MEDICAL PHYSICS

Performance of a deformable image registration algorithm for CT and cone beam CT using physical multi-density geometric and digital anatomic phantoms

Anantharaman Ayyalusamy[1,2] · Subramani Vellaiyan[1,3] · Shanmuga Subramanian[2] · Shyama Satpathy[2]

Abstract

Purpose To study the accuracy of deformable registration algorithm for CT and cone beam CT (CBCT) using a combination of physical and digital phantoms.

Materials and methods The physical phantoms consisted of objects over a range of electron densities, shape and sizes. The system was tested for simple and complex scenarios including performance in the presence of metallic artefacts. Clinically present deformations were simulated using a set of five geometric and anatomic virtual phantoms.

Results The system could not account for large changes in size, shape and Hounsfield units. Deformations of low intensity structures and small objects were highly inaccurate, and errors were prominent for volume reduction scenario than volume growth. The presence of artefacts did alter the performance of the algorithm. Objects of low density and that close to artefacts were affected the most. Overall, deformations to CBCT were poor. In virtual phantoms, the system could not handle gas pockets and deformation errors in inverse direction were higher than that in forward direction.

Conclusion The algorithm was tested for several non-clinical and clinical scenarios. The performance was acceptable for realistic and clinically present deformations. However, it is necessary to tread cautiously for structures with small volumes and large reductions in volume.

Keywords Deformable registration · Physical phantom · Virtual phantom · Adaptive radiotherapy

Introduction

Deformable image registration (DIR) has been a promising development in radiotherapy. It is possible to automatically delineate contours using atlases, generate adaptive contours and propagate doses for dose accumulation using deformable image registration [1–3]. Such applications lead to high throughput and thus pave way for adaptive radiotherapy as and when required. DIR algorithms have been described as black boxes [3–5], and it is highly necessary to verify them before clinical use [3–22]. The level of scrutiny is dependent on the application of DIR with dose accumulation requiring intense validation. Studies have validated DIR by using physical phantoms [8, 10, 11, 15–17, 20–22], virtual phantoms [4–6, 9, 14, 19, 23] or a combination of both [7, 12, 18]. Physical phantoms may be geometric or anatomic based. Geometric phantoms are made up of different shapes and densities, to study the basic functionality, while anatomic phantoms are generally constructed based on CT data or pre-defined physical models to simulate real patient anatomy [11, 15, 18]. Physical phantoms can be used for end-to-end testing from image acquisition to validation of DIR algorithm [3]. Wognum et al. [16] studied the performance of DIR algorithms based on a porcine bladder phantom with radiopaque markers and surface matching as a tool to generate the reference deformation. Wu et al. [17] designed an acrylic phantom with two round slots for holding inserts that can be filled with different materials as well be rotated and tilted to test rigid as well as deformable registration algorithms. This phantom was limited by

✉ Subramani Vellaiyan
    vsmani_ananth@yahoo.com

1   Research & Development Centre, Bharathiar University, Coimbatore, Tamil Nadu, India

2   Department of Radiation Oncology, Yashoda Hospitals, Hyderabad, Telangana, India

3   Department of Radiotherapy, All India Institute of Medical Sciences, New Delhi 110029, India

the number of material combinations and the insert shape changes that could be tested. Two- and three-dimensional phantoms that simulate tumour shrinkage and anatomic changes have been constructed [10, 15, 21]. A geometric physical phantom alone cannot be used for DIR validation as performance of algorithms varies with clinical sites. Though site-specific physical phantoms are necessary, phantom construction is time-consuming, and it is impossible to recreate a clinical scenario realistically [12]. This drawback could be overcome using virtual phantoms, modelled for different clinical sites. Virtual phantoms can be created by inducing a known amount of deformation to the clinical image sets using ImSimQA application [4, 14]. The greatest advantage of virtual phantoms is that clinically relevant deformations can be generated easily. Most of the studies have used ImSimQA as it is the only commercial application available at present to generate virtual phantoms [4, 5, 12, 14].

Contour comparison, landmark tracking and voxel analysis have been the commonly used metrics in the literature for DIR validation [4–22]. In landmark tracking, anatomical points or reference points are generated manually in both fixed and moving image sets and the accuracy is determined by comparing mapped location against the reference. Contour analysis is a volumetric analysis method where the deformed contours are compared against the manually delineated reference contours. This is a relatively easy and common method employed for quantitative analysis based on similarity metrics. These methods are sufficient for contour applications as validation is done for selected locations or volumes of interest. In voxel analysis, the generated deformable vector field (DVF) is compared against the ground truth DVF. The changes in intensity and vector displacement in all voxels are analysed to determine the deformation field error [4, 6].

At present, a single method to verify DIR confidently does not exist. Although routine quality assurance (QA) of DIR using virtual phantom is feasible, it must be used in combination with physical phantoms for commissioning as end-to-end testing is not possible [3]. Many studies have used either physical or virtual phantoms for validation. Moreover, studies that have used both physical and virtual phantoms are sparse. In this study, we have used a combination of physical phantom sets consisting of materials over a range of size, shape and Hounsfield units (HU) and ImSimQA generated virtual phantom sets to validate a commercially available DIR system for CT and cone beam CT (CBCT) images. This study is aimed at validating contouring applications of DIR. As per our knowledge, none of the studies in the literature have used a variety of materials with different intensities to determine the accuracy of DIR. We have studied the system performance for several clinical and non-clinical scenarios including performance under the presence of artefacts.

Methods and materials

Physical phantoms

Two phantom sets with materials of different electron densities, shape and size, placed in an air equivalent polyurethane foam base, were constructed in-house for the validation of DIR algorithm. The phantom is easy to assemble, and many combinations of materials can be created to test the algorithm. The first phantom set (P1) was used to study the ability of the system to handle simple changes in material, size and shape, while the second phantom set (P2) was used to check the capability to handle complex deformations like large shape and size changes. The following materials polymethyl methacrylate (PMMA), polyvinyl chloride (PVC), polyoxymethylene (Delrin), nylon, high density polyethylene (HDPE), polypropylene, polytetrafluoroethylene (Teflon) and water-filled balloons were used to study the performance of the system over a range of electron densities. The final configurations of the phantoms were obtained by swapping materials in the initial configuration.

