy. The framework was based on voxel-wise comparison of local image patches for which dissimilarity measures included intensity difference, sum of squared differences, normalized cross correlation, kullback-leibler divergence, histogram intersection, entropy difference and variation of information. All measures were computed and visualized to indicate locally the registration results. It is a more detailed visualization approach than other simple visualization methods. Varadhan et al. [2] compared the accuracy of DIR based on computational modeling. Schnabel et al. [3] simulated physically plausible and biomechanical tissue deformations using finite-element methods. The authors quantified accuracy with respect to the finite-element simulations by co-registering the deformed images with the original images and comparing the recovered voxel displacements with the biomechanically simulated ones. It is a common behavior to use computational modeling to evaluation DIR performance. But the real clinical performance of a specific algorithm cannot be expressed. The surrogate measures including image similarity [4], image difference, inverse consistency error [5] or transitivity error [6, 7] are generally used to compare the DIR accuracy approximately. Image similarity is not reliable for all images, since it depends directly on the pointwise intensity difference [1, 8]. In addition, Vickress et al. [9] evaluated the dose-predictive power of various measures of DIR
error. The result showed that the DDM had a comparable ability with the actual DIR error. However, few studies are available investigating the usage of DDM in evaluating and comparing DIR algorithms.

Inspired by the studies mentioned above, we presented a framework to evaluate three DIR algorithms qualitatively and quantitatively for aligning lung 4D CT, which was based on displacement vector fields (DVFs). The proposed framework in this study is flexible and could be used to other DIR algorithms comparison.

2. Methods and materials

2.1. Lung CT images and landmarks acquisition

Ten lung 4D CT images were obtained from ‘dir-lab’ [10, 11]. Each 4D CT image data includes 10 respiratory phases with 0% corresponding to end-of-inhalation (CT00) and 50% corresponding to end-of-exhalation (CT50). The images were acquired with the participants in the supine position with normal resting breathing. The 4D CT acquisition technique uses the respiratory signal from the Real-Time Position Management Respiratory Gating System (Varian Medical Systems, Palo Alto, CA). Three hundred landmarks were provided for each image study pair matching landmarks in the end-of-inspiration with corresponding points in the end-of-expiration. Landmarks were located across both the left and right lung with no significant bias, and were not limited to only the clinically important regions.

2.2. Deformable image registration algorithms (DIR)

For the moving image (M) and the fixed image (F), DIR is to compute displacement field \( u(v) \) the in order to optimize the system energy equation:

\[
E = \int_{\Omega} S(M(v), F) d\Omega + \alpha^2 \int_{\Omega} R(v) d\Omega
\]  

(1)

Where \( S \) indicates the similarity function and \( R \) represents the smoothness constraint function, \( \Omega \) refers to the image domain, and \( \alpha \) is a constant.

The Deformable image registration is the process of searching displacement field \( u(v) \), and the values of \( u(v) \) of different DIR are different. Three intensity-based DIR algorithms including original flow (Horn and Schunck 1981) (OF) [12], active Demons (AD) [13], symmetric force Demons (SFD) [14] were collected and validated in this study.

Our optical flow implementation is based on an iterative procedure used to solve for the unknown velocity at each voxel:

\[
v_{n+1} = v + \nabla I \left( \frac{\nabla I \cdot \bar{v}_n + \partial I \over \partial t}{\alpha^2 + \| \nabla I \|^2} \right)
\]

(2)

Where \( n \) and \( n + 1 \) are iteration counts and \( \bar{v}_n \) is the average velocity taken over the nearest neighboring voxels.

Wang et al. [13] applied an additional force to the image, proposing an active Demons algorithm. In the algorithm, the force obtained based on the moving image gradient information is taken as a positive internal force, and the force obtained from the fixed image gradient information is taken as a negative internal force. The implementation at each voxel is:

\[
u_{n+1}(v) = G_o \otimes \left( u_n(v) + \frac{(M(v) + S(v))(\nabla S(v) + \nabla M(v))}{\nabla S(v)^T + \alpha^2 (M(v) - S(v))^2} \right)
\]

(3)
Where $G_{\sigma}$ is the gaussian low-pass filtering function and $\sigma$ is the window width, $n$ is the number of iterations.

