Utility of deformable image registration for adaptive prostate cancer treatment. Analysis and comparison of two commercially available algorithms

Michał Posiewnik[1], **Tomasz Piotrowski**[2,3,*]

[1] Department of Medical Physics, Gdynia Oncology Centre, Powstania Styczniowego 1, 81-519 Gdynia, Poland
[2] Department of Electroradiology, Poznan University of Medical Sciences, Garbary 15, 61-866 Poznan, Poland
[3] Department of Medical Physics, Greater Poland Cancer Centre, Garbary 15, 61-866 Poznan, Poland

Received 13 July 2021; accepted 17 October 2021

Abstract

Introduction: *This study aimed to assess and compare the capabilities of two commercially available deformable image registration algorithms implemented in Raystation 9A (A_1) and Velocity AI (A_2) for possible usage in adaptive prostate radiotherapy based on the propagation of anatomical contours from computed tomography (CT) images to cone-beam CT (CBCT) images.*

Material and methods: *Ten patients were retrospectively selected from a group treated for localized prostate cancer. The propagation of rectum contours was analyzed in a set of CT-CBCT pairs. Two independent observers carried out qualitative analysis using the two-level descriptive scale (meet/fail). Quantitative analysis was done using landmark points distances based on implanted markers as navigation points and differently obtained contours (manually and automatically using DIR algorithms). Quantitative analysis was taken on sets preselected by qualitative analysis.*

Results: *Qualitative analysis shows that 83.7% of the rectum contours were scored identically (meet or fail) for both algorithms, from which 53.5% and 55.8% are failed results for A_1 and A_2, respectively. For the rectum size (RWD parameter), differences between referenced and deformation-based values were 5.5 and 5.8 mm, and for the rectum wall, the prostate marker distance (WMD parameter) was 4.5 and 5.5 mm for A_1 and A_2, respectively. The differences between the WMD parameters were statistically significant ($p = 0.007$).*

Conclusions: *In both tested algorithms, neither effectiveness nor measured uncertainties in the propagation of rectum contour process in prostate patient cases were satisfactory. Careful selection of input images followed by case/set-based verification of every deformable registration is a substantial step to avoid inappropriate results.*

Keywords: Deformable registration, Prostate adaptive radiotherapy, DIR assessment

* Corresponding author: Tomasz Piotrowski, Department of Electroradiology, Poznan University of Medical Sciences, Garbary 15, 61-866 Poznan, Poland.
E-mail: tomasz.piotrowski@me.com (T. Piotrowski).

Z Med Phys **32** (2022) 369–377
https://doi.org/10.1016/j.zemedi.2021.10.001
www.elsevier.com/locate/zemedi

1 Introduction

Future progress in adaptive radiotherapy (ART) depends on assessment and development of precise fraction dose determination and accumulation [1]. Improvement in this respect is conventionally associated with progress in curved image registration according to non-rigid changes that occur inside human body [2–4]. In general, image registration is a process of finding transformation that aligns two images representing a similar morphological area. There are three groups of transformations: (a) rigid that preserves the Euclidean distance between points and includes translations and rotations, (b) affine that preserves parallelism of lines and includes global scaling, and (c) deformable image registration (DIR) [5]. The DIR cannot be represented as a sum or combination of rigid and affine transformations. The result of rigid transformations is used as pre-registrations that define the starting orientation of images before the DIR and the affine transformation is commonly used for scaling images during the pre-registration procedure [3]. The DIR process is based on a few steps that are focused on similarity metrics, optimization, regularization, and transformation. The manner of implementing individual steps depends on a mathematical model used during the DIR [6].

Nowadays, there are easily accessible DIR algorithms both free and in combination with commercial software that represent diverse approaches to the issue [5]. While recommendations are given in report AAPM TG 132 [7] about using DIR in the planning and image-guided procedures of radiotherapy, dose accumulation, and propagation is constantly a field of fierce debate [8]. On one hand, there are growing expectations about dose calculation accuracy and quality of the image-based contouring. On the other hand, there are needs of daily imaging for daily dose adaptation and accumulation. DIR algorithms complexity, significance of results and large scope of possible clinical challenges drive us to introduce quality assurance not only at commissioning but also on the clinical or patient level. This opens a wide field for tailored quality assurance (QA) tools that are sufficiently fast and reliable on a daily clinical basis [7].

Prostate ART based on daily cone-beam computed tomography (CBCT) images is one of the most explored and developed technique but still challenging especially with DIR [4,9]. Unique combination of low contrast structure boundaries (between the prostate and rectum or bladder), daily differences in structure shapes and positions [10,11] and CBCT low quality [12] makes it an interesting example.

In the present study we introduce a direct qualitative and quantitative performance analysis and comparison of two commercially available DIR algorithms. For the qualitative analysis, we used a simple two-level (meet/fail) scale and for the quantitative analysis, we uses selected landmark points and a rectum contour. As complementary volumetric parameter, we used the Dice similarity coefficient.

