

1 Deformable Image Registration and Validation

1.1 Introduction

Most modern clinics today regularly use different imaging techniques including CT, 4D-CT, cone-beam CT, MRI and PET. Registration is needed to overlay different image acquisitions, such as image taken at different days or between imaging modalities or to quantify anatomical changes in a time-resolved image acquisition (4D-CT). While most commercial treatment planning software provide rigid registration between different images, deformable image registration (DIR) is currently rarely used. DIR can quantify anatomical and biological variations better compared to rigid registration [Sarrut, 2006]. It opens exciting new options, such as 4D optimization [Trofimov et al., 2005], 4D dose calculation [Flampouri et al., 2006] or contour propagation [Lu et al., 2006]. DIR usage spreads even into other categories, besides radiotherapy [Cleary and Peters, 2010, Her, 2012, Nithiananthan et al., 2011, Naini et al., 2010]. Several different DIR algorithms are available, such as B-spline [Rueckert et al., 1999], Demons [Thirion, 1998], linear elastic finite element [Venugopal et al., 2005], optical flow [Zhong et al., 2007] or viscous fluid [Christensen et al., 1996].

One of the reasons why DIR is not used in commercial software is the lack of proper DIR quality assurance (DIRQA). While several different DIRQA methods exist, none of them are definitive and most of them are time consuming. It is possible to evaluate DIR with deformable phantoms, where the type and size of deformation is known [Kashani et al., 2007, Kirby et al., 2011]. However this effort is prohibitive in everyday clinical work flow. DIR validation can also be based on landmark positions, specifically their location before and after registration. In absence of externally planted markers, locating landmarks in patient anatomy can be time-consuming and it can be difficult to identify landmarks in low-contrast regions [Varadhan et al., 2013]. Another option is to compare delineated contours with the propagated ones with dice similarity coefficient [Varadhan et al., 2013] or Hausdorff distance [Huttenlocher et al., 1993]. It is more efficient technique than landmark checks, however it requires additional delineation and it does not address the region within the contour.

In this chapter tools to handle DIR and DIRQA will be presented. Tools were constructed for the open-source software 3D Slicer, described in Section 1.2.1. In Section 1.2.2 details on how DIR was performed will be explained. Next, Section 1.2.3 will address the DIR validation - an

extensive tool was created with various functionalities to tackle DIRQA. DIR and DIRQA tools were tested on a real patient data and results will be presented in the Section 1.3.

1.2 Implementation

1.2.1 3D Slicer

3D Slicer (Slicer) is a software platform for analysis and visualization of medical images [Slicer, 2016a, Fedorov et al., 2012]. Slicer is a free, open-source software (BSD-style license) available on Windows, MacOSX and Linux operating systems. It comes with a vast variety of tools, such as:

- Handling of a variety of image formats, including DICOM, NRRD and MHA
- Visualization of voxel images, polygonal meshes and volume renderings
- Image registration (rigid and non-rigid) and display of vector fields
- Automatic image segmentation
- Analysis and visualization of diffusion tensor image data
- Device tracking for image-guided procedures

The source code of Slicer is written in C++ and it's wrapped with Python to provide rapid, iterative development. The graphical user interface is based on Qt. Visualization is based on VTK, a graphical library commonly used in scientific research.

Slicer is a research tool and as such allows implementation of new functionalities in the form of extensions (modules). They can either be as external command-line programs, as scripts to automate Slicer processes or as unique modules with new features. In the next sections, different Slicer modules will be presented, which were all developed in the scope of this thesis. The purpose of these modules is to perform DIR and provide validation of the obtained results.

Registration nomenclature

To provide a clear and consistent description of methods used, an overview of the expressions is given here.

- **Reference image** - image that serves as a reference position in registration (image that is registered to).
- **Moving image** - image that is matched to the reference image (image that is registered from).
- **Warped image** - result of applying a transformation map from registration to the moving image. It should be as close to the reference image as possible.
- **True registration** - registration from moving to reference image. Similar, everything connected to true registration will use `true`(true vector field, true warped image, true absolute difference, true Jacobian, etc.).
- **Inverse registration** - registration from reference to moving image (opposite or inverse of true registration). As in true registration, the term inverse can be used for everything connected to it (inverse vector field, inverse warped image, inverse absolute difference, inverse Jacobian, etc.).

In radiotherapy true registration is used for dose propagation and consequential 4D Dose calculation, whereas with inverse registration contours can be propagated from reference to moving phase.

1.2.2 Registration

Plastimatch [Shackleford et al., 2010] is a commonly used software for registration in medical research. It is a free and open-source software, available as a command-line executable program. Plastimatch B-spline registration is also available in Slicer as part of an extension SlicerRT [Pinter et al., 2012]. The integration of Plastimatch in Slicer brings the advantage of a graphical user interface and hence a quick modification of parameters and display of results. However, for a large number of registration an automatic process is needed. For a complete 4DCT registration there are $2(N - 1)$ registrations required - from reference phase to each of N phases of 4DCT and vice versa, except for the reference phase itself. Typical 4DCTs consist of 10 phases, therefore a automatic registration of a 4DCT is a necessity.

Automatic DIR was achieved with a Python class to handle image locations, store DIR parameters, to perform DIR in the Plastimatch module, to use correct naming conventions and to store all output files (vector fields and warped images). The Python class is shown in Appendix 1.5.

1.2.3 Registration Quality Checks

In order to provide visual and quantitative assessment of the registration quality a **Deformable Image Registration Quality Assurance** module was created. It provides different image checks (inverse color, checkerboard, absolute difference, flicker, movie and landmark positions) and vector checks (Jacobian and inverse consistency error). Details on all different checks will be explained in this section. Reference and warped image, true and inverse vector are used as inputs for DIRQA module. Additionally, landmarks and region of interest (ROI) can also be used as an input.

Absolute difference, Jacobian and inverse consistency error were build using tools from the ITK library [Yoo et al., 2002].

Inverse color

Overlaying two different images will highlight the difference between them. However, since CT scans are usually displayed in grayscale color code, the differences can be indistinguishable. Especially if the images are quite similar, as reference and warped image should be. With applying opposite color codes to overlaying images two things are achieved. First, regions where the registration was successful will be in grayscale. And second, the differences between images will be in the color of the image they originate from. In the module we used red and cyan color code for reference and warped image, respectively. See Fig. 1.1 for details.

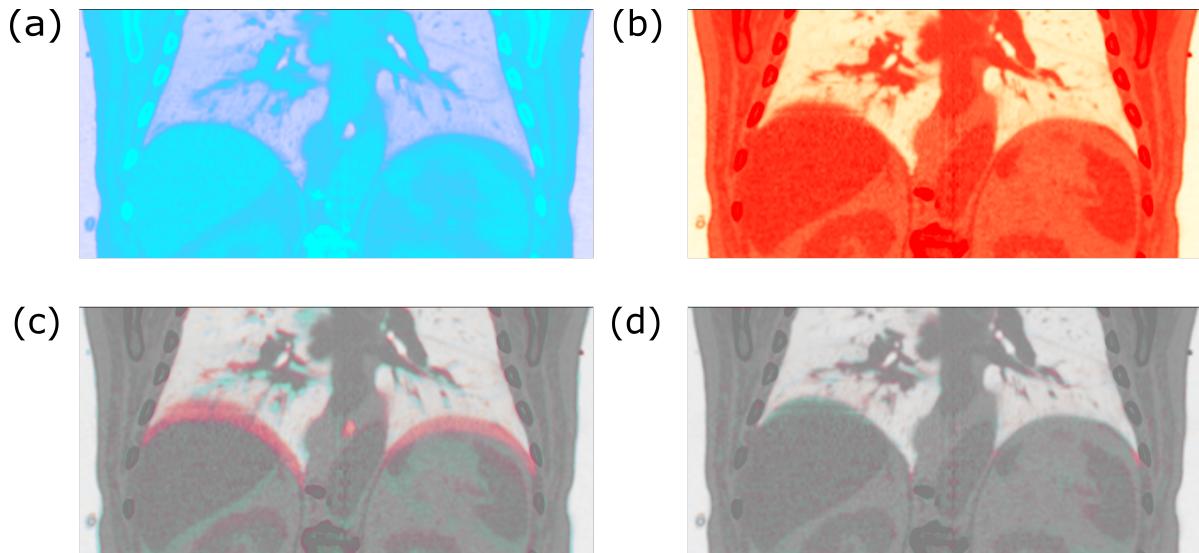


Figure 1.1: Example of a inverse color overlay. Images (a) and (b) show red and cyan color code, respectively, for a CT scan. (c) displays overlaid inverse colored images before and (d) after registration.

Checkerboard

As the name suggests, checkerboard creates an image of tiles. Each tile alternates between reference and warped image, as shown in Fig. 1.2. The differences between two images become apparent if there is no smooth transition from one tile to the next. The number of tiles can be manually selected. While the checkerboard offers clear indication of differences, it requires a user to spot them.

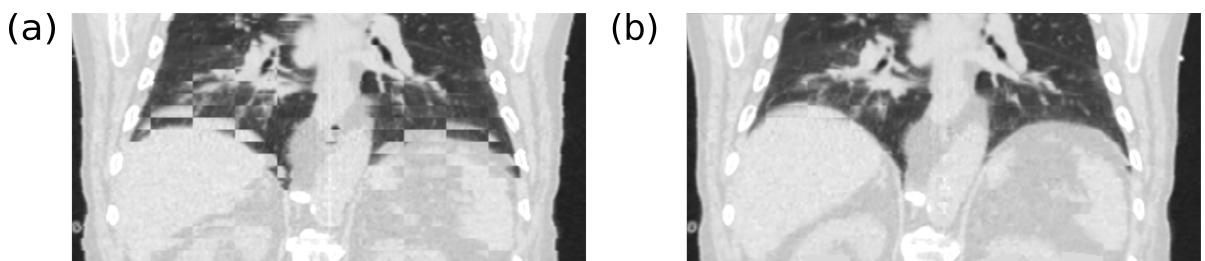


Figure 1.2: Example of a checkerboard image. It consists of tiles alternating between two images. Tiles in image (a) alternate between scans before registration and in image (b) after registration.

Absolute difference

To stress the difference between reference and warped image an absolute difference between voxel values is calculated and displayed. A new image is generated with voxels populated as the absolute difference between reference and warped image voxel values, as shown in Fig. 1.3. Furthermore, statistical values of absolute differences are calculated (mean, standard deviation, minimum and maximum) for quantitative assessment of registration quality (in the ideal case all values would be 0).

Three absolute differences can be calculated: between reference and moving image, called default absolute difference; between true warped and reference image, called true absolute difference; between inverse and moving image, called inverse absolute difference. Usually, registration works on minimizing absolute difference (mean square error metric), so it is a direct indicator of registration success.

To spare computational time or to focus on a specific region, absolute difference can also be calculated just on a specific ROI (if used as an input). Usually the patient body is selected as ROI, without air and couch.

