A multiple-image-based method to evaluate the performance of deformable image registration in the pelvis

Ziad Saleh\textsuperscript{1,5}, Maria Thor\textsuperscript{1}, Aditya P Apte\textsuperscript{1}, Gregory Sharp\textsuperscript{2}, Xiaoli Tang\textsuperscript{1}, Harini Veeraraghavan\textsuperscript{1}, Ludvig Muren\textsuperscript{3,4} and Joseph Deasy\textsuperscript{1}

\textsuperscript{1} Department of Medical Physics, Memorial Sloan Kettering Cancer Center, 1275 York Ave., New York, NY 10065, USA
\textsuperscript{2} Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA, USA
\textsuperscript{3} Department of Medical Physics, Aarhus University / Aarhus University Hospital, Aarhus C, Denmark
\textsuperscript{4} Department of Physics and Technology, University of Bergen, Bergen, Norway

E-mail: salehz@mskcc.org

Received 3 March 2016, revised 11 May 2016
Accepted for publication 7 June 2016
Published 29 July 2016

Abstract

Deformable image registration (DIR) is essential for adaptive radiotherapy (RT) for tumor sites subject to motion, changes in tumor volume, as well as changes in patient normal anatomy due to weight loss. Several methods have been published to evaluate DIR-related uncertainties but they are not widely adopted. The aim of this study was, therefore, to evaluate intra-patient DIR for two highly deformable organs—the bladder and the rectum—in prostate cancer RT using a quantitative metric based on multiple image registration, the distance discordance metric (DDM). Voxel-by-voxel DIR uncertainties of the bladder and rectum were evaluated using DDM on weekly CT scans of 38 subjects previously treated with RT for prostate cancer (six scans/subject). The DDM was obtained from group-wise B-spline registration of each patient’s collection of repeat CT scans. For each structure, registration uncertainties were derived from DDM-related metrics. In addition, five other quantitative measures, including inverse consistency error (ICE), transitivity error (TE), Dice similarity (DSC) and volume ratios between corresponding structures

Original content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

\textsuperscript{5} Author to whom any correspondence should be addressed.
from pre- and post-registered images were computed and compared with the DDM. The DDM varied across subjects and structures; DDM$_{\text{mean}}$ of the bladder ranged from 2 to 13 mm and from 1 to 11 mm for the rectum. There was a high correlation between DDM$_{\text{mean}}$ of the bladder and the rectum (Pearson’s correlation coefficient, $R_p = 0.62$). The correlation between DDM$_{\text{mean}}$ and the volume ratios post-DIR was stronger ($R_p = 0.51; 0.68$) than the correlation with the TE (bladder: $R_p = 0.46$; rectum: $R_p = 0.47$), or the ICE (bladder: $R_p = 0.34$; rectum: $R_p = 0.37$). There was a negative correlation between DSC and DDM$_{\text{mean}}$ of both the bladder ($R_p = -0.23$) and the rectum ($R_p = -0.63$).

The DDM uncertainty metric indicated considerable DIR variability across subjects and structures. Our results show a stronger correlation with volume ratios and with the DSC using DDM compared to using ICE and TE. The DDM has the potential to quantitatively identify regions of large DIR uncertainties and consequently identify anatomical/scan outliers. The DDM can, thus, be applied to improve the adaptive RT process for tumor sites subject to motion.

Keywords: deformable image registration, distance discordance metric (DDM), uncertainties, adaptive radiotherapy

(Some figures may appear in colour only in the online journal)

Introduction

Deformable image registration (DIR) is essential to ensure accurate delivery of radiotherapy (RT) for tumor sites subject to considerable motion, changes in tumor volume and normal anatomy due to patient’s weight loss (Jaffray et al 2010, Kadoya 2014). Given increased use of in-treatment-room volumetric imaging, DIR has the potential to be used routinely for detection of organ motion and anatomical changes over the course of RT (Lu 2006, Zhang 2007, Samant 2008, Wu et al 2009, Jaffray et al 2010, Kadoya 2014).