Rogelj proposed a symmetric Demons. In the algorithm, the force obtained based on the average of the gradients of the fixed image and the moving image, not just the gradient using the fixed image. The implementation at each voxel is:

$$u_{n+1}(v) = G_{\sigma} \otimes \left( \frac{u_n(v) + 2(S(v) - M(v))\nabla S(v)}{\left[ \nabla S(v) + \nabla M(v) \right]^2 + \left( M(v) + S(v) \right)^2} \right)$$

(4)

Where $G_{\sigma}$ is the gaussian low-pass filtering function and $\sigma$ is the window width, $n$ is the number of iterations.

Four multigrid levels were chosen and level 4 was the highest resolution. The iteration numbers were 30 (level 1), 30 (level 2), 20(level 3) and 20 (level 4), and the numbers of passes were 5, 4, 3 and 2, respectively. The stop condition for iteration and multiple passing were set to 0.002 and 0.001. The $\alpha^2$ parameter in the optical flow algorithms and the Gaussian low-pass filter window size in Demons were set to 0.2 and 3. After each pass, the deformation vector field computed by this pass could be smoothed by a Gaussian low-pass filter which sigma was set to 0.5. These values were in unit of voxel. The max filter was chosen to compute the image intensity during image down sampling for the multigrid approach.

2.3. Target registration error (TRE) comparison

Three DIR algorithms were successively applied to generate dDVF. The mDVF was calculated as following:

$$mDVF = P_{50} - P_{00}$$

(5)

Where $P_{50}$ and $P_{00}$ refers to the spatial positions of corresponding landmarks in CT$_{50}$ and CT$_{00}$. The TRE was defined as the deviation between dDVF and mDVF. It was calculated as following:

$$TRE = mDVF - dDVF$$

(6)

Each element in the two related DVFs in equation (6) is a 3D vector which contains displacement vectors in medial-lateral (ML), anterior-posterior (AP) and superior-inferior (SI) directions, respectively. Thus, each element of TRE is also a 3D vector. Three individual directions of TRE and 3D TRE of each landmark were calculated.

2.4. Study on relationship of target registration error and lung motion

The motion of tumor is closely related to the motion of lung tissue [15]. In this study, the correlation scatter plot was drawn and the Pearson Coefficient was calculated for 6 DIR algorithms. The motion of the lungs could be roughly represented by the motion of the landmarks. The degree of overall correlation can be observed, which can reveal the ability of different algorithms to overcome the effects of lung movement. A good DIR algorithm may potentially have a small correlation value between 3D TRE and lung motion, since the registration accuracy were not be easily affected by the lung motion.

2.5. Distance discordance metric (DDM) comparison

The DDM [9] used in this study was calculated as:

$$X_{p_i}^{00} = X_{p_i}^{00} + DVF_{00-p_i}(X_{00})$$

(7)

$$X_{50}^{p_i} = X_{p_i}^{00} + DVF_{p_i-50}(X_{p_i}^{00})$$

(8)
\[
DDM(X_{i0}) = \frac{1}{8} \sum_{i=1}^{8} |X_{i}^{p} - X_{i0}^{00}|
\]  

(9)

Where \(X\) represents a voxel located at coordinate \((x, y, z)\), and \(X_{i0}\) represents the corresponding voxel in CT\(_{i0}\). \(P_i\) refers to a specific phase in 8 additional phases (10\% - 40\% and 60\% - 90\%). \(X_{i}^{p}\) indicates the voxel in \(P_i\) relative to \(X_{i0}\). \(X_{i0}^{00}\) indicates the point in CT\(_{50}\) relative to \(X_{i}^{p}\). \(X_{i0}\) represents the corresponding point in CT\(_{50}\) deformed from CT\(_{i0}\) using the DVF\(_{00,50}\). The mean DDM and corresponding SD values were calculated over all evaluation points for different algorithms, respectively. And the DDM frequency distribution histogram was drawn for comparison.

3. Results

Figure 1 shows an example of 4D CT. The tumor is indicated by the red arrows in 10 phases. The tumor motion is visible throughout the respiratory cycle according to the dashed line. Figure 2 shows sampled DVFs between CT\(_{50}\) and CT\(_{00}\) confined in the lung area in coronal plane. As aforementioned, CT\(_{00}\) served as the moving image to obtain the calculated DVFs. The displacement fields calculated by 3 DIR algorithms showed great differences, which would be quantified through the evaluation framework proposed in this study.