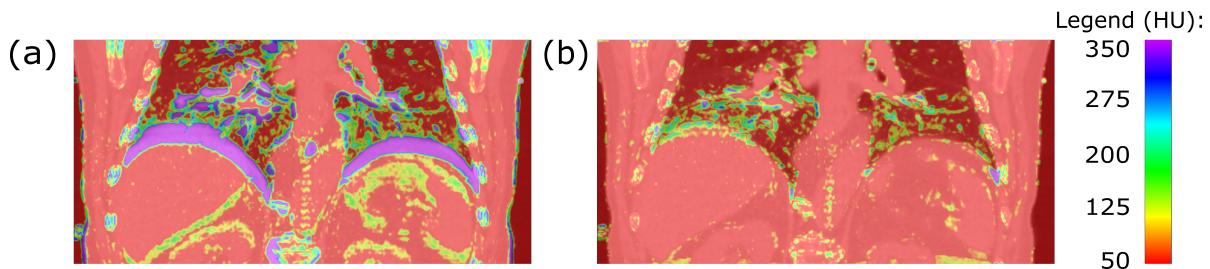


Figure 1.3: Absolute difference image before (a) and after registration (b). The mean absolute difference before registration (default absolute difference) is 62 HU and 31 HU after registration (true absolute difference).

Movie

Medical images are usually quite large - typical CT image consist of $512 \times 512 \times 100$ pixels, which makes inspecting image checks (inverse color, checkerboard, absolute difference) a time-consuming task. Movie feature allows for smoother display of different image slices. User selects which view he would like to inspect (axial, sagittal or coronal) and presses start. Selected views then start scrolling from one limit to the other. It allows user to focus on registration details, rather than scrolling through slices.

Movie and flicker (explained in the next section) do not offer any specific registration checks, but improve the process of DIRQA.

Flicker

While it is possible to display two images side by side in Slicer, it can sometimes be hard to see fine differences between the two images. Flicker alternates between reference and warped image on a single display at a 0.5 s rate.

Landmark distances

Distances between landmark before and after registration are often used to determine registration spatial accuracy [Castillo et al., 2009]. Landmarks can either be a specific feature in patient anatomy or an external marker. The position of landmarks in the warped image would ideally be at the same position as in reference image. The module measures the Euclidean norm between the landmark positions in reference and warped image.

User has to manually select landmarks in reference and moving image. For landmarks based on patient anatomy a selection from physician is required. Landmarks from moving image can then be automatically transformed to warped image with registration vector field or they can be manually selected in warped image itself. An example is shown in Fig. 1.4.

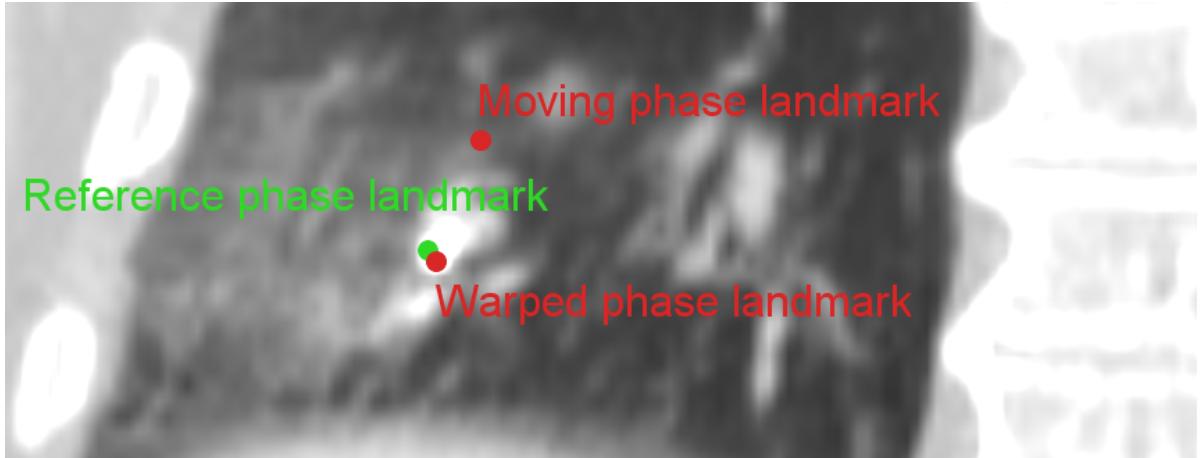


Figure 1.4: Display of landmarks in three phases - reference, moving and warped phase. Distance before registration (between reference and moving landmark) is 22 mm and after registration (between reference and warped landmark) is 2 mm.

Jacobian determinant

The Jacobian determinant or Jacobian (J) of the vector field u is calculated as:

$$J = \begin{vmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} & \frac{\partial u_z}{\partial y} & \frac{\partial u_z}{\partial z} \end{vmatrix} \quad (1.1)$$

The Jacobian is used to validate physical behavior of the registration [Leow et al., 2007]. The Jacobian of vector field should be positive, since negative Jacobian values correspond to organ folding, which is physically unrealistic for patient anatomy (organs can not be folded) [Rey et al., 2002, Chen et al., 2008]. Expansions and contractions around a point are indicated by Jacobian values of greater and less than 1, respectively.

DIRQA module calculates and displays the Jacobian of the vector field, as shown in Fig. 1.5. Statistical values of Jacobian are also calculated. Similar to absolute difference, Jacobian can be calculated inside a ROI rather than on a whole image.

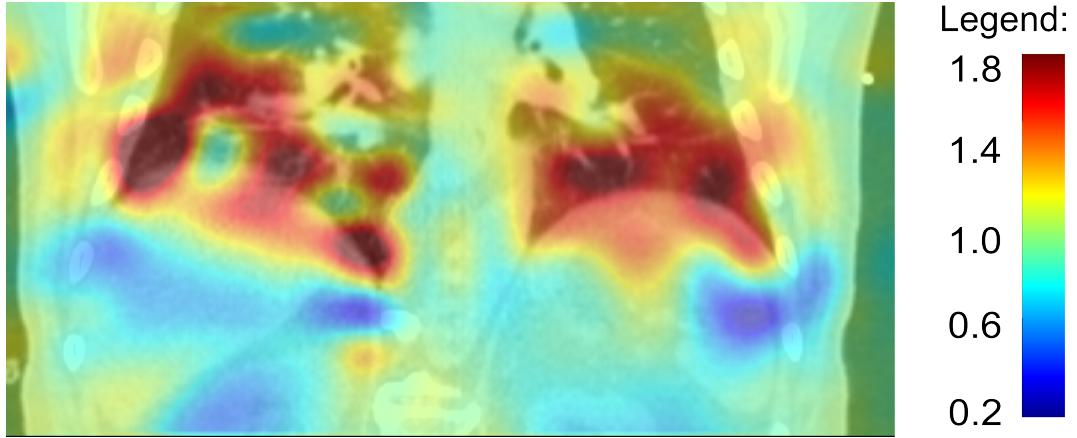


Figure 1.5: Image of Jacobian overlaid on a CT scan. The average value of the displayed Jacobian is 0.98 with 0.09 standard deviation.

Inverse Consistency Error

Inverse consistency error (ICE) is consistently used in literature as one of the main vector field checks [Christensen and Johnson, 2001, Bender and Tomé, 2009]. The principle is as follows. Suppose we have two vector fields: u_{AB} obtained from registration of image A to B and u_{BA} from registration of image B to A . The two registrations should be preformed separately. In an ideal scenario, u_{AB} would be a direct inverse of u_{BA} . However, DIR algorithms do not yield perfectly inverse consistent vector fields. Therefore, the differences between true and inverse vector fields have to be examined.

To check for ICE, an algorithm was created that first transforms a point x in image A , using u_{AB} . The newly obtained point x' is then transformed with the inverse vector field, u_{BA} which yields x'' . The ICE is defined as Euclidean norm between x and x'' :

$$ICE = |x - x''| = |x - u_{BA}(x')| = |x - u_{BA}(u_{AB}(x))| \quad (1.2)$$

Points x' and x'' can have an arbitrary position in space, while vector fields u_{AB} and u_{BA} are positioned on a grid. To apply transformation $u_{BA}(x')$, x' is interpolated on a u_{BA} grid. A tri-linear interpolation is used in this module.

As in Jacobian, ICE image is calculated and displayed, along with statistical values and ROI feature. An example is shown in Fig. 1.6.

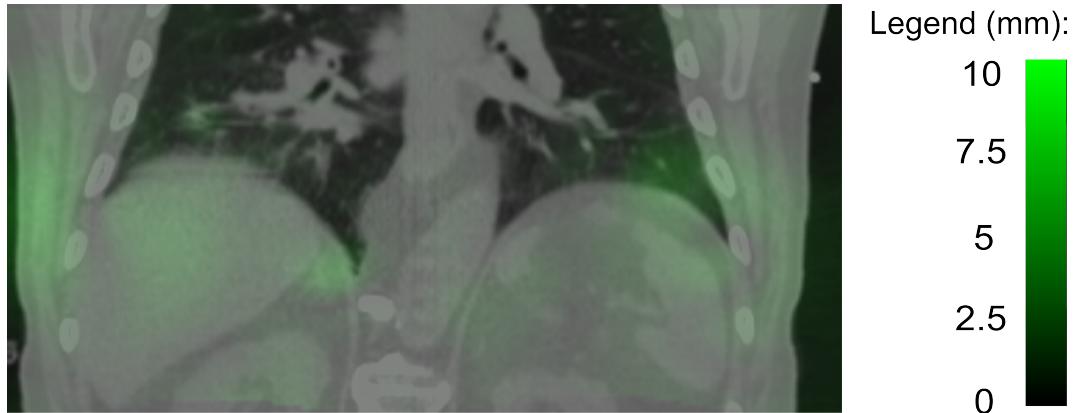


Figure 1.6: Image of the inverse consistency error (ICE) overlayed on a CT scan. The average value of displayed ICE is 1.0 mm with 0.8 mm standard deviation.

1.3 Verification

Several extensions for Slicer were created to perform DIR and DIRQA. All extensions have to undergo testing to prove their functionalities. Furthermore it is necessary to verify if extensions could be used in a typical clinical work flow. Especially DIRQA, which could facilitate the transition of DIR in everyday clinical work flow.

DIR and DIRQA were done on lung 4DCT patient data. As part of an animal study [Lehmann et al., 2015a] DIR and DIRQA were integrated in a simulated clinical work flow.

1.3.1 Registration of lung 4DCT patient data

Chapter **REF** and **REF** present studies on simulating active scanning carbon ion treatment (CiT) for non-small cell lung cancer patients. The effects of interplay can drastically change the dose distribution for CiT and it is necessary to quantify the effects of motion with DIR and transfer results into treatment planning software (TRIP4D in our case). An automatic procedure is required to perform DIR and DIRQA on a large number of patients. This was achieved with Slicer modules described in Section 1.2.

Materials and Methods

A time-resolved CT (4DCT), consisting of 10 motion phases (0 - 9) with approx. 1 mm pixel and 2 mm slice spacing was acquired with either Philips Brilliance BigBore 16-slice (Philips Healthcare, Eindhoven, Netherlands) or a Philips Gemini PET-CT 16-slice scanner. Phase 0 corresponds to the max end-inhale and phase 5 to max end-exhale breathing state. Phase 0 was chosen as a reference phase. 23 4DCTs of lung cancer patients were used in this study.