Description of physical phantom—P1

The initial configuration of phantom set P1 consisted of PMMA cylindrical and rectangular rods, two PVC cylindrical rods of equal thickness and different lengths, a cylindrical Delrin rod, a nylon rod of large diameter, a Teflon rectangular cuboid, two HDPE plates of 10 mm and 15 mm thickness, two polypropylene rods and a water-filled balloon of large volume. All objects except Delrin cylinder were replaced in the final configuration. Cylindrical PMMA rod was replaced with a rod of very small diameter and length, while the rectangular PMMA rod was swapped with a cylinder of large diameter. The PVC rods were replaced by a smaller PVC rod of equal thickness and a cylindrical Teflon pipe of similar length. The nylon and polypropylene rods were replaced by similar rods of smaller diameter, while the volume of water balloon was reduced. An additional Teflon rectangular cuboid was added to the pre-existing one, while HDPE plates of 10 mm and 15 mm thickness were reduced to similar plates of 5 mm thickness. Initial configuration of phantom P1 is depicted in Fig. 1a.

Description of physical phantom—P2

The initial configuration of the second phantom set P2 consisted of a cylindrical PVC rod, three PVC pipes, a Teflon pipe, a rectangular HDPE plate of 5 mm thickness, a water-filled balloon with air gap and six fiducials. In the final configuration, the PVC rod was retained, and

Fig. 1 **a** Initial configurations of phantom P1, **b** phantom P2
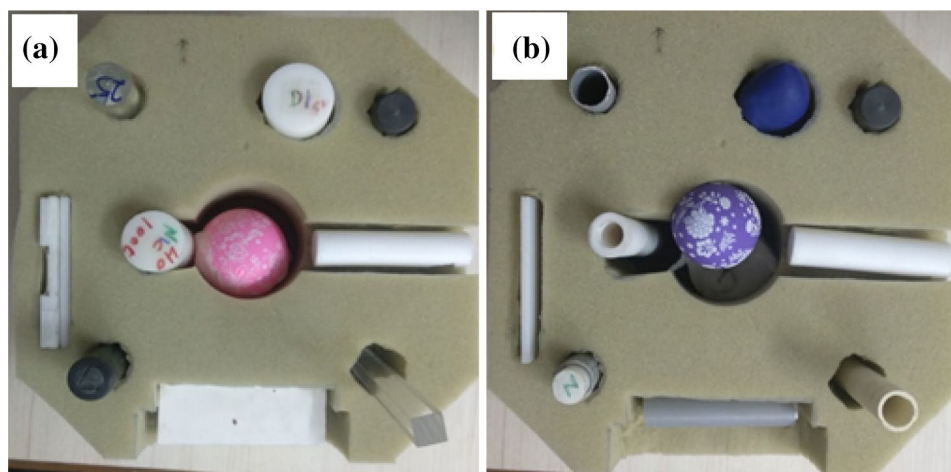
the remaining objects were modified. Three fiducials were removed, while a water balloon without air gap was positioned at a different position. The first PVC pipe was replaced by a double PVC pipe ('8' shape), the second one was substituted by a 'L' bend PVC pipe, while the thickness of the third pipe was doubled. The Teflon pipe was replaced by a twisted PVC pipe, whereas the rectangular HDPE plate was machined to create a 'C' shaped structure. In addition, a new PVC pipe was introduced into the phantom. Initial configuration of phantom P2 is depicted in Fig. 1b.

QA of imaging systems

QA of the image acquisition systems is necessary to avoid errors. Image acquisition process especially image quality can affect the deformation output. Poor acquisition may alter anatomical information, thereby causing differences between images being registered and resulting in erroneous output [3]. The CT scanner checks part of the routine QA of the CT simulation process carried out periodically and also after a major breakdown like tube replacement. Electromechanical, image quality and radiation safety are the three broad categories tested based on the American Association of Physicists in Medicine (AAPM) task group (TG) report 66 using vendor supplied water phantom, CATPHAN 500 (Phantom laboratory, USA) and Unfors X-ray test equipment (RaySafe, USA). Gantry laser alignment and CT number accuracy of water are checked daily. The On-Board Imager system (Varian Medical Systems, Palo Alto, USA) used for CBCT image acquisition is tested for geometric accuracy and image quality based on AAPM TG-179. The kilo-voltage (kV) geometric image stability is verified daily. A detailed imaging QA is beyond the scope of this study and is described in the literature [24–26].

Image acquisition

The initial and final configuration CT images of the phantom sets P1 and P2 were acquired at 120 kVp using a 128 slice GE Optima CT scanner (GE Medical Systems, USA). Another image of the final configuration of phantom set P1 was acquired at 120 kVp with a metallic part that caused streaking artefacts. For the phantom P2, CT and CBCT images of the initial and final configurations were also acquired. The CBCT images were also acquired at 120 kVp using low dose pelvis protocol with half-fan acquisition mode. As the CBCT field of view (FOV) did not cover the phantoms longitudinally, images of superior and inferior parts of phantoms were acquired separately with an overlapping region and then stitched using Slicer3D software (version 10.0). It is an open-source software platform for medical image analysis, visualisation and research in image-guided treatments [27]. The stitched images were imported back into Eclipse treatment planning system, and all the objects were delineated in the contouring workspace.