![Figure 1. Representative 4D CT images in coronal and sagittal views of case 8. Dash lines were drawn to facilitate the visualization of respiratory motion.](image)

![Figure 2. The coronal DVFs confined in the lung area calculated by the OF (a), AD (b) and SFD (c) algorithms for case 8. The CT image at phase T = 0% (CT\(_{00}\)) was used as the moving image and T = 50% (CT\(_{50}\)) was used as the fixed image to derive the DVF.](image)

![Figure 3. Illustration of correlation between mDVF magnitude and 3D TRE for 3 DIR algorithms.](image)

Table 1 shows an overall statistic TRE. Generally, large 3D TRE values were observed in group 2 and small values were observed in group 1. For 10 cases, the mean (max) percentage of 3D TRE falling in the range from 0 mm to 2.5 mm were 81\% (97\%), 81\% (96\%) and 73\% (95\%) for OF, AD and SFD algorithms, respectively. The 3D TRE was more concentrated within a range of the slice thickness of CT images.
Figure 3 shows the correlation between mDVF and 3D TRE calculated by 3000 landmarks. A linear relationship between the two was observed. The mean correlation coefficient values for SFD, OF and AD algorithms were 0.69, 0.58 and 0.44, respectively. The AD algorithms were less affected by lung motion.

Table 2 illustrates the results of the mean DDM (SD) calculated by 3 algorithms. Large DDM values were observed in group 2. The maximum mean DDM was observed in the OF algorithm. The AD algorithm showed better registration in DDM comparison.

Table 1. Comparison of TRE (mm) between CT\textsubscript{00} and CT\textsubscript{50} using 3 DIR algorithms

| Algorithm | ML Mean (SD) | AP Mean (SD) | SI Mean (SD) | 3D Mean (SD) |
|-----------|--------------|--------------|--------------|--------------|
| Group 1   |              |              |              |              |
| OF        | 0.83 (0.72)  | 0.70 (1.00)  | 1.40 (1.58)  | 1.29 (1.50)  |
| AD        | 0.70 (0.74)  | 0.86 (1.02)  | 1.32 (1.38)  | 1.26 (1.34)  |
| SFD       | 0.92 (0.84)  | 0.82 (1.26)  | 1.72 (1.64)  | 1.61 (1.71)  |
| Mean      | 1.21 (1.20)  | 1.15 (1.79)  | 3.74 (5.17)  | 3.45 (4.87)  |
| Group 2   |              |              |              |              |
| OF        | 1.16 (1.00)  | 0.86 (1.86)  | 3.33 (4.03)  | 3.08 (4.03)  |
| AD        | 1.80 (1.86)  | 2.21 (4.48)  | 4.73 (4.73)  | 4.26 (4.89)  |
| SFD       | 0.92 (1.07)  | 1.33 (2.10)  | 3.30 (3.10)  | 1.94 (3.11)  |
| Mean      | 0.85 (0.92)  | 1.17 (1.41)  | 2.66 (1.81)  | 2.61 (2.61)  |

Table 2. Comparison of DDM (mm)

| Algorithm | ML Mean (SD) | AP Mean (SD) | SI Mean (SD) | 3D Mean (SD) |
|-----------|--------------|--------------|--------------|--------------|
| Group 1   |              |              |              |              |
| OF        | 0.85 (0.92)  | 1.07 (1.33)  | 2.10 (3.10)  | 1.94 (3.11)  |
| AD        | 1.16 (1.00)  | 1.87 (3.33)  | 4.03 (3.08)  | 4.03 (4.03)  |
| SFD       | 1.19 (1.11)  | 1.50 (1.69)  | 2.49 (3.24)  | 2.40 (3.27)  |
| Mean      | 0.85 (0.92)  | 1.07 (1.33)  | 2.10 (3.10)  | 1.94 (3.11)  |

ML = medial-lateral; AP = anterior–posterior; SI = superior–inferior; 3D = three-dimensional; SD = standard deviation; Group1 includes case 1-5, 9 and 10; Group 2 includes case 6, 7 and 8.