DIR was done for each patient between each phase and a reference phase and vice versa (true and inverse DIR). Each 4D-CT had 18 DIRs, leading to 414 DIRs in total.

B-Spline Plastimatch module in Slicer was used for DIR (see Section 1.2.2). DIRs were done in two stages with details given in Table 1.1.

Table 1.1: Parameters used for B-Spline Plastimatch DIR. A mean squared error metric was used. Details for each parameter can be found in [Plastimatch, 2016].

Parameter	Stage 1	Stage 2
Resolution	4,4,2	1,1,1
Grid size	50	15
Regularization lambda	0.005	0.005
Iterations	200	100

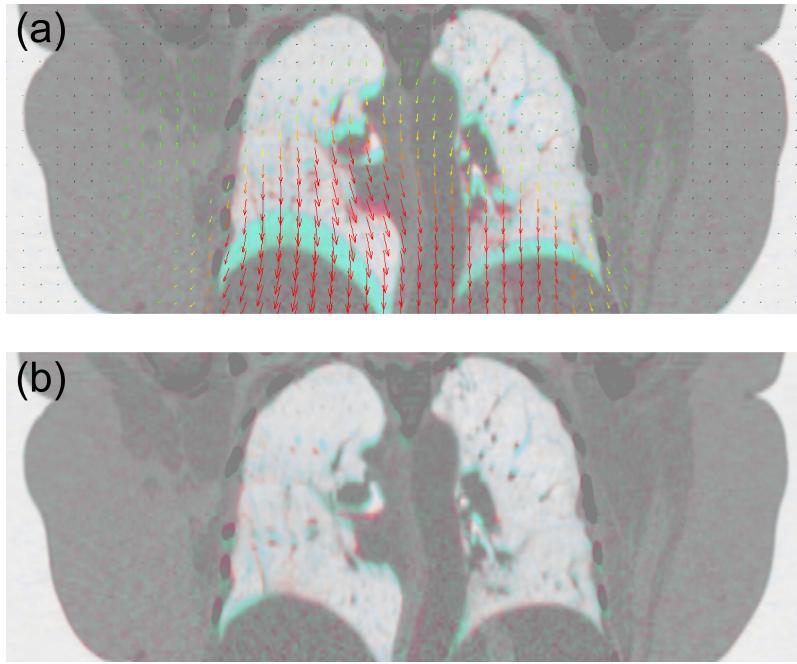


Figure 1.7: Inverse color overlay of two phases before (a) and after (b) DIR. Vector field is displayed on image (a) as arrows.

A ROI around patient body was manually created in Slicer by visually inspecting each reference phase of a 4D-CT. The ROI was then used for calculation of absolute difference, Jacobian and ICE.

Default, true and inverse absolute difference was calculated (see Section ??). In total 621 absolute differences were calculated and all images were down-sampled by a factor of 2 before calculation to save computer time. Similarly, 414 vector fields were also down-sampled by a factor of 2 before calculating Jacobian and ICE. Jacobian and ICE checks were calculated on all vector fields. Additionally, each vector field magnitudes were analyzed for mean, standard deviation (STD) and maximum (max) values. Paired t-tests were performed to compare statistical values of true and inverse vector field magnitudes. A p-value < 0.05 was considered significant.

For each patient it took around 20 min for all 18 DIR and around 30 min for complete DIRQA on the same 18 DIR. A Linux computer with 8 central process units cores and 32 GB RAM was used for DIR and DIRQA.

Results

An example of DIR is displayed in Fig. 1.7. Vector fields statistical analysis is shown in Table 1.2. There was no statistical difference between true and inverse vector field magnitudes. The biggest contribution to vector field magnitude was from superior-inferior direction (around 50%), followed by anterior-posterior direction (around 30%) by left-right direction (around 20%).

Table 1.2: Data of vector magnitudes. Values are presented as mean (range).

	True vector field	Inverse vector field
Mean	0.38 (0.01 - 1.28)	0.38 (0.01 - 1.3)
STD	0.95 (0.04 - 3.17)	0.98 (0.04 - 3.55)
Max	9.67 (0.61 - 28.56)	10.17 (0.56 - 37.11)

True and inverse absolute difference dependence on default absolute difference is shown in Fig. 1.8. It also shows default absolute difference distribution across 9 phases.

Distribution of true and inverse Jacobian and ICE data are displayed in Fig. 1.9 and 1.10. Dependence of maximum true Jacobian on minimum inverse Jacobian and vice versa was tested and results are displayed in Fig. 1.11. Parameters from linear fit in Fig. 1.11 were used to calculate so-called scaled Jacobian for points that deviated from linear fit. Each voxel, x , in scaled Jacobian, S was calculated from true and inverse Jacobian, J_{True} and $J_{Jacobian}$, as:

$$S(x) = 1.7 \ln(J_{True}(x)) + \ln(J_{Inverse}(x)) \quad (1.3)$$

Example of scaled Jacobian is shown in Fig. 1.11.

True and inverse maximum and minimum Jacobian and maximum ICE values were plotted against maximum vector magnitudes and fitted with linear function. Results are shown in Fig. 1.12.

All linear fits used in Fig. 1.8, 1.11 and 1.12 were statistically significant ($p < 0.05$).

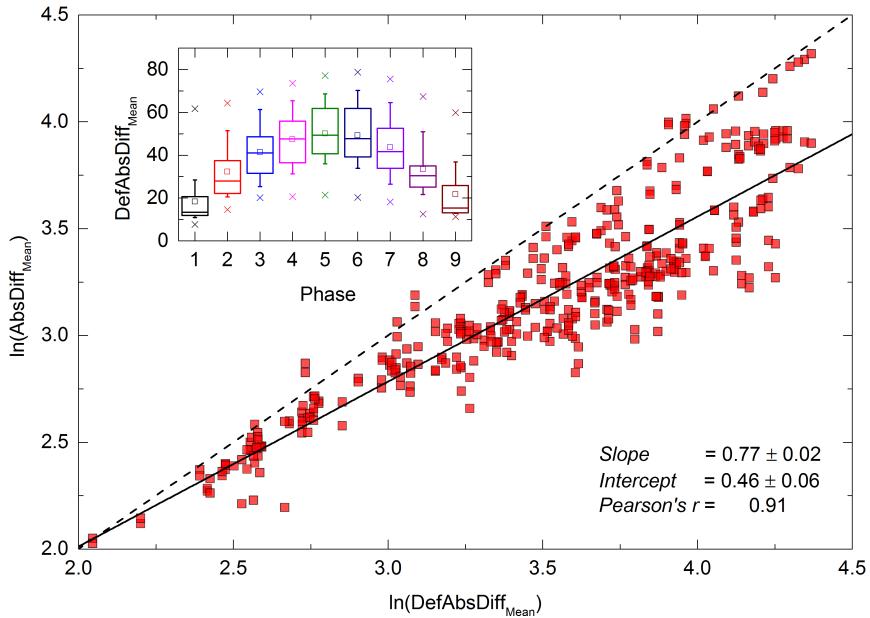


Figure 1.8: Natural logarithm of true and inverse absolute difference (AbsDiff) plotted against default absolute difference (DefAbsDiff). Solid line shows linear fit, with parameters written in corner. Dashed line shows $y(x) = x$ plot. Inset is a box plots of DefAbsDiff mean distribution across nine 4DCT phases. Boxes represent 25-75%, whiskers 10-90% of data, median is shown with solid line, mean with square and outliers with crosses.

Discussion

A DIR was successfully made for 23 4D-CTs, producing 414 vector fields with true and inverse registration. All 414 DIR underwent a DIRQA consisting of vector field magnitudes, absolute difference, Jacobian and ICE.

Vector field magnitudes confirm previously published data that the biggest motion for lungs is in superior-inferior direction [Seppenwoolde et al., 2002, Britton et al., 2007, Liu et al., 2007]. The mean vector field magnitude is small (in submillimeter range), because the ROI included the whole patient body, not just the lungs where most of the motion occurs. Vectors and inverse vectors are similar, which was expected.

There was a good correlation (Pearson's $r = 0.91$) between absolute difference before and after DIR. The greater the difference between reference and moving image, the better DIR can lessen it. Few cases have slightly higher absolute difference after DIR, which should not occur, since DIR was minimizing absolute difference metric. The reason lies in construction of warped images. Vector field is used to transform moving image to warped image. If the vector field is large on the edges, an empty space will be left there in warped image. Absolute difference will therefore seem larger after DIR. A future modification of absolute difference module should handle empty spaces at the edges.

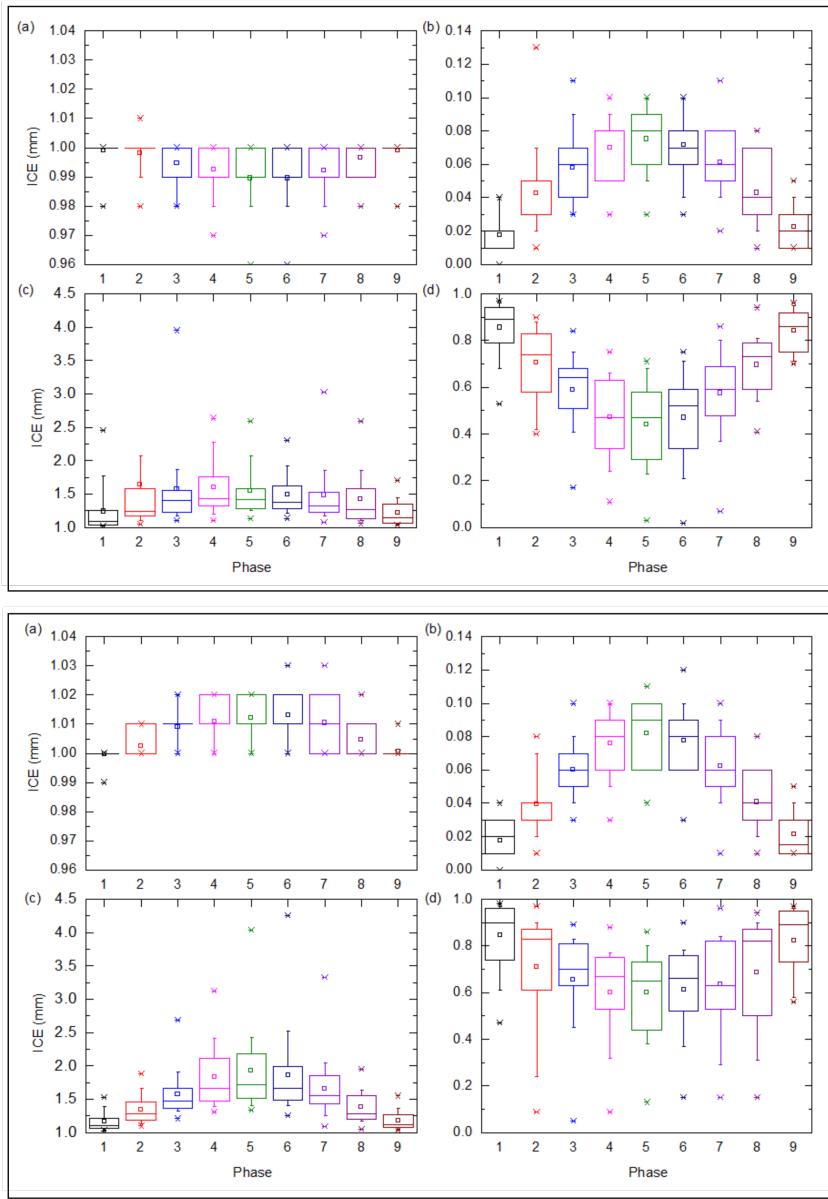


Figure 1.9: Data for true (top) and inverse (bottom) Jacobian for 9 4DCT phases (reference phase 0 is excluded) for 23 lung cancer patients. Mean, STD, maximum and minimum are represented as (a), (b), (c) and (d), respectively. Boxes represent 25-75%, whiskers 10-90% of data, median is shown with solid line, mean with square and outliers with crosses.