Virtual phantoms

The virtual phantoms used in this study were generated using ImSimQA (Oncology Systems Limited, Shrewsbury, UK). It is a third-party software used for quality assurance (QA) of radiotherapy treatment planning systems (TPS) and contains a total of 15 virtual phantoms. Detailed description of ImSimQA software and its workflow have been described in the literature [4, 5, 12, 14]. In this study, virtual phantoms were generated from ImSimQA geometric and anatomical phantom library and from real CT image sets of patients. A total of 5 phantoms that included 1 geometrical and 4 anatomical phantoms were generated. The generated anatomical virtual phantoms simulated scenarios that are typically encountered during radiotherapy treatments like, changes in bony anatomy position, weight loss, changes in bladder and

Table 1 Virtual phantom characteristics

S. no.	Phantom	Site	Description	Scenario
1	Geometric	–	QUASAR phantom	Geometrical checks
2	Anatomic	Brain	ImSimQA dataset	Tumour shrinkage
3	Anatomic	Head and neck	Clinical dataset	Change in neck position
4	Anatomic	Abdomen	ImSimQA dataset	Change in tumour position
5	Anatomic	Pelvis	ImSimQA dataset	Prostate movement

Table 2 Selected structures in virtual phantoms

| Phantom	Structures
Quasar	Sphere large, sphere medium, sphere small, cone, air, square
Brain	GTV
Head and neck	Brainstem, cord, eye Lt, eye Rt, parotids, mandible, oesophagus, larynx
Abdomen	GTV Lt, GTV Rt
Pelvis	Bladder, rectum, prostate, femur

rectum filling and reduction in tumour volume. The characteristics of virtual phantom image sets are described in Table 1, and the structures used for evaluation in this study are in Table 2.

The QUASAR, brain and abdomen phantoms were deformed manually with the help of marker points, and a global deformation was applied. The abdomen phantom selected in the study simulated a patient with liver metastases. Head and neck and pelvis phantoms were deformed automatically using ImSimQA task scheduler. The task scheduler allows for choosing the magnitude of deformation required and automatically generates a clinically present deformation. A large magnitude of global deformation was chosen for head and neck and pelvis phantoms. The geometric QUASAR virtual phantom is shown in Fig. 2. The anatomic brain, head and neck, abdomen and pelvis virtual phantoms are shown in Fig. 3. Contours of structures were delineated prior to the generation of reference deformation. The initial and reference images with structure sets were exported to Eclipse planning system to test SmartAdapt algorithm.

Deformable image registration

The deformable registrations were carried out in SmartAdapt module of Eclipse treatment planning system (Version 13.6, Varian medical systems, Palo Alto, USA). It is based on a modified, *accelerated demons algorithm* proposed by Wang et al. [7] which utilises the intensity information from the fixed image to determine the demons force required to deform the moving image. The dependence of DIR on image artefacts and changes in shape, size and image intensities were studied using the phantom P1. The initial image set was deformed to the final configuration with and without metallic part to understand the performance of algorithm in the presence of streaking artefacts.
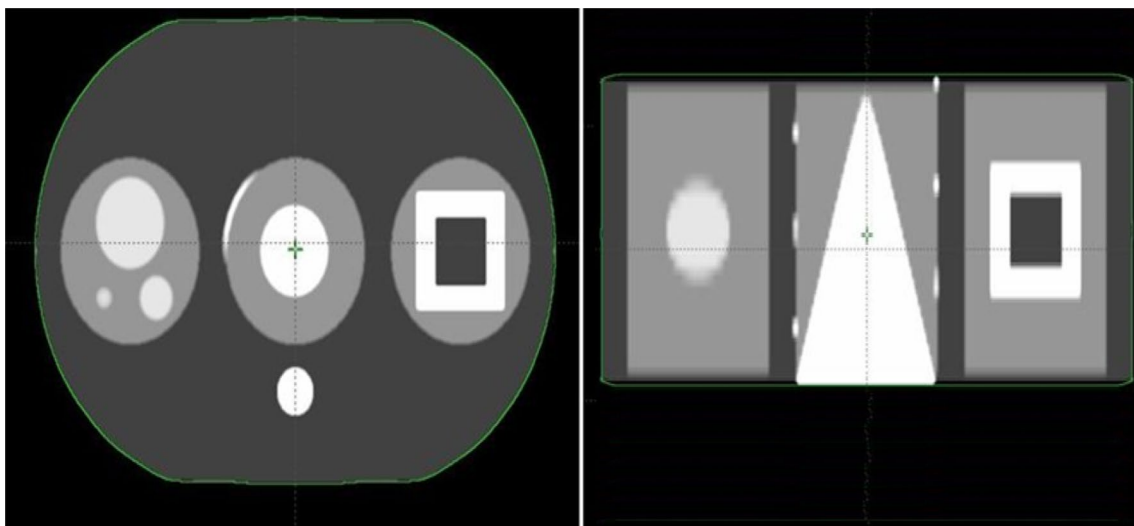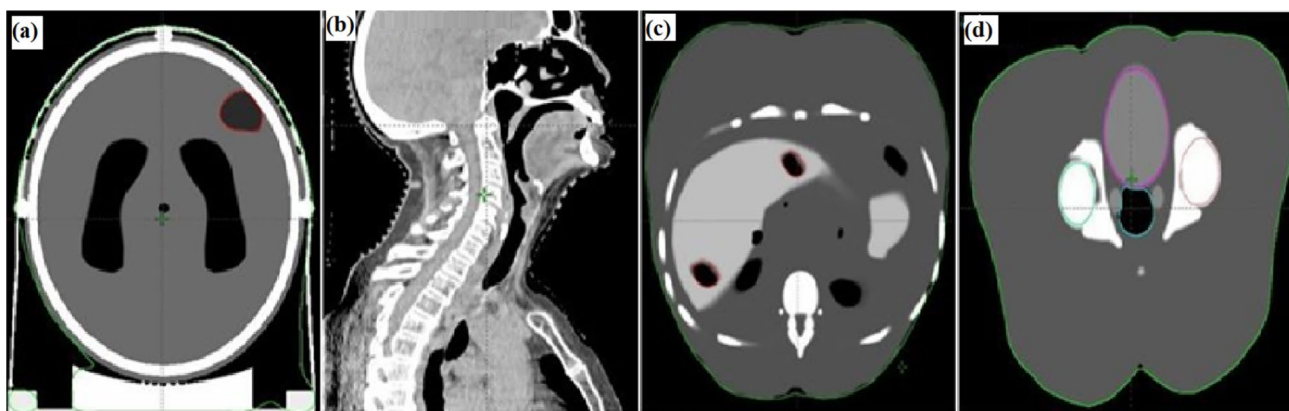