2 Material and methods

The study compares the utility of two commercially available DIR algorithms for matching CT and CBCT images during the treatment of low- or intermediate-risk prostate cancer. The analysis was performed to check the registration accuracy of anatomical regions including the prostate and rectum.

2.1 Algorithms

The following algorithms were used in this study:

A_1: Anatomically constrained deformation algorithm (ANACONDA) implemented in Raystation 9 A (RaySearch Laboratories AB, Stockholm, Sweden) [13].
A_2: Diffeomorphic registration algorithm implemented in Velocity AI ver. 4 (Varian Medical Systems Inc., Palo Alto, CA, USA) [14].

Both investigated algorithms are classified as free form deformation algorithms and utilize B-splines for image interpolation. In A_1 two options of similarity measure are available: mutual information (MI) and Pearson's correlation coefficient (PCC). In this study we used PCC as it is recommended for CT-CBCT registration [15]. PCC is a straightforward measure that directly compares both image intensities. It checks if there is a linear dependence between them. In A_2 similarity measure was automatically selected for CT-CBCT registration as MI [16].

2.2 Patients

Ten patients were retrospectively selected from a group treated for localized prostate cancer with total dose of 76 Gy in 38 fractions. All patients had three gold anchor markers (Naslund Medical AB, Huddinge, Sweden) implanted in the prostate gland before radiotherapy. The planning CT images were taken on Somatom Definition AS (Siemens Medical Solutions, Erlangen, Germany) using standard pelvic protocol for radiotherapy (120 kV and 162 mAs). Every patient was scanned and treated in a supine position with a knee-fix immobilization system (Combi-fix; CIVCO Radiotherapy, Coralville, IA, USA). The CBCT images were taken on the TrueBeam accelerator with an on-board imaging system v. 2.5 (Varian Medical Systems, Palo Alto, CA) using a standard protocol for pelvis imaging (full fan, 125 kV and 100 mAs). Patients are instructed to drink 500 mL of water half hour before CT imaging and, subsequently, before every treatment fraction. Similarly, to prepare the rectum, patients were instructed to follow an easily digestible diet.

2.3 Rectum delineation

Delineation of the rectum on CT images following the ESTRO ACROP guidelines [17] using magnetic resonance images as guidance. The rectum was contoured from the recto-sigmoid junction up to approximately 2 cm below the lowest prostate apex contour. The contours were done by a single physician and were supervised and accepted by the department case conference. To obtain contours on CBCT sets, 3 different methods were used:

M_0: manual delineation. The delineation process was divided into two parts i.e., manual delineation was done by a single observer on all CBCT sets and then inspected by another observer.
M_1: contours obtained from the first (**A_1**) deformable registration algorithm (ANACONDA).
M_2: contours obtained from the second (**A_2**) deformable registration algorithm (Velocity).

2.4 Registration scheme

Both algorithms require initial rigid registration as a starting point for the deformation process. In A_1 we automatically registered all pairs if images using point-based registration with 3 implanted markers as points and one virtual point created in the mass center of the marker system. In A_2, because of the utilization of the adaptive monitoring navigator, the control on rigid registration was limited to translation/rotation tools that allow operations based on information from the whole or specified parts of registered images, not from points marked on it.

In A_1 the user can manually choose grid resolution before starting the algorithm (from 5 mm in all directions to 1.3 mm in the in-plane directions and 3 mm in the longitudinal direction). We choose, always, the finer possible resolution. In A_2 we choose a multi-pass option which means that algorithms work in an iterative process through three resolutions (coarse, medium, and fine resolution) where the result from one resolution is the starting point for the next step. This process, called a multi-resolution optimization strategy [18], allow obtaining the final deformable grid resolution equal to 5 mm.

In general, when using A_1, the deformation takes place in the whole image's overlapping region. Nevertheless, the hybrid mode is possible, where additional constraints, such as the region of interest (ROI) or points of interest (POI), are used for output refinement. Therefore, to fine-tune the registration process for A_1, we used ROI defined as the cuboid volume created based on previously outlined structures, i.e. the prostate gland and the rectum, expanded by a 1 cm margin in the longitudinal direction. A_2 requires that the ROI is defined before every deformation (when the navigator is starting). In contrast with A_1, this is the only region where the algorithm works. Both ROIs (for A_1 and A_2) included the same volume. To prevent unwanted marker displacement after rigid registration due to deformation in A_1, we selected three fiducial markers as controlling POIs (points of interest) which became bonds for the whole process.

2.5 Analysis

The contours (M_1 and M_2) that were morphed and moved from CT to CBCT images as the result of applying automatic methods (A_1 and A_2) were evaluated by visual inspection by two independent observers (Obs_1 and Obs_2) using a simple two-level descriptive scale (meet/fail). Unacceptable result (fail) was registered when the contour had to be corrected on at least three scans, or the outline's structural error was observed. Set is considered to meet the criteria only if there is consistency of results for both observers, otherwise the set is scored as fail. The agreement between the observer ratings for each method (i.e., Obs_1 *vs* Obs_2 for A_1 and Obs_1 *vs* Obs_2 for A_2) as well as between the methods (i.e., A_1 *vs* A_2) was analyzed using Fisher's exact test. The Cohen's Kappa coefficient was used to measure inter-observer reliability and the correlation between scores obtained for the algorithms. Tests were performed at a significance level of 0.05.