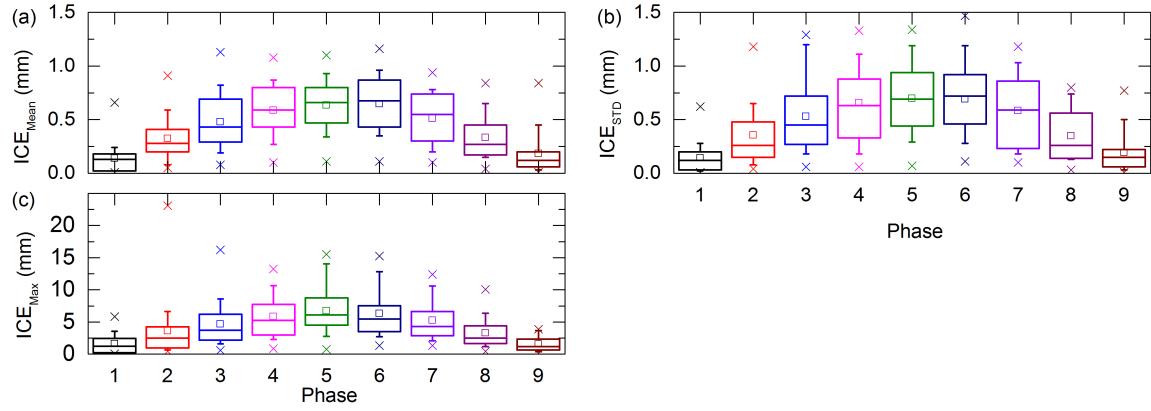


Figure 1.10: Data for ICE for 9 4DCT phases (reference phase 0 is excluded) for 23 lung cancer patients. Mean, STD, maximum are represented as (a), (b) and (c), respectively. ICE Minimum is 0 throughout all phases and patients. Boxes represent 25-75%, whiskers 10-90% of data, median is shown with solid line, mean with square and outliers with crosses.

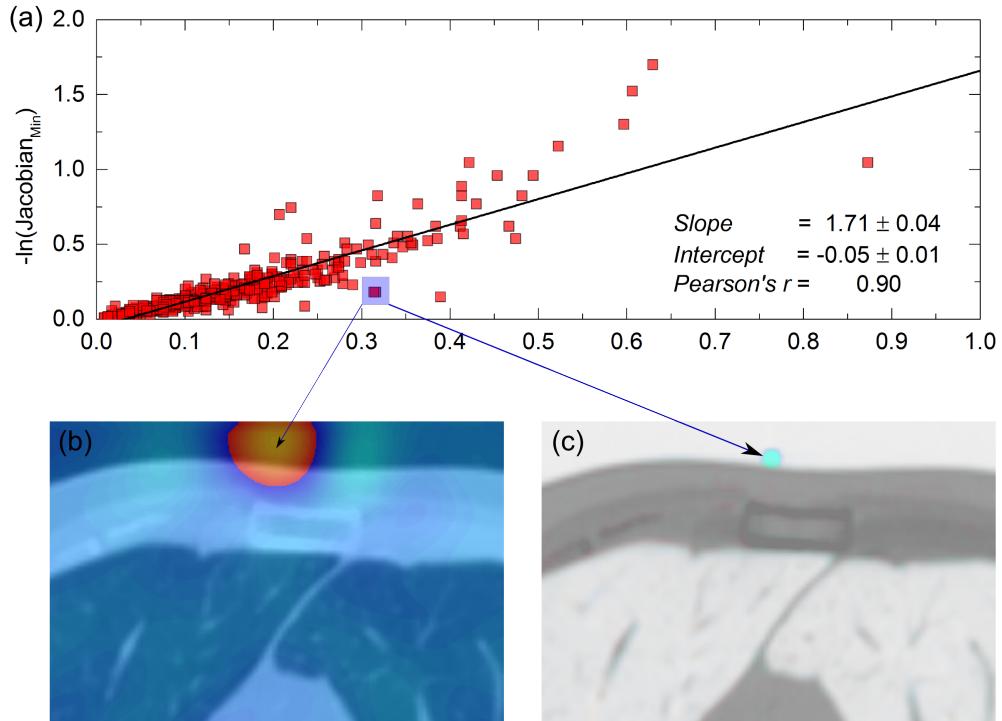


Figure 1.11: (a) Plot of negative natural logarithm of minimum inverse (true) Jacobian versus natural logarithm of maximum true (inverse) Jacobian. Linear fit is displayed with solid line and its parameters are given in the corner. A deviation from linear fit (highlighted with blue square) was used as an example of scaled Jacobian (see text), shown in (b). (c) shows the CT the phases highlighted in (a) in inverse color, where an artifact is clearly seen in one phase and not the other.

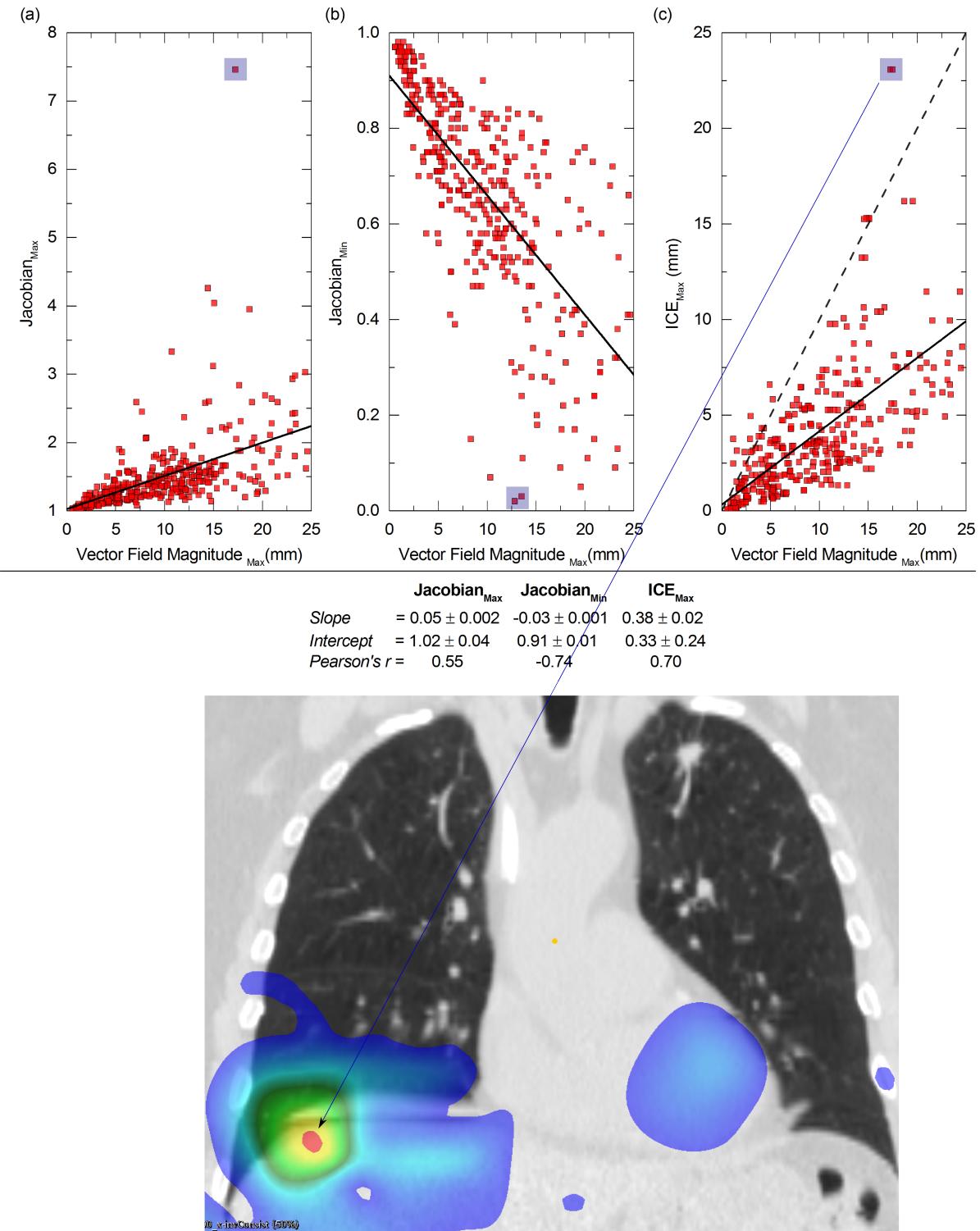


Figure 1.12: Values of maximum Jacobian (a), minimum Jacobian (b) and maximum ICE (c) plotted against maximum vector magnitudes. Linear fit is displayed with solid line and parameters are written below the plots. Dashed line in (c) shows $y(x) = x$ plot. ICE on image (d) is taken from patient highlighted with blue squares in (a)-(c). ICE is displayed using color table as displayed in legend.

Due to small mean vector field magnitudes, average values for true and inverse Jacobian were 10.05, which indicates that most of the patient body does not change during the 4DCT scan. This is also a direct consequence of a small vector field magnitude. However, patient expansions and contractions can be seen on maximum and minimum Jacobian, with average values around 1.50 and 0.65 respectively. If a part of a patient body contracts from reference to moving image, then it expands in inverse direction and vice versa. The correlation was confirmed in Fig. 1.11a, with a high Pearson's r (0.90). Furthermore outliers from linear fit spot inconsistencies in DIR as shown in Fig. 1.11b and c, where a small artifact was found in one patient phase solely from the deviation from linear fit in Fig. 1.11. A scaled Jacobian could be used as a DIRQA check.

Mean and STD ICE are in submillimeter range, due to the correlation between vector field and ICE (see Eq. 1.2). The maximum ICE values can be large, with up to 2.3 cm, but still smaller as average maximum vector values.