Fig. 2 QUASAR virtual phantom

Fig. 3 Anatomic virtual phantoms **a** brain, **b** head and neck, **c** liver and **d** male pelvis

Complex deformation and double deformation errors were studied using phantom P2. Deformations were carried out between the initial and final configurations of CT and CBCT phantom images, and four combinations were generated. In addition, three combinations of double deformation were also generated. The deformation in one direction immediately followed by its inverse is double deformation. The CT_{FINAL} to $CT_{INITIAL}$, CT_{FINAL} to $CBCT_{INITIAL}$, $CBCT_{FINAL}$ to $CBCT_{INITIAL}$ deformations were taken as the reference against which the inverse deformations of CT_{INITAL} to CT_{FINAL}, $CBCT_{INITAL}$ to CT_{FINAL}, $CBCT_{INITAL}$ to $CBCT_{FINAL}$ were compared. The deformations were evaluated quantitatively using contour analysis metrics and qualitatively by checking for the presence of fiducials and water balloon.

The initial and ImSimQA modified virtual phantom CT images were imported into Eclipse treatment planning system. The initial image sets were deformed to ImSimQA generated image sets and vice versa. Eclipse DVFs and deformed structure sets were exported to ImSimQA for quantitative analysis. The deformation was evaluated by contour and deformation field error analysis methods. The deformations were also inspected visually.

Analysis metrics

Contour analysis for physical phantom deformations was done using Slicer3D software. Dice similarity index (DSI) and Hausdorff distance (HD) were computed to evaluate physical phantom deformations. Virtual phantom deformations were evaluated in ImSimQA using mean distance to conformity (MDC), DSI and deformation field error metrics. DSI is a volume-based similarity measure used to evaluate quantitatively any two sets of contours. It is used to determine the degree of overlap between the manually contoured and system delineated structures. DSI is the standard measure for volume mismatch and does not measure surface

errors. If A and B be the volume of deformed and reference structures, then DSI is defined as

$$DSI = \frac{2(A \cap B)}{(A + B)}$$

A value of '0' indicates no overlap, and value of '1' indicates absolute overlap. HD is a measure of degree of mismatch between two image sets based on contour boundaries. It is defined as the maximum distance between a point in one image set and the corresponding point on another image set. MDC is the mean of distances between points in one image to the corresponding points in another image. It is analogous to mean HD value for all the points in consideration. The deformation field error was determined by subtracting Eclipse generated DVFs from ImSimQA reference DVFs. The maximum and mean deformation field errors were computed. A short description of the methodology is shown in Table 3.

Results

Physical phantom P1

The ability of algorithm to handle basic changes in size, shape, material and image quality was studied using phantom P1. DSI values between deformed and reference structures were determined using Slicer3D software. The DSI for phantom P1 without and with artefact is shown in Fig. 4. HDPE2, PMMA CS, Teflon cylinder and PMMA rectangle had poor DSI scores in both normal and artefact image series. The artefacts did affect the deformation. In the artefact image set, differences in DSI could be observed for nearly all the structures. Magnitude of difference in scores varied among the structures. Large decrease in nylon score was observed. The HD values for phantom set P1 without

Table 3 Short description of the methodology

Phantoms		Type	Simulated deformation	Image modality	Analysis software	Metrics
Physical phantom	P1	Geometric	Simple	Plain and artefact CT	Slicer3D	DSI, HD
	P2		Complex	CT and CBCT	Slicer3D	DSI, HD
Virtual phantoms	Brain, head and neck, abdomen, pelvis and quasar	Anatomic and geo-metric	Clinically present	CT	ImSimQA	DSI, MDC, CGD

DSI dice similarity index, *HD* Hausdorff distance, *MDC* mean distance to conformity, *CGD* centre of gravity distance
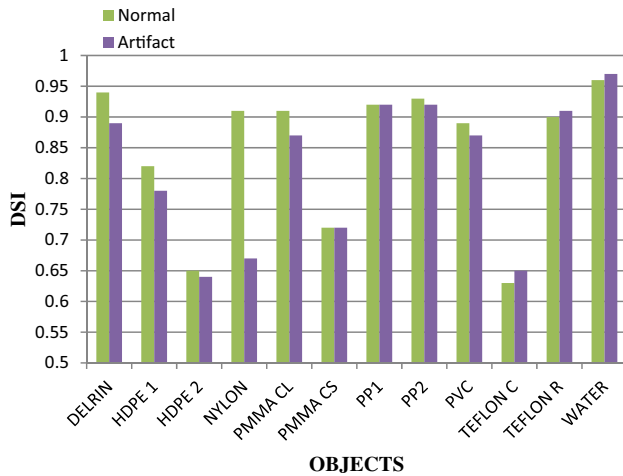

Fig. 4 DSI variation with artefact. *PMMA CL* PMMA cylinder of large diameter, *PMMA CS* PMMA cylinder of small diameter, *PP*1 polypropylene cylinder 1, *PP*2 polypropylene cylinder 2, *TEFLON C* Teflon cylinder, *TEFLON R* Teflon rectangle, *PMMA R* PMMA rectangle

Fig. 5 HD variation with artefact. *PMMA CL* PMMA cylinder of large diameter, *PMMA CS* PMMA cylinder of small diameter, *PP*1 polypropylene cylinder 1, *PP*2 polypropylene cylinder 2, *TEFLON C* Teflon cylinder, *TEFLON R* Teflon rectangle, *PMMA R* PMMA rectangle

and with artefact are shown in Fig. 5. HDPE2 and PMMA CS had high HD values in normal image set. A remarkable increase in HD values was noted for nylon, PMMA CL and PMMA CS in the artefact scan. Low DSI scores correlated with high HD values for most structures.