Quantitative analysis was performed for the rectum contours from CBCT sets that were scored positively for both methods (A_1 and A_2) and by both observers (Obs_1 and Obs_2) during qualitative evaluation. The contours manually delineated (M_0) were selected as the reference. The three landmark points related to the rectum (RP) and three associated with implanted gold markers (MP) were selected for every qualified CBCT set. Each single point in the rectum (RP) was connected by a straight line to one point in the prostate (MP). The shortest distance between RP and MP was the determining factor of points connection. In the places where each of the lines crossed the contour of the rectum, another three points were inserted to represent the reference position of the rectal wall (WP). Figure 1 shows the schematic view of location of the points and lines.

The points locations determined two kinds of distances that were measured: RWD – the distance between a point located in the rectum (RP) and point on the rectum wall (WP), and WMD – the distance between point located on the rectum wall (WP) and point in the prostate (MP). The distances RWD and WMD were measured for referenced contours (M_0) as well as for the contours that were generated by the A_1 and A_2 algorithms (i.e., M_1, M_2). The differences between these distances obtained for M_0 and M_1 or M_2 were calculated according to the following formulas:

$$\Delta RMD_i^{Ma} = RWD_i^{Mo} - RWD_i^{Ma} \quad \text{and}$$
$$\Delta WMD_i^{Ma} = WMD_i^{Mo} - WMD_i^{Ma} \tag{1}$$

where $i = 1, 2, 3$ is a point number, *Mo* is the manual contouring method and *Ma* is the automatic method (*M*) number ($a = 1, 2$).

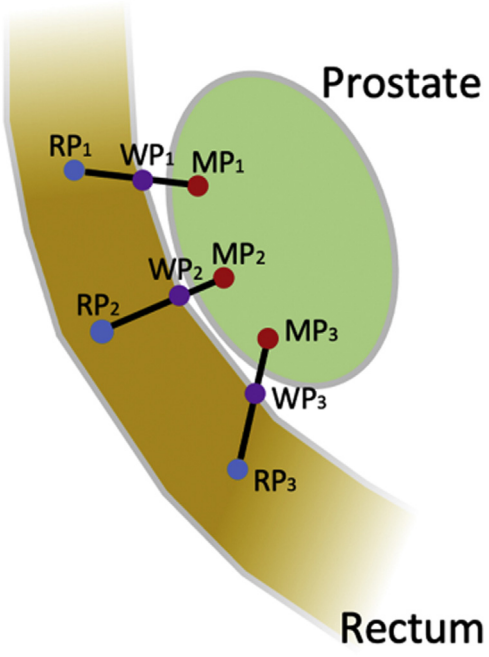

Figure 1. Schematic view of the points used to determine the position of wall of the rectum adjacent to the prostate. RP – points located in the rectum, MP – points in the prostate associated to markers, WP – points located on the intersection of the rectal wall defined by contour and of the lines that connect the RP and MP points.

While the potential differences of RWDs suggest changes in the rectum volume, the differences of WMDs suggest changes in the rectum position related to the prostate. Finally, the sum of differences was counted for every CBCT set:

$$RWD(M_a) = \sum_{i=1}^{3} \Delta RWD_i^{Ma} \quad \text{and}$$

$$WMD(M_a) = \sum_{i=1}^{3} \Delta WMD_i^{Ma} \tag{2}$$

The mean Dice similarity coefficient (DSC) [19] between the rectum obtained using automatic (M_1, M_2) and manual delineation (M_0) was chosen as an alternative and complemented our landmark-based method. The DSC was calculated according to the formula:

$$DSC = 2 \times \frac{V(R_{Mo}) \cap V(R_{Ma})}{V(R_{Mo}) + V(R_{Ma})} \tag{3}$$

where $V(R_{Mo})$ is the volume (V) of the rectum (R) defined on the basis of manual delineation (Mo), and $V(R_{Ma})$ is the volume (V) of the rectum (R) delineated based on the automatic method (M) number ($a = 1, 2$). DSC assumed values were from 0 (no overlap) to 1(full overlap).

In all analysis, the Wilcoxon test at significance level of 0.05 was used to check the statistical significance of differences between the automatic methods (M_1 and M_2).

3 Results

We evaluated 10 CT sets (one for every patient) and 111 CBCT sets which were taken during the treatment course (from 5 to 12 per patient). As a preliminary preparation, all sets were reviewed for quality reasons (artifacts). 14 CBCT sets were excluded from further analysis due to metal artifacts produced by markers implanted in the prostate gland (2 sets) and air moving in the rectum (12 sets).