DIR is offered by most commercial RT treatment planning systems and is being used clinically for multi-modality image fusion and atlas-based segmentation (Sims 2009, Teguh 2011, Thor et al 2011, Hardcastle 2012, Daisne and Blumhofer 2013, Asman et al 2014). However, DIR-induced uncertainties are challenging to interpret and the interpretation is rather subjective to the viewer. Moreover, the lack of a ground truth together with registration errors, owing to organ motion and anatomical differences, e.g. variable bladder filling and variable amount of bowel gas (Brock 2010, Thor et al 2011, Varadhan et al 2013, Zambrano 2013, Kadoya 2014), limit the usefulness of DIR to ultimately adapt treatments (Brock 2010, Jaffray et al 2010, Zambrano 2013, Kadoya 2014, Rigaud et al 2015).

The most commonly used metrics to evaluate the performance of DIR involve the Dice similarity coefficient (DSC), the Hausdorff distance, and the mean surface distance, or by identification of landmarks (Castillo 2009, Kirby et al 2013, Latifi et al 2013, Varadhan et al 2013). However, the former three metrics typically rely on the availability of manually delineated structures (Brock 2010, Varadhan et al 2013), and the landmark technique is limited in regions of soft tissue where robust identification of landmarks is challenging (Dice 1945, Castillo 2009, Li 2013). Other DIR performance metrics include the inverse consistency error (ICE) and transitivity error (TE) (Christensen and Johnson 2001, 2003, Bender and Tomé 2009, Bender et al 2012), mean squared error (MSE), and the Jacobian (Latifi et al 2013, Varadhan et al 2013). The ICE and TE rely on registration between image pairs without
consideration to other images in the data set. Meanwhile, MSE relies on the underlying image intensities and does not contain any spatial information. Jacobian, on the other hand, can only provide information about tissue expansion and shrinkage without conveying any information about DIR uncertainties.

In our previous study we introduced the multiple-image-based distance discordance metric (DDM), and showed that the DDM was more strongly correlated with the absolute registration error than ICE and TE when DIR was performed on a digital phantom (Saleh 2014). The current study takes the DDM metric-based evaluation beyond phantom studies and we explore the performance of DDM in the context of intra-patient DIR of pelvic organs (the bladder and the rectum) in a series of subjects treated with RT for prostate cancer where repeat imaging CT data was acquired.

**Materials and method**

*Imaging data*

The imaging data consisted of CT scans from 38 subjects previously treated for prostate cancer at Haukeland University Hospital, Bergen, Norway (Thor 2013, Thor et al 2013). The data were collected within a clinical trial that was approved by the relevant ethics committee (REK Vest). Each subject had a planning CT scan (pCT), and also received weekly repeated CT (wCT) scans over the course of RT. All scans were acquired in supine position as close to the treatment session as possible and no filling/emptying protocol was applied to the bladder, nor to the rectum. The bladder and the rectum were manually contoured on all scans under supervision of the same radiation oncologist to limit inter-observer variability. For this study, the first six acquired weekly scans for each patient (wCTs) were used, resulting in a total of $38 \times 6$ scans; each with a scan resolution of $1 \times 1 \times 3$ mm$^3$. The pCTs were not used in this study due to the systematic use of bladder contrast.

*DIR and related uncertainties*

Group-wise DIR was performed using a B-spline algorithm with MSE cost function as implemented in the Plastimatch software (Sharp 2009, Shackleford 2012). Rigid registration was initially performed to align the images followed by deformable registration. The DIR imposes a regularization parameter with the purpose of generating accurate deformations. DIR was performed between all pairs of wCTs for each patient, and voxel-by-voxel uncertainties of the generated displacement vector fields (DVFs) were assessed by the DDM (Bender et al 2012). The DDM describes the mean of the distances among set of voxels as they get registered across different image sets. Suppose that a set of voxels from different image sets (wCT$_2$, wCT$_3$ ...) are co-registered to the same location on an image (wCT$_1$), these voxels will be distributed at nearby locations when the image sets are registered to an arbitrary image (wCT$_n$). If the registration is reasonable, then the distances between these voxels will be small. On the contrary, if the registration is bad, then the distances between the voxels will be larger. Therefore, small DDM corresponds to regions of good registration meanwhile large DDM values correspond to regions of bad registration. As the number of images increase, the performance of DDM will improve since it can capture more variations among images (Saleh 2014).