Physical phantom P2

The tests using phantom P2 determined the ability of algorithm to handle complex deformation and double deformation error. DSI and HD for the set of four possible complex deformations are given in Table 4. Complex shape and size changes resulted in poor DSI scores. The double deformation errors were evaluated using DSI and HD metrics. The DSI and HD scores for the three combinations of double deformation are depicted in Figs. 6 and 7, respectively. An increase in magnitude of complicated changes resulted in poorer metric scores. Double deformation errors were more prominent for complex structures. Intra-CBCT deformation scores were comparable to intra-CT deformation. Deformations with CBCT image as the target were less accurate than
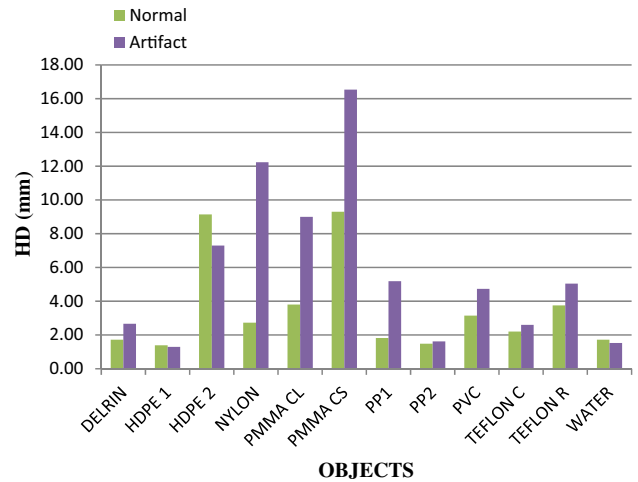
that with CT as the target image. Deformation to CBCT images with drastic changes produced bad results. The system could annul the three fiducials completely in deformations to CT, while it could not nullify them in deformations to CBCT. In double deformation, only the reference fiducials were retained. In addition, the water structure could not be accounted for perfectly. It could only shrink the water structure partially but not null it perfectly.

Virtual phantoms

The deformation field errors were analysed for all the virtual phantoms. The maximum and mean deformation errors were computed within the body structure. The deformation field error in forward and inverse directions for QUASAR, brain, head and neck, abdomen and pelvis phantoms is represented in Table 5. The Eclipse deformed structure sets were compared against ImSimQA reference structure sets in the contour comparison module of ImSimQA. MDC, CGD and DSI scores were computed to determine the accuracy of SmartAdapt DIR algorithm. The metric scores for Quasar,

Table 4 DSI and HD for complex deformation

Material		Metric	CT initial to CT final	CT initial to CBCT final	CBCT initial to CBCT final	CBCT initial to CT final
HDPE		DSI	0.7	0.13	0.67	0.62
		HD (mm)	7.67	29.57	7.17	17.94
PVC	Cylinder	DSI	0.92	0.62	0.96	0.87
		HD (mm)	1.95	6.13	1.33	5.81
	Pipe	DSI	0.68	0.22	0.6	0.59
		HD (mm)	5.46	28.9	5.09	18.53
	Double pipe	DSI	0.42	0.05	0.71	0.59
		HD (mm)	12.38	23.37	5.75	6.01
	L bend	DSI	0.39	0.22	0.15	0.58
		HD (mm)	1.33	6.13	13.54	23.3
	Thick pipe	DSI	0.55	0.2	0.48	0.8
		HD (mm)	18.09	83.57	72.32	5.81
	Twisted	DSI	0.23	0.13	0.32	0.19
		HD (mm)	147.92	150.36	129.23	116.19

brain, head and neck, abdomen and pelvis phantoms are given in Table 6.

Discussion

Automation in radiotherapy is heavily dependent on DIR. Validation of deformable registration is essential prior to its clinical use. We have used physical phantoms fabricated in-house, with materials over a range of electron densities and a set of virtual phantoms to validate the DIR algorithm in Eclipse SmartAdapt module. The basic and complex deformations in physical phantoms were used to study the ability of algorithm to account for changes in shape, size, quality and density. Clinically relevant deformations were simulated using virtual phantoms.

In the physical phantom P1, poor metric scores for materials HDPE2, PMMA CS, TEFLON cylinder and PMMA rectangle can be attributed to the drastic change in material HU and large volume differences in final configuration image. HDPE1 with 5mm reduction in thickness was accounted easily, whereas HDPE2 with 10 mm reduction resulted in inaccurate deformation. Similarly, a large change in thickness of PMMA CS affected the DIR output. Simple changes like variation in length as in PP1, PP2 and PVC rod and volume reduction as in water balloon could be handled easily, while large size variation resulted in errors. This is in line with that reported in the literature studies [12, 18]. Stanley et al. [18] have concluded that the accuracy of the algorithm is determined by the magnitude of deformation and large displacements result in large errors.

In our study, we found that errors were more pronounced for volume reduction than volume growth and the system could not account moderate to large changes in

HU. HDPE2 plate could not be shrunk effectively, whereas growth in PMMA cylinder of large diameter and rectangular Teflon cuboid size could be handled. The DIR algorithm could not nullify or create an air gap, when a solid material was replaced by a hollow cylinder or vice versa. Brock et al. [13] have also reported that deformation in the presence of gas pockets is challenging. The presence of artefacts likewise did cause an appreciable difference in the DIR output though a trend similar to that in 'null artifact' image set was observed. Metallic artefacts alter the electron density information and increase noise levels thereby affecting the DIR performance. Metrics for structures close to metallic artefact exhibited a larger deviation. Nylon was the most affected due to the distance factor. Low density structures were more susceptible to the presence of artefacts than high density structure like PVC rod.