The comparison of the observer ratings for A_1 and A_2 was taken on 97 CBCT sets. In the case of A_1, the rectum contour was accepted by both observers for 45 CBCT sets (46.4% of all sets analyzed) and rejected for 46 (47.4%) CBCT sets. For 6 (6.2%) CBCT sets, the observer ratings were different. Analyzing the observers' agreement for A_2, the rectum contour was accepted for 42 (43.3%) and rejected for 50 (51.5%) CBCT sets. For 5 (5.2%) CBCT sets, the observer ratings were different. The Fisher exact test confirm high agreement between observers' ratings ($p < 0.001$ for both analysis). According to the scale proposed by Landis and Koch [20], the Cohen's Kappa coefficients were noted as "*almost perfect*" and were 0.876 for inter-observer analysis for A_1 and 0.896 for A_2. Table 1 shows the agreement between observers' scores for accuracy of the rectum contours morphing using the A_1 and A_2 algorithms.

Direct comparison between the methods (i.e., A_1 *vs* A_2) was taken on the CBCT sets where both observers had compliant results (86 out of 97) and for the whole group of 97 observations (Table 2). Performance obtained through the algorithms confirmed the similarity of their efficacy ($p < 0.001$, Fisher exact test) in both situations. For the whole group of the CBCT sets (Table 2b), the Cohen's Kappa coefficient was noted as "*moderate*" and was 0.521. For selected observations (i.e., 86 CBCT sets equally scored by observers, Table 2a), the Cohen's Kappa coefficient was higher than for the whole group (i.e., 0.672) and was defined as "*substantial*". Analysing this selected group of observations (Table 2a), we noted that in 72 out of 86 sets (83.7%) both rectum contours were scored identically: positively for 32 (37.2%) CBCT sets, and negatively for 40 (46.5%). For 14 (16.3%) CBCT sets differences in scored outcomes were obtained. Considering direct outcome comparison of algorithms performance, the A_1 received positive results in 40 (46.5%) cases and A_2 in 38 (44.2%) cases.

Quantitative analysis was taken on 32 CBCT sets scored positively by both Obs_1 and Obs_2 for both algorithms (A_1, A_2). The mean summed differences RWD(M_1) and RWD(M_2) were, respectively, 5.8 and 5.5 mm and were statistically comparable ($p = 0.640$). Mean summed difference WMD(M_1) was 4.5 mm and was significantly smaller ($p = 0.007$) than mean summed difference WMD(M_2) that was 5.5 mm. The

Table 1
Agreement between observers' scores (Obs1, Obs2) for accuracy of the rectum contours morphing using A_1 and A_2 algorithms.

Algorithm A_1 *Fisher exact test: $p < 0.001$* *Cohen's Kappa: $R = 0.876$*		Observer 2	
		Meet	Fail
Observer 1	Meet	45 (46.4%)	2 (2.1%)
	Fail	4 (4.1%)	46 (47.4%)

Algorithm A_2 *Fisher exact test: $p < 0.001$* *Cohen's Kappa: $R = 0.896$*		Observer 2	
		Meet	Fail
Observer 1	Meet	42 (43.3%)	3 (3.1%)
	Fail	2 (2.1%)	50 (51.5%)

Table 2
Direct comparison between A_1 and A_2 algorithms: (a) for 86 CBCT sets equally assessed by observers, and (b) for the whole group of 97 CBCT sets.

(a)	**A_1 vs A_2** for 86 CBCT sets equally assessed by observers			
	Fisher exact test: $p < 0.001$ *Cohen's Kappa: $R = 0.672$*		Algorithm A_2	
			Meet	Fail
	Algorithm A_1	Meet	32 (37.2%)	8 (9.3%)
		Fail	6 (7.0%)	40 (46.5%)

(b)	**A_1 vs A_2** for the whole group of 97 CBCT sets			
	Fisher exact test: $p < 0.001$ *Cohen's Kappa: $R = 0.521$*		Algorithm A_2	
			Meet	Fail
	Algorithm A_1	Meet	32 (33.0%)	13 (13.4%)
		Fail	10 (10.3%)	42 (43.3%)