Large vector field magnitudes will produce more errors in DIR as shown in Fig 1.12. Linear fit was used to estimate increase (decrease) of Jacobian and ICE. As a rough DIRQA check, ICE should always be smaller as maximum vector field values. To confirm this, all cases above dashed line in Fig 1.12c were investigated and were found to have areas of poor DIR. An extreme case (highlighted in Fig. 1.12a-c) had a large image artifact present in one phase (shown in Fig. 1.12d) leading to inconsistencies in DIR.

1.3.2 Registration of pig heart 4DCT data

Atrial fibrillation is an unorganized atrial activity, causing a quivering motion of the heart. Heart is therefore not able to sustain a healthy pumping rhythm. Atrial fibrillation is not a deadly disease, however it worsens the patients quality of life and increases the risk of a stroke [Benjamin et al., 1998]. A common method for treating atrial fibrillation is catheter ablation [January et al., 2014], where the success rate is still limited and can even lead to major complications or even death of a patient [Cappato et al., 2005, R et al., 2010].

As an alternative treatment, a carbon-ion therapy was proposed [Bert et al., 2012] and later feasibility on a beating heart was shown experimentally [Lehmann et al., 2015b]. In 2014 a pilot experiment was performed at GSI using large animal model (pigs) and scanned carbon-ion to verify the treatment *in vivo* [Lehmann et al., 2015a].

To estimate and compensate motion of the heart during irradiation DIR of 4DCT data was required. Furthermore, because of the actual irradiation of live pigs a DIRQA had to be made, to ensure the validity of DIR. Description of the procedure will be given here, alongside with the results.

Materials and Methods

Pig irradiation experiment

DIR and DIRQA procedures will be given here, while a detailed description of the whole pig irradiation experiment can be found elsewhere is given here [Lehmann et al., 2015a]. A cardiac gated contrast enhanced CT scans (4DCT-cardiac) were made on 15 pigs with a multidetector 64 row Siemens Somatom Definition Flash scanner (Siemens Healthcare, Forchheim, Germany) with 1 mm voxel and 1 mm slice spacing. There was no breathing motion present, since a breath-hold technique was used. Cardiac motion was divided into 10 sequential phases (0-10%) and a field of view of 400 mm for skin-to-skin images was used. 8 pigs had a pacemaker implemented, because the irradiation was planned to disturb the heart rhythm and a pacemaker should compensate for that. Pigs are therefore divided into two groups, with pacemaker (PM) and without one (noPM).

After CT acquisition, DIR on 4DCT-cardiac was made using B-Spline Plastimatch module and patient hierarchy in Slicer (see Section 1.2.2 and 1.5). Details on parameters used for DIR can be found in Table 1.3. Phase 0% was chosen as a reference phase. All other phases were registered to reference phase with inverse registration as well. A checklist was made to follow DIR and DIRQA for quality assurance. An example of a used checklist is shown in Fig. 1.13a.

Based on lung patient DIR and because of the time pressure, DIRQA was made only on DIR from one phase, phase 50%. DIRQA consisted of true absolute difference¹, true Jacobian and ICE. DIRQA results were stored in a text file (example shown in Fig. 1.13b) and users checked if the values did not exceed expected ones: Absolute difference mean should be smaller than 1; Jacobian mean should be 1 with not too high extreme values; ICE mean should be smaller than 2 mm with not too high extreme values. ROI was manually created to encompass pig body and then used in all DIRQA checks.

After successful DIR and DIRQA vector fields were used in treatment planning and the resulting plans were used in pig irradiation experiment.

For each pig it took around 20 min for all 18 DIR and additional 30 min for DIRQA on 18 DIR. An interactive Linux computer with 8 clusters and 32 GB RAM was used for DIR and DIRQA.

Table 1.3: Parameters used for Plastimatch registration. A mean squared error metric was used. Details for each parameter can be found here [Plastimatch, 2016].

Parameter	Stage 1	Stage 2
Resolution	4,4,2	2,2,1
Grid size	50	15
Regularization lambda	0.005	0.005
Iterations	200	100

Post-experiment analysis

After pig irradiation experiment, a more detailed DIRQA was made, with all phases included in DIRQA. In addition to original checks explained in previous section, absolute difference checks were made on inverse warped image, inverse Jacobian and vector magnitudes were analyzed.

¹ Relative difference between moving and warped image absolute difference was displayed in DIRQA file rather than just absolute values.

(a)

Checkliste Bestrahlungsplanung

DOSE: 406 TARGET: PV SCHWEIN: Uniform

Scripts path: AlXd/user/motion/Beamtime/GSI1407/Simulations/SCRIPTS

Heartbeats during CT: N:45 bpm Finalmotion: sin_3mm_1200ms0

Task	Script	Done?	Name	Date
Create Header		✓	AE	17.07.
Copy CT data to PatientData2		✓	AE	17.07.
Sort CTs (contrast/native)		✓	AE	"
Sort motion phased		✓	AE	"
Check # of files in each phase	CTX.sh	✓	AE	"
Check contrast and native slices	"	✓	AE	"
Create CTX	"	✓	AE	"
Create MHA	"	✓	AE	"
Check BB coordinates and write Header		✓	AE	"
Copy Contours		✓	AE	18.07.
Create VDX		✓	AE	"
Change 0 to 1 for Target (VDX)		✓	AE	"
Check voi names in VDX		✓	AE	"
FalseColorCheck contrast<->native (00)		✓	AE	18.07.
Contours native ok?				
Registration contrast -> 4DCT		✓	AE	17.07.
Registration ok?		✓	AE	17.07.

(b)

DIRQA for: Registration Node Contrast 4D

```

InvVector
x: 4.01
y: 3.59
z: 4.1
Vector
x: 4.53
y: 4.11
z: 4.49
AbsoluteDifference
Mean: 0.19
STD: 0.41
Max: 1446.0
Min: 0.0
Jacobian
Mean: 1.0
STD: 0.05
Max: 1.48
Min: 0.65
InverseConsistency
Mean: 0.11
STD: 0.17
Max: 1.97
Min: 0.0

```

Figure 1.13: (a) Part of the checklist for quality assurance during pig irradiation. DIR and DIRQA part is highlighted in red and consisted of two steps. First DIR was made on 4DCT-cardiac and afterwards DIRQA was made on DIR from phase 50%. End result was presented as text shown in (b). Maximum values for absolute difference were disregarded, because there can be large deviations in individual voxels.

Results

All DIRs were successful and during experiment all DIRQA checks were positive. An example of DIR is shown in Fig. 1.14. Statistical analysis on vector field magnitudes is shown in Table 1.4. No statistical difference was observed between true and inverse vector fields. However, significant difference was observed between vector field magnitudes of PM and noPM. Contributions to vector field magnitudes from three axis were equal.

Table 1.4: Data for vector magnitudes. Values are presented as mean (range).

	PM		noPM	
	True vector field	Inverse vector field	True vector field	Inverse vector field
Mean	0.08 (0.03 - 0.16)	0.08 (0.03 - 0.14)	0.07 (0.0 - 0.18)	0.06 (0.0 - 0.17)
STD	0.4 (0.09 - 0.78)	0.36 (0.08 - 0.68)	0.3 (0.05 - 0.77)	0.28 (0.04 - 0.71)
Max	8.24 (1.6 - 17.33)	7.98 (0.7 - 17.76)	5.9 (0.97 - 15.91)	5.38 (1.08 - 12.42)

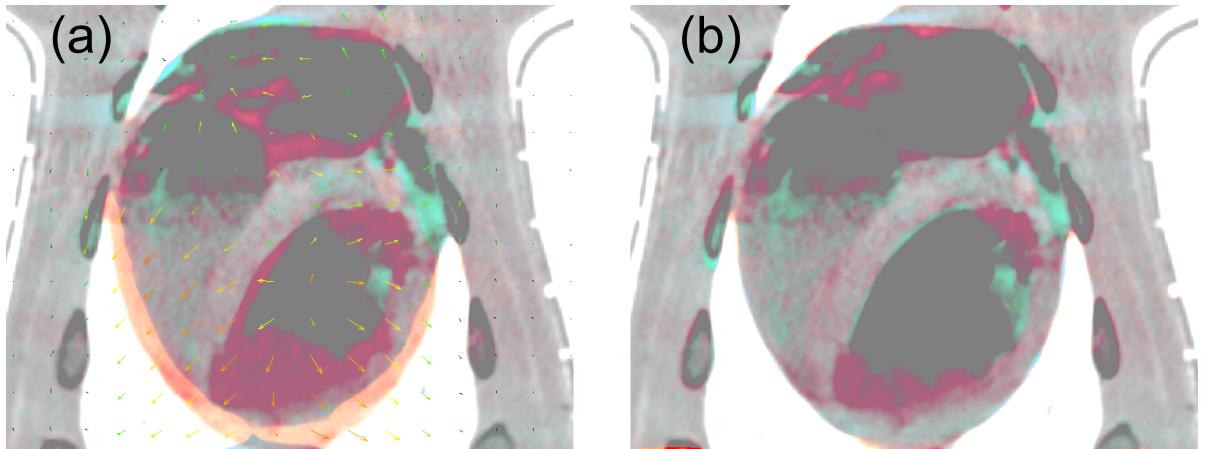


Figure 1.14: Inverse color overlay of two phases before (a) and after (b) DIR. Vector field is displayed on image (a) as arrows.

Dependence of true and inverse absolute difference on default absolute difference with a linear fit is shown in Fig. 1.15. Default absolute difference distribution across 9 phases can be seen in inset in Fig. 1.15.

Distribution of Jacobian and ICE results are shown in Fig. 1.16 and 1.17. Maximum true (inverse) Jacobian was plotted against minimum inverse (true) Jacobian and results were fitted with a linear function, see Fig 1.18. Maximum values of true and inverse Jacobian and maximum ICE were tested against maximum vector magnitudes and fitted with linear function. Results are plotted in Fig. 1.19.

All linear fits in Fig. 1.15, 1.18 and 1.19 were statistically significant ($p < 0.05$).