The capability of DIR algorithm to handle complex deformations was tested using phantom P2. Deformations to CT were better than deformations to CBCT because the DIR output is highly dependent on noise in the fixed image set. Noise in CBCT scans affected the accuracy of deformable registration. Stanley et al. [18] added noise to phantom to simulate CBCT images and also concluded that accuracy degraded with all phantoms. In our study, we also compared intra-modality deformations. Performance of the system is comparable in case of intra-modality deformations because of similar noise levels in both fixed and moving image sets. Hence, there are similar results for intra-CBCT and intra-CT deformations. The system could not create fiducials and objects in a previously empty region. Such unrealistic deformation cannot be expected. The DIR system could not revert back the deformation to original state and hence poor DSI and HD metrics in double deformation.
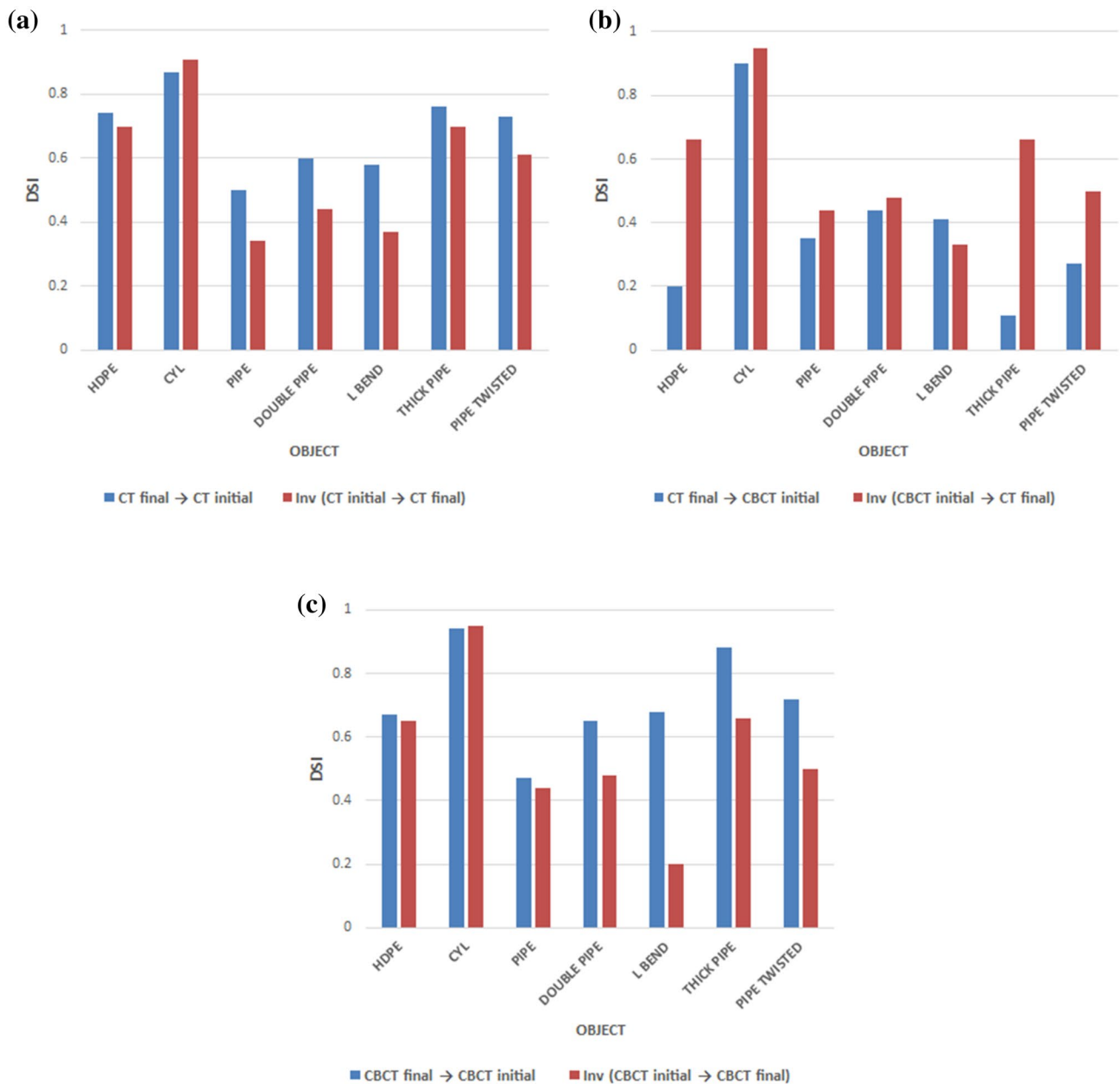
Fig. 6 DSI for three combinations of complex double deformation **a** final and initial CT, **b** final CT and initial CBCT and **c** final and initial CBCT