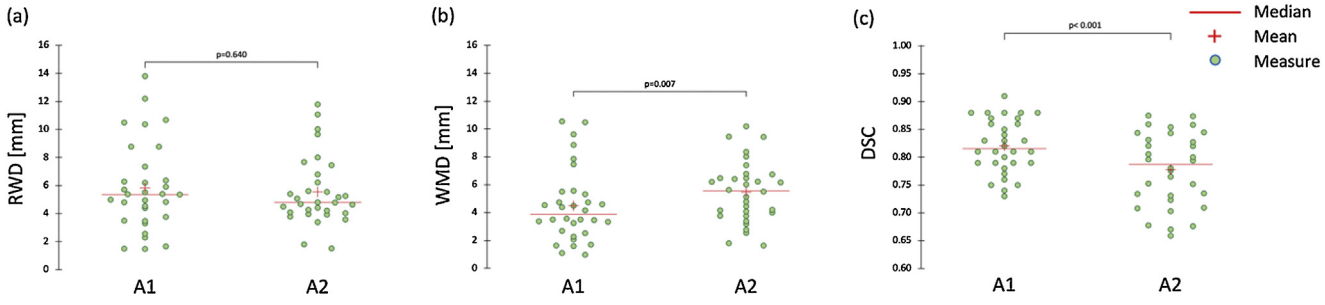

Figure 2. Summed differences of (a) RWD – the rectum center to the rectal wall distance, (b) WMD – the rectal wall to the prostate marker distance, and (c) DSC – Dice similarity coefficients for the rectum obtained from comparisons of manually created contours (M_0) and the contours generated automatically (M_1, M_2) by A_1 and A_2 algorithms.

mean DSC calculated for the rectum significantly differed ($p < 0.001$) and were 0.82 and 0.78 for A_1 and A_2 algorithms, respectively. The scattergrams presented on Figure 2 show the RWD, WMD and DSC values obtained for A_1 and A_2 algorithms.

4 Discussion

Contouring quality on CBCT images, mainly organ borders, is affected by its intrinsic poor quality in soft tissue regions. The most significant discrepancies detected by imaging modalities superior to CBCT (i.e., magnetic resonance or ultrasonography) are in the bladder-prostate border, the seminal vesicles, and the prostate apex [21]. To assess contours properly in this region, contrast medium (for bladder-prostate border) is introduced, or other artificial fiducial markers used [17]. The retrospective data used in this study do not include the contrast increasing the visibility of the prostate-bladder border or additional fiducials for the precise location of seminal vesicles. Therefore, we chose to evaluate the rectum contour related to fiducial markers in regions of the prostate with relatively good visibility.

To conduct a qualitative analysis, we adapted the 4-level scale used by Huyskens et al. [22] to construct a two-level descriptive scale (meet/fail). Based on our clinical experience, if any correction to contour is necessary, it means a new contouring process for clinicians, which makes it useless in daily routine. Therefore, our simplified scale is such that contours meeting criteria can be used without any correction for radiotherapy planning. The obtained results show both high consistency between tested algorithms and low efficacy (Table 2). One can state that there are unsolvable problems common for both algorithms in the prostate area, which indicates that we need to seek mutual properties leading to unacceptable results.

Image registration as an inherently incorrectly posed problem has, apart from the case of identical images, no single solution. Even for simple translation-rotation transformation, there can be many solutions that lead to identical results [23]. If we allow non-rigid transformation every solution based on permutation of equal intensity points has the same weight in similarity measure terms.

To select physically and functionally proper solutions of deformation, field regularizations are introduced. Different processes for expected and accurate results demand different construction of the cost function [2]. In both algorithms used in the study, there is no choice of cost function terms (regularizations), that are fixed which directly contributes to image registration failure or success [24]. Both algorithms used are in the same solution class [25] and, therefore, can be expected to deliver similar results. We can look for a domain of images where this type of algorithm meets our expectations, and track and explain situations where it fails. In our opinion, the relation or process combining test and reference image must express bounds imposed by cost function terms on the deformation field and has a crucial impact on results.

When CT and CBCT images are almost undifferentiated, the CT is a proper indicator for the whole treatment. We can state that in such situations the benefit of using a deformable image can be irrelevant and results from this analysis can be affected by the internal algorithm and methodical uncertainties. On the other hand, considering that periods between fractions in radiotherapy are long compared to physiological processes, rectum filling, or voiding (with air or content), which results in global mass changing (not due to movement of mass) is genuinely likely to happen. Proper patient preparation is a sufficient but not a necessary condition to avoid that [26]. In terms of image properties, this can be described as appearing or disappearing new voxels with the same or different intensity (air bubbles, new/different content) inside. Such cases do NOT represent a physically determined deformation process. If we apply the algorithm to such a situation, regularization bounds will prevent appropriate results but even if we get an adequate result the deformation field cannot be correct. The type of terms that keeps the algorithm from such non-physical deformation is called topology preservation terms and relates to the invertibility of deformation mapping. There must be no vanishing or overlapping elements. While in A_1 analogy to harmonic functions [15] is used, in A_2 the topology is constructed as an analogy to the bending of a thin sheet of metal and is used to achieve a smooth transformation [27].

Also, both algorithms incorporate terms related to volume or shape preservation. In A_2, the tissue incompressibility term is motivated by the observation that many tissues in the human body are approximately incompressible for small deformations and short time periods [14,27]. Related term penalizes deviation from $J_T(x) = 1$ (Jacobian transformation), which means transformation that preserves volume. Value less than 1 represents a compression and above 1, expansion. Different strategy is developed in A_1 where shape-based regularization is introduced to penalize non-physical and non-physiological changes in anatomical regions defined as a certain type of ROIs. Figure 3a and 3b represent extreme but common examples that blindly deviate a physical deformation process. It is simply impossible to connect CT and CBCT with any physically regularized image deformation relation. Cases c and d in Figure 3 are not so evident because rectal air (regions with certain intensities) occurs in both CT and CBCT. Nevertheless, in case c from Figure 3, volume preservation of DIR counteracts such significant volume changes and in case d from the same figure continuity prevents it from splitting into more structures.