The resulting DDM map was overlaid on the first CT (wCT$_1$). Similarly to the DDM, the ICE and TE voxel-wise maps (Christensen and Johnson 2001, 2003) were calculated between all image pairs of each patient and the mean value at each voxel was overlaid on wCT$_1$ for comparison with the DDM.
For each structure (bladder/rectum) we defined the following two volume ratios:

For Pre-DIR: \[ V_{\text{pre}/V_{\text{ref}}} = \left| V_{\text{wCT}_i} - V_{\text{wCT}/V_{\text{wCT}_i}} \right|; \ i = [2 \ldots 6] \]

For Post-DIR: \[ V_{\text{post}/V_{\text{ref}}} = \left| V_{\text{dCT}_i} - V_{\text{wCT}_1}/V_{\text{wCT}_i} \right|; \ i = [2 \ldots 6] \]

Where \( V_{\text{wCT}_i} \) represents the volume of manually delineated contour on \( wCT_i \), and \( V_{\text{dCT}_i} \) the deformed contour from the \( wCT_i \) to \( wCT_1 \). The volume ratio will result in a small value if the volume of the deformed structure is comparable to the volume of the manual contour which indicates a good registration.

The Pearson’s correlation coefficient (\( R_p \)) was applied between \( (V_{\text{pre}/V_{\text{ref}}}) \), \( (V_{\text{post}/V_{\text{ref}}}) \), or DSC and the DDM, ICE, and TE. A weak, modest, and high correlation was indicated by \( R_p \leq 0.35 \), \( 0.36 \leq R_p \leq 0.67 \) and \( R_p \geq 0.68 \)–1.00, respectively (Deasy et al 2003). All metrics were compared using the Wilcoxon rank-sum test, and significance level was defined at a two-sided 5% level. All DIRs were conducted in Plastimatch under Linux, and data extraction and post-processing of the DVFś were performed in MATLAB (R2011a) and in CERR (Taylor 1990).

Results

Within the entire DDM map, regions with the highest DDM values were observed near the skin and in the bladder and the rectum (figure 1). The population median (range) DDM was 6.6 (1.5–14) mm and 5.0 (1.1–15) mm for the bladder and rectum, respectively. There was a moderate correlation between DDM\(_{\text{mean}}\) in the rectum and the bladder (\( R_p = 0.62 \)).

The population median (range) values for the bladder and the rectum using ICE were 7.4 (1.5–15) mm, and 5.4 (0.2–11) mm, respectively, whereas the corresponding values using TE were 3.5 (0.8–13) mm, and 6.4 (1.3–18) mm. There was, however, a wide distribution of the DDM, ICE, and the TE values across all subjects (figure 2).

A strong correlation was observed between these three metrics, and with the highest correlation being observed between TE and ICE (\( R_p = 0.95 \)), followed by DDM and TE (\( R_p = 0.93 \)), and DDM and ICE (\( R_p = 0.84 \); figure 3).

Subjects with a DDM\(_{\text{mean}}\) in the rectum above the population median (>5.0 mm) had significantly larger post-DIR volume ratios than subjects with a DDM\(_{\text{mean}}\) below the median (\( p = 0.001 \); table 1). A similar pattern was observed for both ICE (median > 3.5 mm) and TE (median > 6.4 mm). For the bladder, however, the differences in the volume ratios were statistically significant for DDM (median > 6.6 mm, \( p = 0.04 \)) and marginally significant for ICE (median > 5.4 mm, \( p = 0.1 \)) and TE (median > 7.4 mm, \( p = 0.1 \)).