4. Discussion

In this study, a framework of DIR validation which was based on point-to-point DVF comparison was proposed. The validation was performed by using three metrics, which could be summarized into two orientations: TRE and DDM metrics for quantitative assessment of spatial error and DVF magnitude correlation metric for qualitative comparing the spatial trend of displacements. For all DIR algorithms participated in this study, the AD showed small TRE and DDM values, and its registration effect was less affected by the lung movement. Hence, the AD outperformed other DIR algorithms, which indicated that the proposed method had the ability of validating DIR performance and an appropriate algorithm could be determined for aligning lung 4D CT images.

There is no uniform standard for DIR assessment, since each algorithm has specific registration mechanism and even the tumor location is uncertain. As stated by Yeo et al. [16], multiple studies will be needed to provide sufficient evidence, and hence confidence, to validate the wide variety of applications of DIR. Thus, an evaluation framework was proposed to find algorithms with good performance by comparing multiple indicators in this study.

Inevitably, there are limitations in this study. The mean TRE and DDM values across whole lung were compared to evaluate algorithms. Although the overall performance of the algorithm for lung can be compared, the specific conditions of different pulmonary lobes are ignored. This may lead to the inability of selecting the most appropriate registration algorithm based on different tumor sites.
5. Conclusion
This study presented a framework of DIR validation, which was based on point-to-point DVF comparison by calculating variable metrics. The preliminary results in 10 lung 4D CT images demonstrated that the proposed DIR validation method had the potential ability in selecting appropriate registration algorithm for lung 4D CT images.

References
[1] Schlachter M, Fechter T, Jurisic M, et al. (1994) Visualization of Deformable Image Registration Quality Using Local Image Dissimilarity. IEEE Trans Med Imag., 35(10):2319-2328.
[2] Varadhan R, Karangelis G, Krishnan K, Hui SJJoaclmp. (2013) A framework for deformable image registration validation in radiotherapy clinical applications. J Appl Clin Med Phys., 14(1):192-213.
[3] Schnabel JA, Tanner C, Castellano-Smith AD, et al. (2003) Validation of nonrigid image registration using finite-element methods: application to breast MR images. IEEE Trans Med Imag., 22(2):238-247.
[4] Yang J, Li D, Yin Y, Zhao F, Wang HJTicr, treatment. (2016) Evaluation of Demons-and FEM-Based Registration Algorithms for Lung Cancer. Technol Cancer Res Treat., 15(2):275-284.
[5] Yang D, Li H, Low DA, Deasy JO, El Naqa I. (2008) A fast inverse consistent deformable image registration method based on symmetric optical flow computation. Phys Med Biol., 53(21):6143-6165.
[6] Bender ET, Tome WA. (2009) The utilization of consistency metrics for error analysis in deformable image registration. Phys Med Biol., 54(18):5561-5577.
[7] Bender ET, Hardcastle N, Tome WA. (2012) On the dosimetric effect and reduction of inverse consistency and transitivity errors in deformable image registration for dose accumulation. Med Phys., 39(1):272-280.
[8] Rohlfing T. (2012) Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. IEEE Trans Med Imag., 31(2):153-163.
[9] Vickress J, Battista J, Barnett R, Yartsev S. (2017) Representing the dosimetric impact of deformable image registration errors. Phys Med Biol., 62(17):N391-N403.
[10] Castillo R, Castillo E, Guerra R, et al. (2009) A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys Med Biol., 54(7):1849-1870.
[11] Castillo E, Castillo R, Martinez J, Shenoy M, Guerrero T. (2010) Four-dimensional deformable image registration using trajectory modeling. Phys Med Biol., 55(1):305-327.
[12] Horn BKP, Schunck BG. (1981) Determining optical flow. Artif Intell., 17(1):185-203.
[13] Wang H, Dong L, O'Daniel J, et al. (2005) Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. Phys Med Biol., 50(12):2887-2905.
[14] Rogelj P, Kovačič S. (2006) Symmetric image registration. Med Image Anal., 10(3):484-493.
[15] Smith RL, Yang D, Lee A, Mayse ML, Low DA, Parikh PJ. (2011) The correlation of tissue motion within the lung: implications on fiducial based treatments. Med Phys., 38(11):5992-5997.
[16] Yeo UJ, Supple JR, Taylor ML, Smith R, Kron T, Franich RD. (2013) Performance of 12 DIR algorithms in low-contrast regions for mass and density conserving deformation. Med Phys., 40(10):101701.