The second part of the study was a quantitative rectum analysis that was done only on selected sets which, in favor of the previous discussion, means physically proper sets which eliminate most uncertainties not related to algorithm properties but to patient/image properties.

There are many measurement tools that allow the overlapping of the different regions/contours to be checked (e.g.,

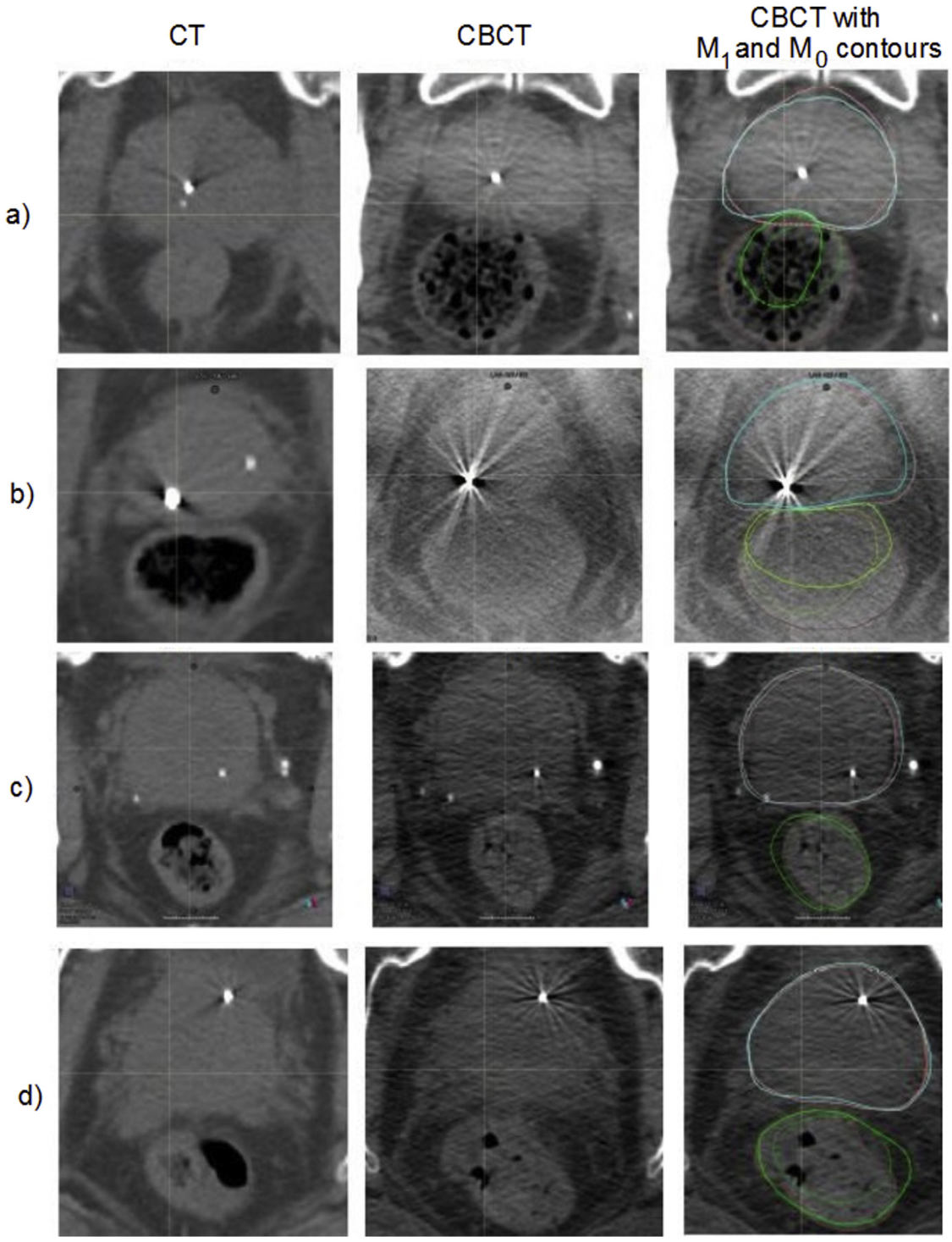

Figure 3. Examples of non-physical processes that occur in the rectum. Rows (a) and (b) are examples of drastic but not infrequent changes that occur in rectal filling. Row (c) represents a slight but substantial difference in filling consistency. Row (d) is an example of random air moving inside the rectum that occurs both on CT and CBCT sets. All situations lead to a deformation algorithm failure represented by incorrect contour propagation (solid green contour in column 3).