Deformation field error and contour analysis metrics were used to assess the virtual phantom deformations. The deformation field error was estimated for a global deformation inside the body outline. We have used maximum and mean deformation errors. The maximum errors were seen near bone–tissue interface, tissue–air interface and in regions of uniform density as in brain. The errors in the inverse direction were slightly higher for all phantoms. Contour analysis revealed that DIR of small volumes as well as low intensity structures in low-contrast regions like oesophagus were inaccurate. The performance improved in regions with high intensity differences. Yeo et al. [20] have shown that DIR algorithms fare poorly in regions of low contrast and uniform density. Poor performance in low-contrast regions is an inherent drawback of the original demons algorithm [18]. Although the *accelerated demons* algorithm used in the study is much more accurate and does perform better, its performance in uniform low-contrast regions has not been up to the mark. The poor scores for oesophagus in head and neck virtual phantom may be attributed due to the same reason. The algorithm used in this study is an *intensity-based demons algorithm*.
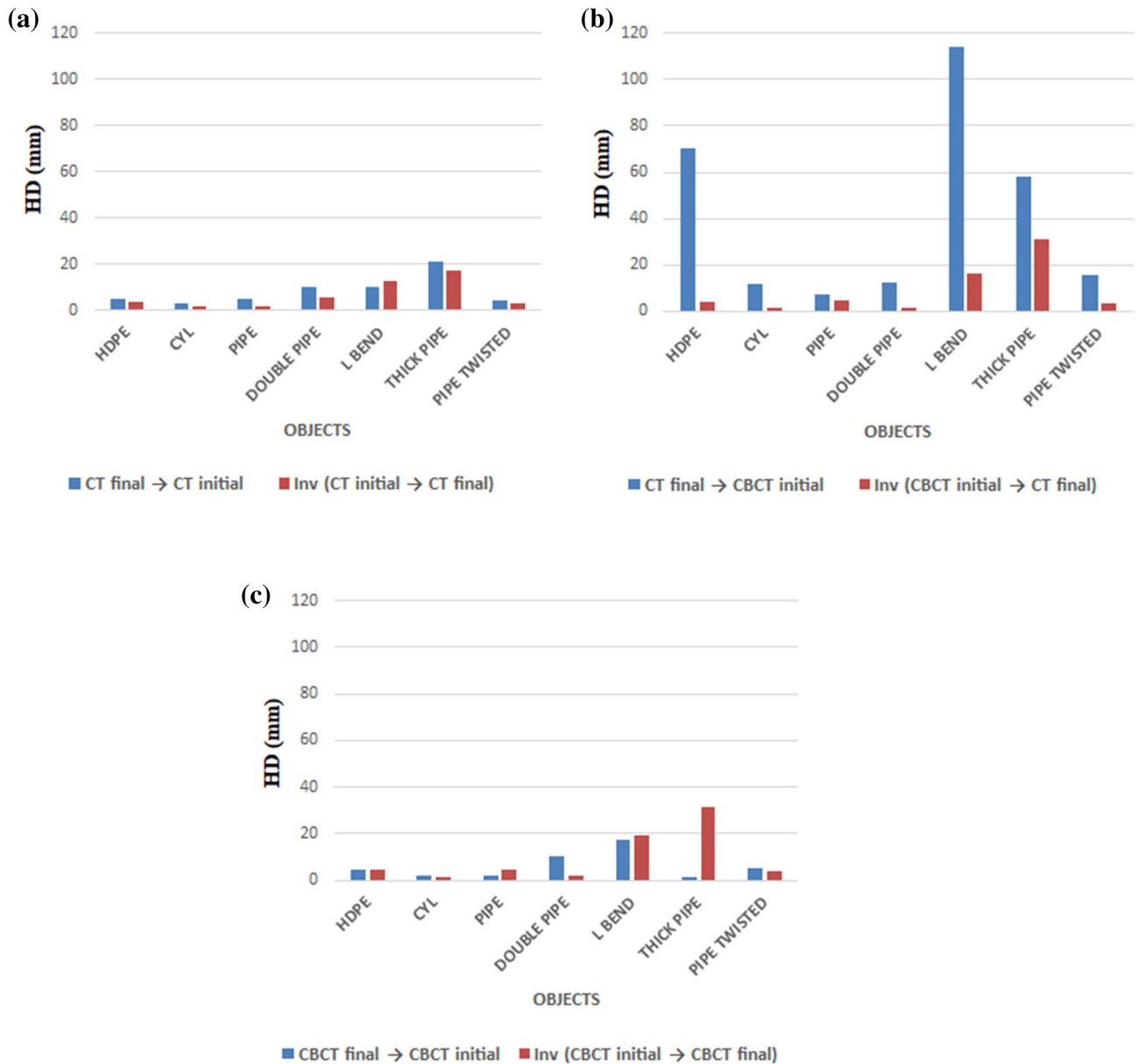
(a)

(b)

(c)

Fig. 7 HD for three combinations of complex double deformation **a** final and initial CT, **b** final CT and initial CBCT and **c** final and initial CBCT

Table 5 Deformation field error analysis for forward and inverse directions

S. no.	Phantom	Forward direction (mm)		Inverse direction (mm)	
		Maximum	Mean	Maximum	Mean
1	QUASAR	25.90	13.00	27.41	13.70
2	Brain	32.97	16.49	33.50	16.80
3	HN	34.50	17.20	35.12	17.67
4	Abdomen	13.40	6.70	13.9	6.86
5	Pelvis	35.20	17.60	35.45	17.93

The presence of non-uniform intensity gradients improves the algorithm performance. The tumours could be clearly distinguished from normal liver due to intensity differences. Prostate movement is due to the changes in bladder and rectal filling. Brock et al. [13] in their MIDRAS study have evaluated 21 DIR algorithms for 4D CT images of lung, liver and MR images of prostate. They concluded that DIR for lungs and liver was acceptable, whereas for prostate, the deformation of prostate due to the introduction of gas in rectum was not accurate. In our study, DIR could account for prostate motion; however, it could not handle large gas pockets introduced in rectum.

Table 6 Contour comparison metrics for virtual phantom SmartAdapt contours

Phantom	Structures	Metrics		
		MDC (mm)	CGD (mm)	DSI
Quasar	Sphere Large	3.07	2.52	0.86
	Sphere Medium	2.59	1.17	0.56
	Sphere Small	3.62	2.62	0.21
	Cone	4.81	2.51	0.89
	Air	2.61	1.85	0.88
	Square	3.76	2.84	0.57
Brain	GTV	3.5	4.09	0.71
HN	Brainstem	1.96	1.38	0.85
	Cord	2.05	1.76	0.82
	Eye Lt	2.54	1.64	0.88
	Eye Rt	2.55	1.57	0.87
	Parotid Lt	3.92	2.31	0.83
	Parotid Rt	3.86	2.25	0.82
	Mandible	2.44	1.98	0.91
	Oesophagus	4.36	3.65	0.61
	Larynx	3.99	3.54	0.69
Abdomen	GTV Lt	1.32	0.37	0.93
	GTV Rt	3.05	2.03	0.80
Pelvis	Bladder	3.20	2.55	0.88
	Rectum	4.22	1.52	0.79
	Prostate	2.65	1.50	0.82
	Femur Lt	2.58	1.50	0.96
	Femur Rt	2.54	1.52	0.96

MDC mean distance to conformity, *CGD* centre of gravity distance *DSI* dice similarity index

The validation presented in the study is focussed on the prominent DIR applications, namely adaptive contouring and auto-segmentation. Dose deformation is the most promising application as well as the missing link in online adaptive radiotherapy. Although dose deformation in radiotherapy is available commercially, it is still in nascent stages. Verification of dose deformation is complex and requires stringent validation and QA. We have not validated the algorithm for dose deformation as it requires a detailed deformation field error data, especially structure-specific information. In the future, we consider validating the same.