The correlation between DDM\(_{\text{mean}}\) and \( (V_{\text{post}/V_{\text{ref}}}) \) was modest to high (rectum: \( R_p = 0.68 \); bladder: \( R_p = 0.53 \)), and slightly stronger compared to ICE and TE (table 2). The population median (range) of the DSC was 0.81 (0.51–0.92) for the bladder and 0.72 (0.62–0.84) for the rectum (figure 4). The DSC correlation with DDM, ICE and TE was correspondingly higher in the rectum (\( R_p = -0.63; -0.56; -0.53 \)) compared to the bladder (\( R_p = -0.23; -0.22; 0.18 \); table 2). The weakest overall correlations were observed with \( V_{\text{pre}/V_{\text{ref}}} \) (\( R_p < 0.10 \) for all metrics).

Discussion

Our multiple-image based DIR-uncertainty metric, the DDM, as applied to intra-patient DIR indicated considerable variability across the two investigated organs and across the
38 investigated subjects. Within the generated DDM map, the most pronounced variations were observed in regions of the bladder and the rectum, which are both subject to motion due to bladder filling or absence/presence of air/feces in the rectum. The DDM values were slightly higher in the bladder than in the rectum, and the highest values were observed in the superior part of the bladder and the regions invaded by bowel gas in the rectum. Meanwhile, regions of high contrast such as the bony anatomy showed the lowest variations. These results are consistent with the fact that regions of high contrast including the bony anatomy are less challenging to register, thereby, resulting in lower values of DDM uncertainty whereas, parts of anatomy prone to large errors in registration (Castillo 2009, Kirby et al 2013) are associated with higher values of DDM. This indicates that the DDM metric is viable for measuring relative DIR uncertainties.

Both the ICE and TE exhibited similarly large values in areas of poor registration. The extent of our registration uncertainties, as assessed by the DDM, ICE, or the TE, is in a similar range as the mean registration errors reported in previous studies (Brock 2010, Kirby et al 2013, Nie et al 2013, Varadhan et al 2013), although it should be pointed out that variations may be present given the choice of DIR algorithm and anatomy. We found a strong correlation between the DDM values of the bladder and the rectum, which might be an indication of the

Figure 1. Color wash representation of the uncertainty maps: ICE (top), TE (middle), and DDM (bottom) overlaid on an axial and sagittal view of a reference CT for an example patient. Deformed contours of the rectum and bladder from weekly CTs are also shown. The largest uncertainties are located near skin, in the rectum due to bowel gas and near top of the bladder due to bladder filling while bony anatomy has the lowest uncertainties.
interplay between motion caused by the bladder filling and bowel gas as illustrated by e.g. Nijkamp et al (2008).

In contrast to the DDM, ICE, and TE values, the DSC was slightly higher for the bladder (DSC mean = 0.81) than for the rectum (DSC mean = 0.72). These DSC values are comparable with those from other studies of the same organs (Thor et al 2011, Varadhan et al 2013, Zambrano 2013). The correlation with DSC using ICE, TE, or DDM was, however, modest for the rectum and weak for the bladder. It should be kept in mind that the DSC includes only volume information such that a higher DSC value does not necessarily indicate a more accurate registration (Kirby et al 2013). On the other hand, DDM, ICE, and TE correlated strongly with the ratio of the volumes of the deformed and the manually delineated structures (\(V_{\text{post}}/V_{\text{ref}}\)) with the strongest correlation for both structures observed with the DDM (rectum: \(R_p = 0.68\); bladder: \(R_p = 0.53\)). The lack of correlation between the DDM, ICE, or TE and the pre-registered volume ratios (\(V_{\text{pre}}/V_{\text{ref}}\)) or the DSC may indicate that this volumetric metric do not fully capture the full extent of the registration uncertainty.

Based on the results from this study, DDM resulted in higher correlations with the investigated volume ratios and the DSC compared to TE and ICE. Therefore, in the absence of a ground truth where absolute registration errors cannot be obtained, DDM can be used to quantify the underlying uncertainties for intra-patient DIR when multiple images (>3) are available. Given the absence of repeat intra-patient imaging, another application of the DDM could be to assess population based uncertainties from inter-patient DIR (Saleh 2014). As such, the generated ‘inter-patient DDM ATLAS’ could be deformed onto a new subject where patient specific uncertainties can’t be obtained due to lack of longitudinal images.