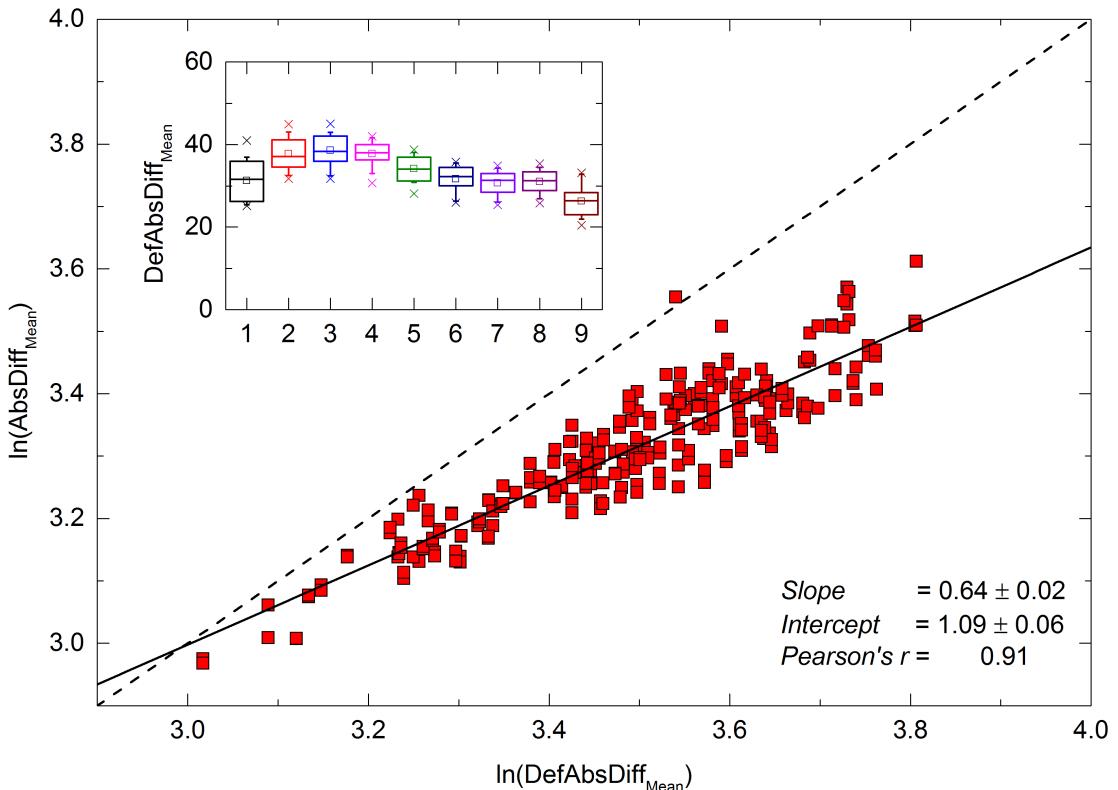


Figure 1.15: Natural logarithm of true and inverse absolute difference (AbsDiff) plotted against default absolute difference (DefAbsDiff). Solid line shows linear fit, with parameters written in the corner. Dashed line shows $y(x) = x$ plot. Inset shows box plots of mean distribution of DefAbsDiff across nine 4DCT-cardiac phases. Boxes represent 25-75%, whiskers 10-90% of data, median is shown with solid line, mean with square and outliers with crosses.

Discussion

Mean vector field magnitudes were small (approx. 0.1 mm), since pigs were in a breath-hold position and the only motion was the beating of the heart. Despite the small mean vector magnitude it was still enough to observe statistical difference between the PM and noPM. Consequently, the difference between the two groups is consistent throughout the DIRQA.

The DIR did well in terms of lowering the absolute difference metric. There was a strong correlation between default versus true and inverse absolute difference. The shape of default absolute difference distribution persist then in other DIRQA check as well.

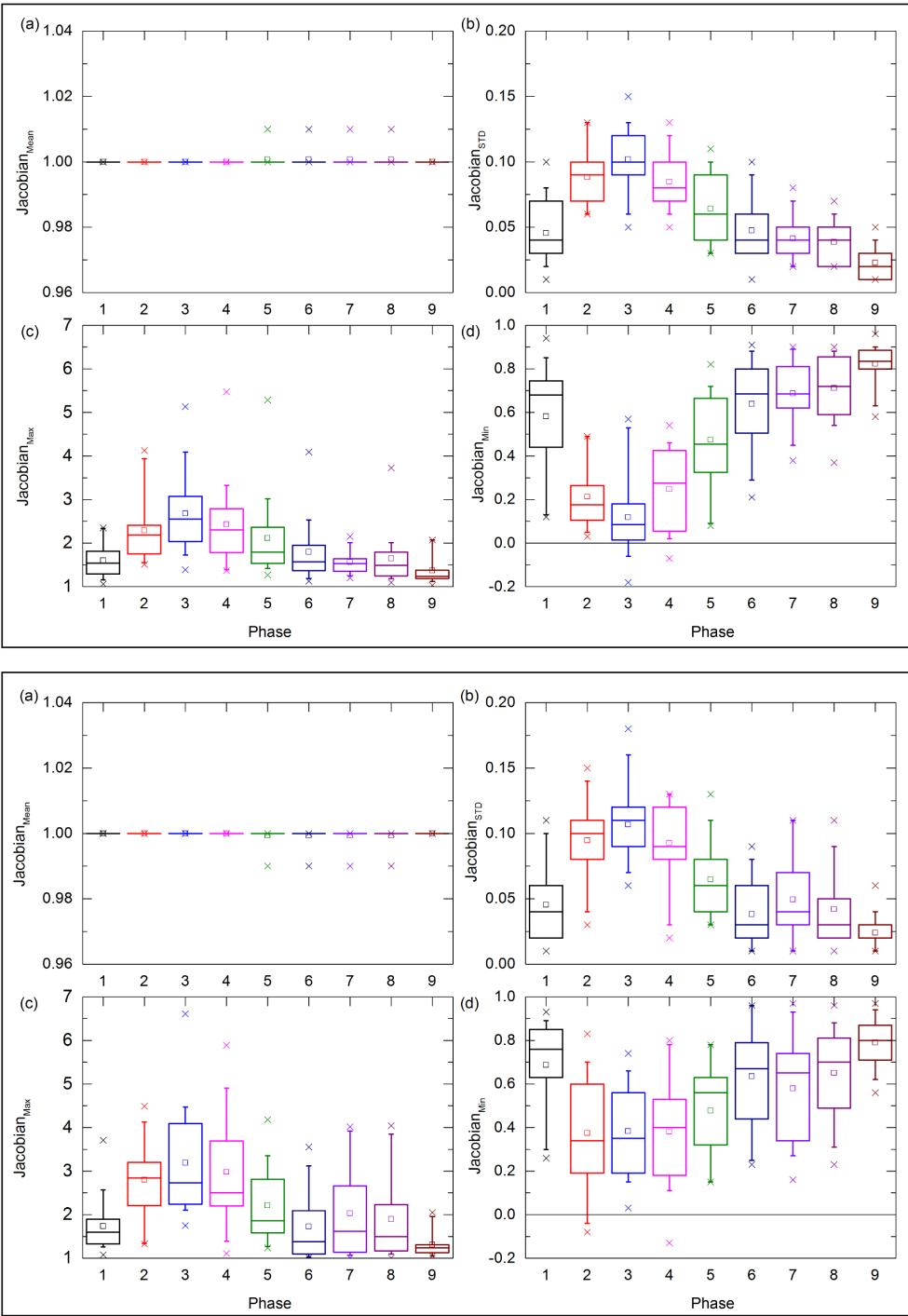


Figure 1.16: Statistical data for true (top) and inverse (bottom) Jacobian for 9 4DCT-cardiac phases (reference phase 0 is excluded) for 15 pigs. Mean, STD, maximum and minimum are represented as (a), (b), (c) and (d), respectively. Boxes represent 25-75%, whiskers 10-90% of data, median is shown with solid line, mean with square and outliers with crosses.

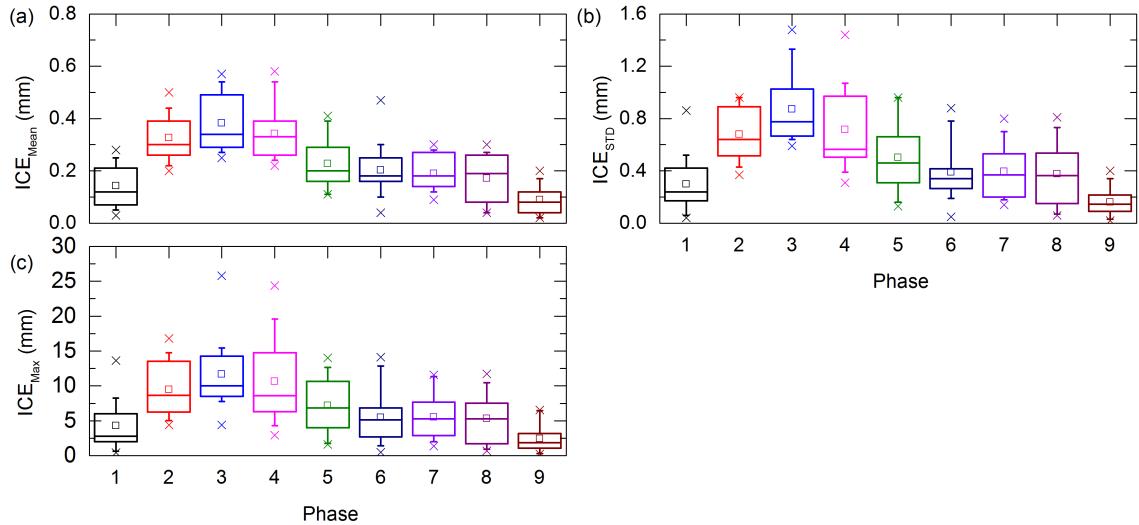


Figure 1.17: Statistical data for ICE for 9 4DCT phases (reference phase 0 is excluded) for 15 pigs. Mean, STD, maximum are represented as (a), (b) and (c), respectively. ICE Minimum is 0 throughout all phases and pigs. Boxes represent 25-75%, whiskers 10-90% of data, median is shown with solid line, mean with square and outliers with crosses.

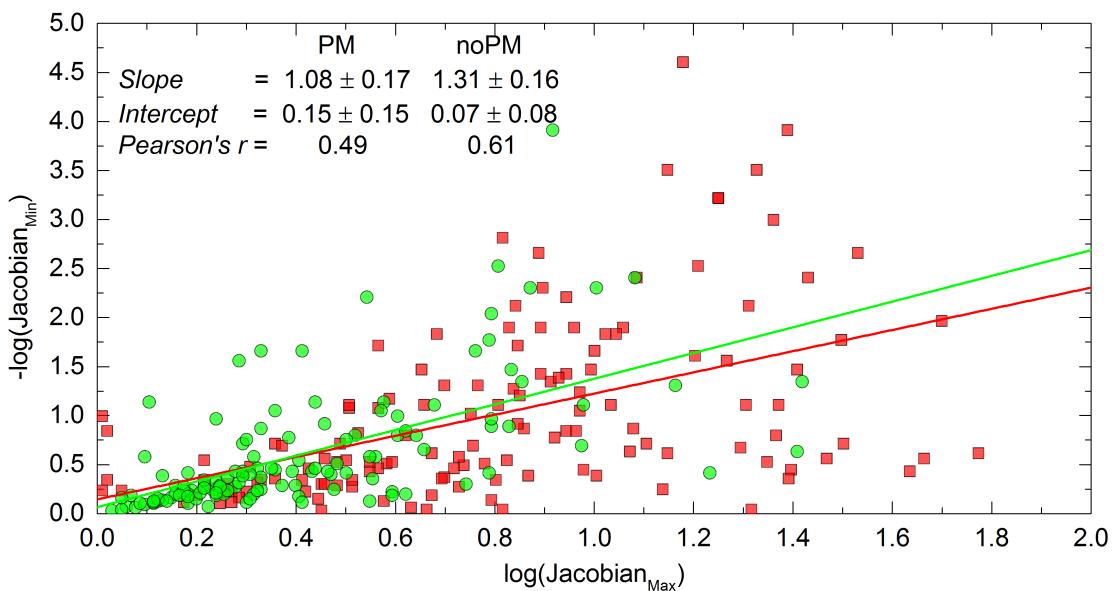


Figure 1.18: (a) Plot of negative natural logarithm of minimum Jacobian versus natural logarithm of maximum Jacobian. Linear fit is displayed with solid line and it's parameters are given in the corner. PM and noPM are shown with red squares and green circles, respectively.