We have used a combination of physical and virtual phantoms for simple, complex and clinically present deformations. Though complex deformations as presented in the study are not seen clinically, we wanted to understand the behaviour of the system under extreme conditions. The DIR algorithm could handle simple deformations effectively over a range of electron densities; however, that involving highly complex shape and very large volume changes, structures with small volume, presence of large gas pockets and

structures with low-contrast and homogeneous intensity resulted in poor accuracy.

Conclusion

The DIR algorithm was capable of handling clinically present deformations for most scenarios, but the performance varied with large changes in size and shape of objects, intensity and artefacts. The deformation produced under such conditions should be scrutinised. The validation method presented helps to understand the strength and drawbacks of the algorithm before clinical use.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standard This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent This article does not contain patient data.

References

1. Kadoya N (2014) Use of deformable image registration for radiotherapy applications. J Radiol Radiat Ther 2(2):1042
2. Oh S, Kim S (2017) Deformable image registration in radiation therapy. Radiat Oncol J 35(2):101
3. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML (2017) Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM Radiation Therapy Committee Task Group No. 132. Med Phys 44(7):e43–e76
4. Varadhan R, Karangelis G, Krishnan K, Hui S (2013) A framework for deformable image registration validation in radiotherapy clinical applications. J Appl Clin Med Phys 14(1):192–213
5. Pukala J, Johnson PB, Shah AP, Langen KM, Bova FJ, Staton RJ et al (2016) Benchmarking of five commercial deformable image registration algorithms for head and neck patients. J Appl Clin Med Phys 17(3):25–40
6. Fatyga M, Dogan N, Weiss E, Sleeman WC IV, Zhang B, Lehman WJ et al (2015) A voxel-by-voxel comparison of deformable vector fields obtained by three deformable image registration algorithms applied to 4DCT lung studies. Front Oncol 4(5):17
7. Wang H, Dong L, O'Daniel J, Mohan R, Garden AS, Ang KK et al (2005) Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. Phys Med Biol 50(12):2887
8. Kashani R, Hub M, Balter JM, Kessler ML, Dong L, Zhang L et al (2008) Objective assessment of deformable image registration in radiotherapy: a multi-institution study. Med Phys 35(12):5944–5953
9. Castillo R, Castillo E, Guerra R, Johnson VE, McPhail T, Garg AK et al (2009) A framework for evaluation of deformable image

registration spatial accuracy using large landmark point sets. Phys Med Biol 54(7):1849

10. Kirby N, Chuang C, Pouliot J (2011) A two-dimensional deformable phantom for quantitatively verifying deformation algorithms. Med Phys 38(8):4583–4586

11. Kirby N, Chuang C, Ueda U, Pouliot J (2013) The need for application based adaptation of deformable image registration. Med Phys 40(1):011702

12. Nie K, Chuang C, Kirby N, Braunstein S, Pouliot J (2013) Site specific deformable imaging registration algorithm selection using patient-based simulated deformations. Med Phys 40(4):041911

13. Brock KK (2010) Deformable registration accuracy consortium. Results of a multi-institution deformable registration accuracy study (MIDRAS). Int J Radiat Oncol Biol Phys 76(2):583–596

14. Pukala J, Meeks SL, Staton RJ, Bova FJ, Mañon RR, Langen KM (2013) A virtual phantom library for the quantification of deformable image registration uncertainties in patients with cancers of the head and neck. Med Phys 40(11):111703

15. Singhrao K, Kirby N, Pouliot J (2014) A three-dimensional head and neck phantom for validation of multimodality deformable image registration for adaptive radiotherapy. Med Phys 41(12):121709

16. Wognum S, Heethuis SE, Rosario T, Hoogeman MS, Bel A (2014) Validation of deformable image registration algorithms on CT images of ex vivo porcine bladders with fiducial markers. Med Phys 41(7):071916

17. Wu RY, Liu AY, Wisdom P, Zhu XR, Frank SJ, Fuller CD et al (2019) Characterization of a new physical phantom for testing rigid and deformable image registration. J Appl Clin Med Phys 20(1):145–153

18. Stanley N, Glide-Hurst C, Kim J, Adams J, Li S, Wen N et al (2013) Using patient-specific phantoms to evaluate deformable image registration algorithms for adaptive radiation therapy. J Appl Clin Med Phys 14(6):177–194

19. García-Mollá R, de Marco-Blancas N, Bonaque J, Vidueira L, López-Tarjuelo J, Perez-Calatayud J (2015) Validation of a deformable image registration produced by a commercial treatment planning system in head and neck. Phys Med 31(3):219–223

20. 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

21. Chang J, Suh TS, Lee DS (2010) Development of a deformable lung phantom for the evaluation of deformable registration. J Appl Clin Med Phys 11(1):281–286

22. Graves YJ, Smith AA, McIlvena D, Manilay Z, Lai YK, Rice R et al (2015) A deformable head and neck phantom with in vivo dosimetry for adaptive radiotherapy quality assurance. Med Phys 42(4):1490–1497

23. Zhong H, Kim J, Chetty IJ (2010) Analysis of deformable image registration accuracy using computational modeling. Med Phys 37(3):970–979

24. Mutic S, Palta JR, Butker EK, Das IJ, Huq MS, Loo LN, Salter BJ, McCollough CH, Van Dyk J (2003) Quality assurance for computed-tomography simulators and the computed-tomography-simulation process: report of the AAPM Radiation Therapy Committee Task Group No. 66. Med Phys 30(10):2762–2792

25. Garcia-Ramirez JL, Mutic S, Dempsey JF, Low DA, Purdy JA (2002) Performance evaluation of an 85-cm-bore X-ray computed tomography scanner designed for radiation oncology and comparison with current diagnostic CT scanners. Int J Radiat Oncol Biol Phys 52(4):1123–1131

26. Bissonnette JP, Balter PA, Dong L, Langen KM, Lovelock DM, Miften M, Moseley DJ, Pouliot J, Sonke JJ, Yoo S (2012) Quality assurance for image-guided radiation therapy utilizing CT-based technologies: a report of the AAPM TG-179. Med Phys 39(4):1946–1963

27. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S et al (2012) 3D Slicer as an image computing platform for the quantitative imaging network. Magn Reson Imaging 30(9):1323–1341