**Figure 2.** Stacked bar plots showing the distribution of the DDM (blue), TE (red), and ICE (green) for the rectum (top) and bladder (bottom) for the investigated 38 subjects. There are large variations among different subjects. ICE values are relatively smaller than DDM values while TE values are relatively larger.
Figure 3. Scatter plot showing the correlation between ICE, TE, and DDM for the rectum (top) and bladder (bottom). Best linear fit is shown in solid lines which are color coded. Highest correlation exists between TE versus ICE followed by TE versus DDM and DDM versus ICE respectively as indicated by $R^2$.

Table 1. Summary of $p$-values for the Wilcoxon rank-sum test statistics.

|            | Rectum       | Bladder      |
|------------|--------------|--------------|
|            | $V_{\text{pre}}/V_{\text{ref}}$ | $V_{\text{post}}/V_{\text{ref}}$ | $V_{\text{pre}}/V_{\text{ref}}$ | $V_{\text{post}}/V_{\text{ref}}$ |
| DDM        | 0.90         | 0.68         | 0.86         | 0.74         |
| TE         | 1.00         | <0.001      | 0.86         | 0.10         |
| ICE        | 0.63         | <0.001      | 0.74         | 0.10         |

Statistical significance level $p$-value < 0.05.

Table 2. Pearson correlations ($R_p$) between DDM, TE, ICE and the DIR volume metrics.

|            | Rectum       | Bladder      |
|------------|--------------|--------------|
|            | $V_{\text{pre}}/V_{\text{ref}}$ | $V_{\text{post}}/V_{\text{ref}}$ | DSC       | $V_{\text{pre}}/V_{\text{ref}}$ | $V_{\text{post}}/V_{\text{ref}}$ | DSC       |
| DDM        | 0.03         | 0.68         | −0.63     | −0.08     | 0.53         | −0.23     |
| TE         | −0.05        | 0.49         | −0.56     | −0.08     | 0.48         | −0.22     |
| ICE        | −0.06        | 0.37         | −0.53     | −0.03     | 0.36         | −0.18     |
Conclusion

Applied to intra-patient DIR for the bladder and the rectum, our automated DIR performance metric, the DDM, was more strongly correlated with post-DIR volume ratios than the commonly used DSC, the ICE or the TE. The DDM could, thus, be used to quantitatively evaluate DIR-related uncertainties and further identify regions of poor DIR both being essential for adaptive RT purposes.

Conflict of interest

None

Acknowledgment

The CT data were collected at Haukeland University Hospital, Bergen, Norway and provided by the responsible oncologist Svein Inge Helle and physicist Liv Bolstad Hysing. Lise Bentzen, Aarhus University Hospital, is acknowledged for approving the manually contoured bladder and rectal structures. This research was partially supported by the MSK Cancer Center Support Grant/Core Grant (P30 CA008748). This project was also supported in part through the NIH Grant R01 CA85181.

References

Asman A J, Bryan F W, Smith S A, Reich D S and Landman B A 2014 Groupwise multi-atlas segmentation of the spinal cord’s internal structure Med. Image Anal. 18 460–71
Bender E T, Hardcastle N and Tome W A 2012 On the dosimetric effect and reduction of inverse consistency and transitivity errors in deformable image registration for dose accumulation Med. Phys. 39 272
Bender E T and Tome W A 2009 The utilization of consistency metrics for error analysis in deformable image registration Phys. Med. Biol. 54 5561–77
Brock K K 2010 Results of a multi-institution deformable registration accuracy study (MIDRAS) Int. J. Radiat. Oncol. Biol. Phys. 76 583–96
Castillo R et al 2009 A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets Phys. Med. Biol. 54 1849–70

Christensen G E and Johnson H J 2001 Consistent image registration IEEE Trans. Med. Imaging 20 568–82

Christensen G E and Johnson H J 2003 Invertibility and transitivity analysis for nonrigid image registration J. Electron. Imaging. 12 106–17

Daisne J F and Blumhofer A 2013 Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation Radiat. Oncol. 8 154

Deasy J O, Blanco A I and Clark V H 2003 CERR: a computational environment for radiotherapy research Med. Phys. 30 979–85