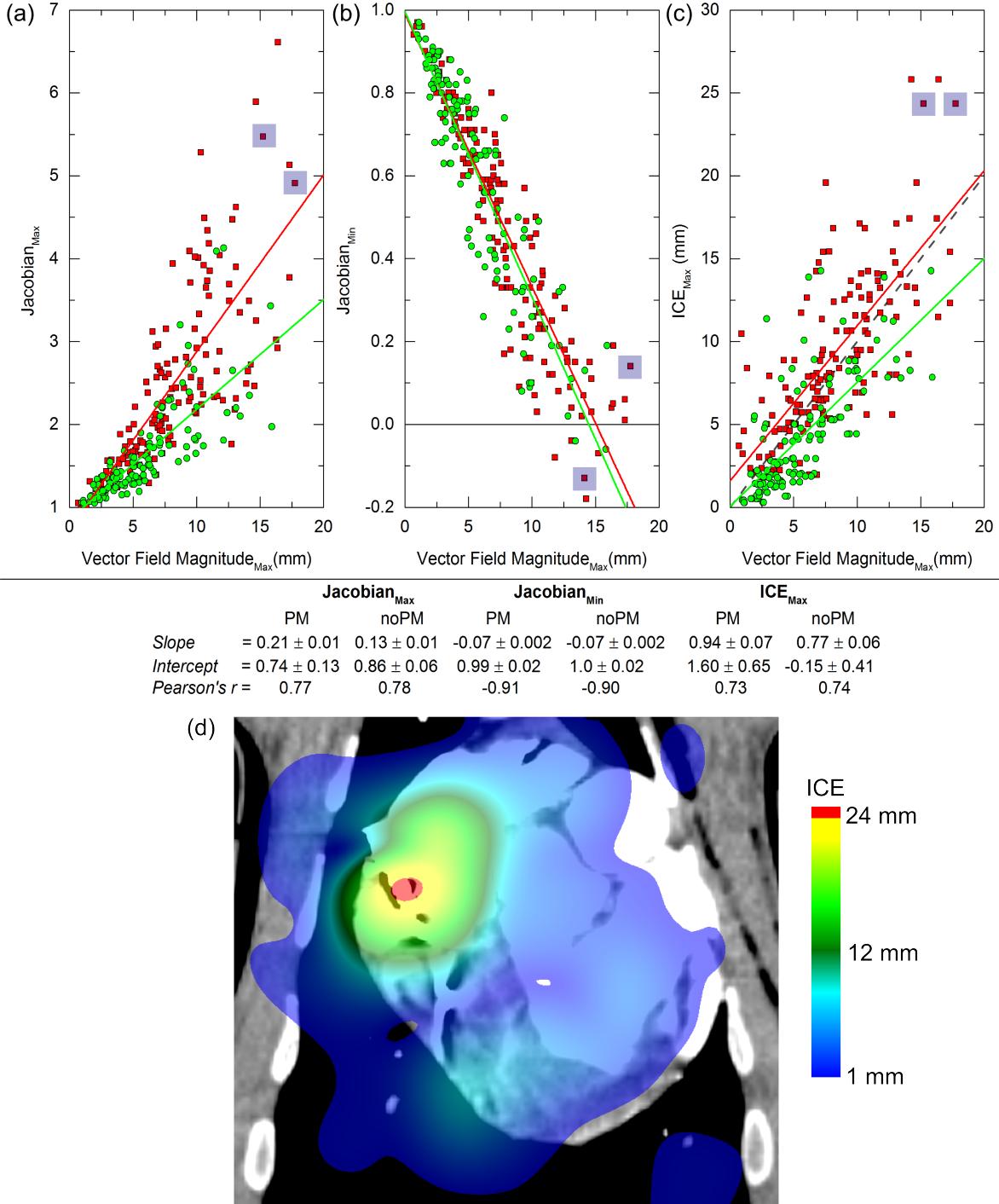


Figure 1.19: Values of maximum Jacobian (a), minimum Jacobian (b) and maximum ICE (c) plotted against maximum vector magnitudes. Linear fit is displayed with solid line and its parameters are written below the plots. Dashed line in (c) shows $y(x) = x$ plot. Values from pig on image (d) are highlighted with blue squares in (a)-(c). PM and noPM are shown with red squares and green circles, respectively. ICE is displayed on (d) using color table as displayed in legend.

A good result in absolute difference does not necessarily mean a good DIR, as can be seen from Jacobian and ICE checks. The mean Jacobian and ICE were 1 and 0, respectively, since the vector

fields were small on average. However there were large deviations present in Jacobian and ICE. Most notably, there were a few cases of negative minimum Jacobian which would suggest organ folding. Since organ folding did not occur during a heart beat, negative minimum Jacobian points to inconsistencies in DIR. The inconsistencies in DIR can be also seen on Fig 1.18, with a poor correlation between minimum inverse and maximum true Jacobian. The correlation is better for noPM, however there are still cases that deviate from linear fit.

The large deviations in Jacobian and ICE can in part be explained with large maximum vector field values, as shown in Fig. 1.19. All linear fits have good correlation, with no difference between PM and noPM in the quality of the fits. The actual linear fit parameters, however, further show the inconsistencies in DIR. The clearest example of inconsistencies in DIR is with linear fit from maximum ICE PM, which lies above $y(x) = x$ function. This basically means that there were points further away from starting point after true and inverse transformation, then just after true transformation. The linear fit for noPM maximum ICE showed better results in this terms, since it lied below $y(x) = x$ function.

It was also displayed that DIRQA should be made on all DIR not just on one phase, since DIRQA from one DIR does not guaranty the quality of the other DIR from 4DCT. Each DIR is performed individually and should be treated as such. However, deviations in DIR from single 4DCT could be used to spot inconsistencies.

1.4 Summary and Discussion

Tools to preform DIR and DIRQA were presented in this chapter. Different modules were written for an open-source software Slicer, that can handle large DIR problems, such as registering whole 4DCT. In addition to DIR the modules can also make DIRQA on the same set as DIR was made with several different checks. The main objective of this work was to provide systematic approach to DIR and to give parameters on DIRQA that can estimate the quality of DIR. A first analysis of DIRQA checks was done on a large DIR database - 684 DIR were checked in total.

Most of the work was based on Slicer, which is well-established software in medical research. To date, there are more than 500 publications that have used Slicer in their research [Slicer, 2016b], with topic ranging from teaching [Pujol et al., 2016], disease staging [Liu et al., 2016, Liu et al., 2015], motion tracking [Behringer et al., 2015] to image reconstruction [Meyer et al., 2015], image registration [Li et al., 2015, Fedorov et al., 2015, ?] and others. Slicer offers a lot of different functionalities and is especially good for research, since it can be modified to specific needs. However, it is important to stress that Slicer is not a medical application and can be used only for research. Additionally, Slicer can sometimes be unstable with unexpected crashes. It is constantly under development and more and more errors are fix with each new release. New releases also bring new functionalities, but there can be problems with backtrack compatibility Even though there are some disadvantages to using Slicer,

it's advantages definetelly outweigh them and make Slicer a useful tool, as was showed in this chapter.

Results shown in this chapter were obtained with DIR made by B-Spline algorithm. Several other algorithms exist, demons most commonly used alongside B-Spline [Thirion, 1998]. Varadhan et al. made comparison between B-Spline and demons DIR for lung case [Varadhan et al., 2013] and showed that B-Spline is superior to demons, especially if there is difference in contrast between images. They used mutual information metric, to account for differences in contrast. Images used in this chapter were either all without (lung case) or all with (pig case) contrast agent, therefore there was no difference in contrast between images. Mutual information metric was tested as well for DIR, however the results were much worse than mean square error metric.

The evaluation of DIR was made by DIRQA module. The main advantage of the DIRQA module is that all different techniques are gathered in a single place and can be used on a specific case. The ease of use is also essential, for DIRQA to find it's way into clinical work flow. A test of using DIRQA in potential clinical work flow was done at GSI during pig experiment, where different users had to use both DIR and DIRQA modules. The experiment was also under time pressure, since there is a scheduled beam time and it cannot be shifted. There were already prepositions for frameworks for DIRQA in clinical work flow [Varadhan et al., 2013], however none were tested in an actual clinical environment

The techniques used in DIRQA can be divided into visible (inverse color, checkerboard) and quantitative (absolute difference, Jacobian and ICE). While the quantitative can be used to pinpoint errors in DIR, the visible can be used to actually see the error and how does it affect the end result, as shown in Fig. 1.11c. All three quantitative checks have been used in literature as a possible DIRQA [Varadhan et al., 2013, Leow et al., 2007, Christensen and Johnson, 2001, Bender and Tomé, 2009]. They all share the same flaw, however, that they are a necessary condition for a successful DIR but not sufficient. One common DIRQA check in literature that our module is currently missing, is comparison of anatomical correspondence - comparison between reference, moving and warped contours. Ideally warped and reference contour should be the same. Two metrics are usually used in contour comparison - dice similarity coefficient [Varadhan et al., 2013] and Hausdorff distance [Huttenlocher et al., 1993]. Slicer already has functionalities for both contour comparison checks, so they could be used. The biggest disadvantage of anatomical correspondence check is that contour delineation is required in both, reference and moving phase, which is scarcely done by physicians, since it takes too much time. Lack of contour in both, reference and moving phase, was also the reason the anatomical correspondence check was not used.

Studies on DIRQA so far have focused on a small number of DIR cases, weather it is phantom [Mutic et al., 2001, Moore et al., 2004] or patient studies [Wu et al., 2008, Varadhan et al., 2013]. With small number of DIRs it is possible to thoroughly examine each

DIR and hence now what to expect from the DIRQA. In this chapter a different approach was used. Rather than examining each DIR individually, a large dataset was analyzed and common traits for DIR were found. Due to differences in anatomical sites, DIRQA parameters have to be found for each anatomical site individually, since they can deviate significantly, as seen by two different cases presented here. However there may be parameters that persist throughout different anatomical sites, scaled Jacobian in particular.

DIR of lung 4DCT can be considered relatively easy, since the changes between phases are small and the contrast between lungs and other tissue is high. This is supported by Fig. 1.11 and 1.12. The correlation between expansion of tissue in reference phase should directly correlate to tissue contraction in moving phase. Even though Fig. 1.11a shows only plot between maximum and minimum Jacobian in two phases, the correlation is still apparent (Pearson's $r = 0.9$). With calculation of difference between maximum and minimum Jacobian in each individual voxel, it is possible to pinpoint the position of DIR inconsistencies, as shown in Fig. 1.11. A small artifact was found in one phase and it was responsible for deviation from linear fit. Deviation from linear fit in Fig. 1.12a-c, was also used to spot DIR inconsistencies and a lung image artifact was found in a different patient.