Sorensen–Dice similarity coefficient (DSC) [19], Jaccard overlap scores [28]). Nevertheless, as shown by Rohlfing [27], they are unreliable for quantification of image registration accuracy and the results based on it can be misleading. Therefore, they can be used only as supplementary tools to other, more precise methods. There is a consensus that a dense set of landmark points is the proper solution [7]. Because there is a lot of displacements in the pelvic region (e.g., rectal wall can move in all directions without relation to the prostate border), there are no natural stable landmark points related to the rectum or prostate. In our analysis, we used fiducial markers as artificial landmarks and their relation to the rectal wall as a natural and clearly visible physiological barrier. We have excluded bladder and prostate walls from the analysis because they are hardly detectable in CBCT images and lead to unwanted sources of uncertainties.

In this study, we chose measures that are related to different and – as is possible in such an environment – separated aspects of rectum behavior. The first parameter, RWD, can be associated with the radius of rectum planar intersection and is linked to rectal filling. In the context of the previous discussion, we can point out that even on carefully chosen sets with proper physical deformations both algorithms (statistically comparable) are sources of non-negligible uncertainties (5.5 and 5.8 mm for M_1 and M_2, respectively). These results are significantly higher than CBCT site-related manual contouring uncertainties. Lütgendorf-Caucig et al. [29] showed in a multi-observer study that large uncertainty in rectum delineation was observed for CBCT images as compared to CT and MR. They found that the additional PRV margin for CBCT contouring was 1 mm in the right-left, 1 mm in the anterior-posterior, and 3 mm in the supine-inferior direction, and the rectum volume was contoured larger on CT than CBCT images. Weiss et al. [21] showed that limitations in CBCT image quality increase manual contouring discrepancies. Especially interobserver contouring of the rectum in CT *vs* CBCT was significantly larger (rectum average surface mesh 1.3 versus 1.5 mm) and patient averaged rectum volumes were 87 cm^3 on CT *vs* 64.4 cm^3 on CBCT. The second parameter, WMD, is defined as the distance between the fiducial marker in the prostate and the rectal wall. The difference between results obtained for A_1 and A_2 algorithms (i.e., 4.5 and 5.5 mm) was statistically significant ($p = 0.007$). Both algorithms due to using B-spline functions allow only local deformations so the initially assigned position strongly affects the algorithm output. In A_1 all pairs are pre-registered using automatic rigid registration based on fiducial markers that are also used as control POIs (Points of Interest) in the deformation process. On the other hand, in A_2 fiducial markers are only a guide for initial manual rigid pre-registration. In our opinion, this is the reason for the significant difference in algorithms outcome for WMD parameters where fiducials are taken as a reference for measurements. Although DSC is not an accurate tool for registration and deformation algorithm accuracy, it can be utilized as complementary analysis for contouring quality. Results for DSC follow previous analysis and show slightly better performance of A_1 vs A_2 (0.82 vs 0.78; $p < 0.001$). Nevertheless, in our opinion, such low DSC scores (<0.90) always require further visual inspection and its assessment quality is tentative.

Deformable image registration for adaptive radiation therapy is still a tool under development. Analysis of our results complies with other studies [22,30] and shows poor effectiveness in both rectum automatic contour propagation and deformation. There are two paths of possible improvement. First is related to using better imaging modality in daily routine, Torresin et al. [31] which facilitates better visibility in the soft tissue region and performance of DIR registration. The other one is using some artificial intelligence solutions to individualize cases in the automatic solution, which is impossible when using standard algorithmic solutions and always requires human interaction (which was the case in our setup) [32].

5 Conclusion

Neither of the tested algorithms delivered satisfactory outcome in terms of effectiveness and measured uncertainties in the propagation of the rectum contouring process. To avoid incorrect results, it is then essential that input images are carefully selected and each deformable registration verified on a case/set basis.

Conflict of interest

None declared.

References

[1] McVicar N, Popescu IA, Heath E. Techniques for adaptive prostate radiotherapy. Phys Med 2016;32:492–8.
[2] Rigaud B, Simon A, Castelli J, Lafond C, Acosta O, Haigron P, et al. Deformable image registration for radiation therapy: principle, methods, applications, and evaluation. Acta Oncol 2019;58:1225–37.
[3] Barber J, Yuen J, Jameson M, Schmidt L, Sykes J, Gray A, et al. Deforming to best practice: key considerations for deformable image registration in radiotherapy. J Med Radiat Sci 2020;67:318–32.
[4] Sarrut D. Deformable registration for image-guided radiation therapy. Z Med Phys 2006;16:285–97.
[5] Czajkowski P, Piotrowski T. Registration methods in radiotherapy. Rep Pract Oncol Radiother 2019;24:28–34.
[6] Oh S, Kim S. Deformable image registration in radiation therapy. Radiat Oncol J 2017;35:101–11.
[7] Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM Radiation Therapy Committee Task Group No 132. Med Phys 2017;44:e43–76.
[8] Schultheiss TE, Tomé WA, Orton CG. It is not appropriate to "deform" dose along with deformable image registration in adaptive radiotherapy. Med Phys 2012;39:6531–3.
[9] Posiewnik M, Piotrowski T. A review of cone-beam CT applications for adaptive radiotherapy of prostate cancer. Phys Med 2019;59:13–21.