Dice L R 1945 Measures of the amount of ecologic association between species Ecology 26 297–302

Hardcastle N et al 2012 A multi-institution evaluation of deformable image registration algorithms for automatic organ delineation in adaptive head and neck radiotherapy Radiat. Oncol. 7 90

Jaffray D A, Lindsay P E, Brock K K, Deasy J O and Tomé W A 2010 Accurate accumulation of dose for improved understanding of radiation effects in normal tissue Int. J. Radiat. Oncol. Biol. Phys. 76 S135–9

Kadoya N 2014 Use of deformable image registration for radiotherapy applications J. Radiol. Radiat. Ther. 2 1042

Kirby N, Chuang C, Ueda U and Pouliot J 2013 The need for application-based adaptation of deformable image registration Med. Phys. 40 011702

Latif K, Zhang G, Stawicki M, van Elmpt W, Dekker A and Forster K 2013 Validation of three deformable image registration algorithms for the thorax J. Appl. Clin. Med. Phys. 14 3834

Li S et al 2013 Voxel-based statistical analysis of uncertainties associated with deformable image registration Phys. Med. Biol. 58 6481–94

Lu W et al 2008 Deformable registration of the planning image (kVCT) and the daily images (MVCT) for adaptive radiation therapy Phys. Med. Biol. 51 4357–74

Nie K, Chuang C, Kirby N, Braunstein S and Pouliot J 2013 Site-specific deformable imaging registration algorithm selection using patient-based simulated deformations Med. Phys. 40 041911

Nijkamp J et al 2008 Adaptive radiotherapy for prostate cancer using kilovoltage cone-beam computed tomography: first clinical results Int. J. Radiat. Oncol. Biol. Phys. 70 75–82

Rigaud B et al 2015 Evaluation of deformable image registration methods for dose monitoring in head and neck radiotherapy Biomed. Res. Int. 2015 726268

Saleh Z et al 2014 The distance discordance metric—a novel approach to quantifying spatial uncertainties in intra- and inter-patient deformable image registration Phys. Med. Biol. 59 733–46

Samant S S et al 2008 High performance computing for deformable image registration: towards a new paradigm in adaptive radiotherapy Med. Phys. 35 5546–53

Shackelford J A et al 2012 Plastimatch 1.6—design, architecture, and future directions MICCAI Workshop

Sharp G et al 2009 Plastimatch—an open source software suite for radiotherapy image processing Computer Engineering

Sims R et al 2009 A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck Radiother. Oncol. 93 474–8

Taylor R 1990 Interpretation of the correlation coefficient: a basic review J. Diagn. Med. Sonog. 1 35–9

Teguh D N et al 2011 Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck Int. J. Radiat. Oncol. Biol. Phys. 81 950–7

Thor M et al 2013 Prediction of rectum and bladder morbidity following radiotherapy of prostate cancer based on motion-inclusive dose distributions Radiother. Oncol. 107 147–52

Thor M, Apte A, Deasy J O and Muren L P 2013 Statistical simulations to estimate motion-inclusive dose-volume histograms for prediction of rectal morbidity following radiotherapy Acta Oncol. 52 666–75

Thor M, Petersen J B, Bentzen L, Høyer M and Muren L P 2011 Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of prostate cancer Acta Oncol. 50 918–25

Varadhan R, Karangelis G, Krishnan K and Hui S 2013 A framework for deformable image registration validation in radiotherapy clinical applications J. Appl. Clin. Med. Phys. 14 4066

Wu Q, Chi Y, Chen P Y, Krauss D J, Yan D and Martinez A 2009 Adaptive replanning strategies accounting for shrinkage in head and neck IMRT Int. J. Radiat. Oncol. Biol. Phys. 75 924–32

Zambrano V et al 2013 Performance validation of deformable image registration in the pelvic region J. Radiat. Res. 54 1120–8

Zhang T et al 2007 Automatic delineation of on-line head-and-neck computed tomography images: toward on-line adaptive radiotherapy Int. J. Radiat. Oncol. Biol. Phys. 68 522–30