[10] Chen Z, Yang Z, Wang J, Hu W. Dosimetric impact of different bladder and rectum filling during prostate cancer radiotherapy. Radiat Oncol 2016;11:1–8.
[11] Nassef M, Simon A, Cazoulat G, Duménil A, Blay C, Lafond C, et al. Quantification of dose uncertainties in cumulated dose estimation compared to planned dose in prostate IMRT. Radiother Oncol 2016;119:129–36.
[12] Stock M, Pasler M, Birkfellner W, Homolka P, Poetter R, Georg D. Image quality and stability of image-guided radiotherapy (IGRT) devices: a comparative study. Radiother Oncol 2009;93:1–7.
[13] Weistrand O, Svensson S. The ANACONDA algorithm for deformable image registration in radiotherapy. Med Phys 2015;42:40–53.
[14] Rueckert D, Aljabar P, Heckemann RA, Hajnal JV, Hammers A. Diffeomorphic registration using B-splines. In: Larsen R, Nielsen M, Sporring J, editors. MICCAI 2006, LNCS 4191. Springer-Verlag Berlin Heidelberg; 2006. p. 702–9.
[15] Lawson JD, Schreibmann E, Jani AB, Fox T. Quantitative evaluation of a cone-beam computed tomography–planning computed tomography deformable image registration method for adaptive radiation therapy. J Appl Clin Med Phys 2007;8:96–113.
[16] Maes F, Vandermeulen D, Suetens P. Medical image registration using mutual information. Proc IEEE 2003;91: 1699–722.
[17] Salembier C, Villeirs G, De Bari B, Hoskin P, Pieters BR, Van Vulpen M, et al. ESTRO ACROP consensus guideline on CT-and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer. Radiother Oncol 2018;127:49–61.
[18] Thevenaz P, Unser MA. Spline pyramids for intermodal image registration using mutual information. In: Wavelet Applications in Signal and Image Processing V, vol. 3169. International Society for Optics and Photonics; 1997.
[19] Andersen ES, Noe KO, Sorensen TS, Nielsen SK, Fokdal L, Paludan M, et al. Simple DVH parameter addition as compared to deformable registration for bladder dose accumulation in cervix cancer brachytherapy. Radiother Oncol 2013;107:52–7.
[20] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[21] Weiss E, Wu J, Sleeman W, Bryant J, Mitra P, Myers M, et al. Clinical evaluation of soft tissue organ boundary visualization on cone-beam computed tomographic imaging. Int J Rad Oncol Biol Phys 2010;78:929–36.
[22] Huyskens DP, Maingon P, Vanuytsel L, Remouchamps V, Roques T, Dubray B, et al. A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer. Radiother Oncol 2009;90:337–45.
[23] Fischer B, Modersitzki J. Ill-posed medicine – an introduction to image registration. Inverse Probl 2008;24:034008.
[24] Miura H, Ozawa S, Nakao M, Furukawa K, Doi Y, Kawabata H, et al. Impact of deformable image registration accuracy on thoracic images with different regularization weight parameter settings. Phys Med 2017;42:108–11.
[25] Holden M. A review of geometric transformations for nonrigid body registration. IEEE Trans Med Imaging 2007;27:111–28.
[26] Smitsmans MH, Pos FJ, de Bois J, Heemsbergen WD, Sonke J, Lebesque JV, et al. The influence of a dietary protocol on cone beam CT-guided radiotherapy for prostate cancer patients. Int J Rad Oncol Biol Phys 2008;71:1279–86.
[27] Rohlfing T. Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. IEEE Trans Med Imaging 2011;31:153–63.
[28] Jaccard P. The distribution of the flora in the alpine zone. New Phytol 1912;11:37–50.
[29] Lütgendorf-Caucig C, Fotina I, Stock M, Pötter R, Goldner G, Georg D. Feasibility of CBCT-based target and normal structure delineation in prostate cancer radiotherapy: multi-observer and image multi-modality study. Radiother Oncol 2011;98:154–61.
[30] Fabri D, Zambrano V, Bhatia A, Furtado H, Bergmann H, Stock M, et al. A quantitative comparison of the performance of three deformable registration algorithms in radiotherapy. Z Med Phys 2013;23:279–90.
[31] Torresin A, Brambilla MG, Monti AF, Moscato A, Brockmann MA, Schad L, et al. Review of potential improvements using MRI in the radiotherapy workflow. Z Med Phys 2015;25:210–20.
[32] Brock KK. Adaptive radiotherapy: moving into the future. Semin Radiat Oncol 2019;29:181–4.