If the DIR of lung 4DCT was considered relatively easy, opposite holds true for DIR of 4DCT-cardiac. The motion of the heart during a heartbeat is complex, with muscles stretching [?] and contracting in different directions. Furthermore, the volume of blood shifts from one ventricle to the other. In the case of 4DCT-cardiac blood carried a contrast agent, therefore different phases in 4DCT-cardiac had different distribution of HU in different parts of the heart. Additionally, it was well established that pacemakers cause several complications in a CT scan [?]. This was confirmed by the differences observed between PM and noPM. Clearest example is the PM linear fit of maximum ICE in Fig. 1.19, which is above $y(x) = x$ curve. The complexity of DIR is further seen in the negative values of Jacobian, which should not occur. The DIR of 4DCT-cardiac is currently under careful investigation and several different solutions, such as artifact removal and different registration order are being tested.

In the future DIRQA module should undergo further testing. In addition to checking DIRQA on different anatomical sites and between different modalities, it should be investigated how good is DIRQA at spotting inconsistencies in DIR, i.e. what is the number of false negatives. Furthermore, with more data analyzed, the parameters in DIRQA checks should get more precise, so outliers could be easier spotted.

1.5 Appendix

Patient hierarchy

Patient hierarchy follows a subject hierarchy principle in Slicer. It was designed for a clear overview of the registration process, DIRQA and all resulting files. Another reason is to track DIR and DIRQA in case if they are interrupted by Slicer crash. DIR and DIRQA files can be quite large and can cause Slicer to run out of memory. With patient hierarchy Slicer is able to continue work from where it was interrupted rather than starting anew.

There are several levels in patient hierarchy. Each level also has different attributes, where details regarding each level can be written.

- Level 1: **Patient name** - separates different patients.
- Level 2: **Registration node** - separates between different registrations, e.g. between different imaging modalities or between 4DCT phases.

Attributes:

- The file directory of images, vector fields and registration quality files.
- Number of phases to be registered.
- Reference phase

- Level 3: **Registration set** - specific registration phase. Registration is done between all phases and the reference one. There have to be at least two phases
- Level 4: **Node** - can be either an image, a vector field, an inverse vector field or any of DIRQA nodes (see Section 1.2.3).

Attributes:

- Exact file paths for specific node.
- Statistical analysis if node is absolute difference, Jacobian or inverse consistency (see Section 1.2.3).

The patient hierarchy can be constructed in two ways. The first option is to manually create the whole patient hierarchy, from top to bottom level, with necessary attributes. Second option is to use an automatic script to look for files on hard drive and create corresponding levels. The second option is possible only by using proper naming conventions for file names and locations.



Bibliography

- [Her, 2012] (2012). Image-guided robotic surgery: update on research and potential applications in urologic surgery. *Curr Opin Urol*, 22(1):47–54.
- [Behringer et al., 2015] Behringer, P., Herz, C., Penzkofer, T., Tuncali, K., Tempany, C., and Fedorov, A. (2015). Open-source platform for prostate motion tracking during in-bore targeted mri-guided biopsy. 18(WS):122–9.
- [Bender and Tomé, 2009] Bender, E. T. and Tomé, W. A. (2009). The utilization of consistency metrics for error analysis in deformable image registration. *Physics in Medicine and Biology*, 54(18):5561.
- [Benjamin et al., 1998] Benjamin, E. J., Wolf, P. A., D'Agostino, R. B., Silbershatz, H., Kannel, W. B., and Levy, D. (1998). Impact of atrial fibrillation on the risk of death: the framingham heart study. *Circulation*, 98(10):946–952.
- [Bert et al., 2012] Bert, C., Engenhart-Cabillic, R., and Durante, M. (2012). Particle therapy for noncancer diseases. *Medical Physics*, 39(4):1716–1727.
- [Britton et al., 2007] Britton, K. R., Starkschall, G., Tucker, S. L., Pan, T., Nelson, C., Chang, J. Y., Cox, J. D., Mohan, R., and Komaki, R. (2007). Assessment of gross tumor volume regression and motion changes during radiotherapy for non-small-cell lung cancer as measured by four-dimensional computed tomography. *Int. J. Radiat. Oncol. Biol. Phys.*, 68(4):1036–1046.
- [Cappato et al., 2005] Cappato, R., Calkins, H., Chen, S. A., Davies, W., Iesaka, Y., Kalman, J., Kim, Y. H., Klein, G., Packer, D., and Skanes, A. (2005). Worldwide survey on the methods, efficacy, and safety of catheter ablation for human atrial fibrillation. *Circulation*, 111(9):1100–1105.
- [Castillo et al., 2009] Castillo, R., Castillo, E., Guerra, R., Johnson, V. E., McPhail, T., Garg, A. K., and Guerrero, T. (2009). A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology*, 54(7):1849.
- [Chen et al., 2008] Chen, M., Lu, W., Chen, Q., Ruchala, K. J., and Olivera, G. H. (2008). A simple fixed-point approach to invert a deformation field. *Med Phys*, 35(1):81–88.
- [Christensen and Johnson, 2001] Christensen, G. E. and Johnson, H. J. (2001). Consistent image registration. *IEEE Transactions on Medical Imaging*, 20(7):568–582.

- [Christensen et al., 1996] Christensen, G. E., Rabbitt, R. D., and Miller, M. I. (1996). Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447.
- [Cleary and Peters, 2010] Cleary, K. and Peters, T. M. (2010). Image-guided interventions: Technology review and clinical applications. *Annual Review of Biomedical Engineering*, 12(1):119–142. PMID: 20415592.
- [Fedorov et al., 2012] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F. M., Sonka, M., Buatti, J., Aylward, S. R., Miller, J. V., Pieper, S., and Kikinis, R. (2012). 3d slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging*, 30(9):1323 – 1341.
- [Fedorov et al., 2015] Fedorov, A., Khallaghi, S., Sánchez, C., Lasso, A., Fels, S., Tuncali, K., Sugar, E. N., Kapur, T., Zhang, C., Wells, W., Nguyen, P., Abolmaesumi, P., and Tempany, C. (2015). Open-source image registration for mri-trus fusion-guided prostate interventions. *Int J Comput Assist Radiol Surg*, 10(6):925–34.
- [Flampouri et al., 2006] Flampouri, S., Jiang, S., Sharp, G. C., Wolfgang, J., Patel, A. A., and Choi, N. C. (2006). Estimation of the delivered patient dose in lung imrt treatment based on deformable registration of 4d-ct data and monte carlo simulations. *Phys Med Biol*, 51(11):2763–2779.
- [Huttenlocher et al., 1993] Huttenlocher, D. P., Klanderman, G. A., and Rucklidge, W. J. (1993). Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863.
- [January et al., 2014] January, C. T., Wann, L. S., Alpert, J. S., Calkins, H., Cigarroa, J. E., Cleveland, Jr., J. C., Conti, J. B., Ellinor, P. T., Ezekowitz, M. D., Field, M. E., Murray, K. T., Sacco, R. L., Stevenson, W. G., Tchou, P. J., Tracy, C. M., and Yancy, C. W. (2014). 2014 aha/acc/hrs guideline for the management of patients with atrial fibrillation report of the american college of cardiology/american heart association task force on practice guidelines and the heart rhythm society. *Journal of the American College of Cardiology*, 64(21):e1–e76.
- [Kashani et al., 2007] Kashani, R., Hub, M., Kessler, M. L., and Balter, J. M. (2007). Technical note: A physical phantom for assessment of accuracy of deformable alignment algorithms. *Medical Physics*, 34(7):2785–2788.
- [Kirby et al., 2011] Kirby, N., Chuang, C., and Pouliot, J. (2011). A two-dimensional deformable phantom for quantitatively verifying deformation algorithms. *Medical Physics*, 38(8):4583–4586.

[Lehmann et al., 2015a] Lehmann, H. I., C, G., A, C., P, L., M, P, D, R., M, T, A, E, R, K., P, S., Fiedler, F, S, H., C, F, N, E, R, R, AK, R, D, T, HA, K, SB, J, KD, P, J, D, T, H, S, A, C, B, M, D, and DL, P (2015a). *Heart Rhythm*, volume Abstract. PO02-97, chapter Arrhythmia Ablation Using Scanned Carbon Ion Pencil Beams in a Porcine Model: First Outcomes from 12C Catheter-free Ablation.

[Lehmann et al., 2015b] Lehmann, H. I., Richter, D., Prokesch, H., Graeff, C., Prall, M., Simoniello, P., Fournier, C., Bauer, J., Kaderka, R., Weymann, A., Szabó, G., Sonnenberg, K., Constantinescu, A. M., Johnson, S. B., Misiri, J., Takami, M., Miller, R. C., Herman, M. G., Asirvatham, S. J., Brons, S., Jäkel, O., Haberer, T., Debus, J., Durante, M., Bert, C., and Packer, D. L. (2015b). Atrioventricular node ablation in langendorff-perfused porcine hearts using carbon ion particle therapy: Methods and an in vivo feasibility investigation for catheter-free ablation of cardiac arrhythmias. *Circulation: Arrhythmia and Electrophysiology*, 8(2):429–438.

[Leow et al., 2007] Leow, A. D., Yanovsky, I., Chiang, M. C., Lee, A. D., Klunder, A. D., Lu, A., Becker, J. T., Davis, S. W., Toga, A. W., and Thompson, P. M. (2007). Statistical properties of jacobian maps and the realization of unbiased large-deformation nonlinear image registration. *IEEE Trans Med*, 26:822–832.

[Li et al., 2015] Li, M., Wittek, A., Joldes, G., and Miller, K. (2015). Fuzzy tissue classification for non-linear patient-specific biomechanical models for whole-body image registration.

[Liu et al., 2007] Liu, H. H., Balter, P., Tutt, T., Choi, B., Zhang, J., Wang, C., Chi, M., Luo, D., Pan, T., Hunjan, S., Starkschall, G., Rosen, I., Prado, K., Liao, Z., Chang, J., Komaki, R., Cox, J. D., Mohan, R., and Dong, L. (2007). Assessing respiration-induced tumor motion and internal target volume using four-dimensional computed tomography for radiotherapy of lung cancer. *Int. J. Radiat. Oncol. Biol. Phys.*, 68(2):531–540.

[Liu et al., 2016] Liu, S., Cai, W., Pujol, S., Kikinis, R., and Feng, D. (2016). Cross-view neuroimage pattern analysis in alzheimer’s disease staging. *Front Aging Neurosci*, 8:23–.

[Liu et al., 2015] Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M., and ADNI (2015). Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer’s disease. *IEEE Trans Biomed Eng*, 62(4):1132–40.

[Lu et al., 2006] Lu, W., Olivera, G. H., Chen, Q., Chen, M. L., and Ruchala, K. J. (2006). Automatic re-contouring in 4d radiotherapy. *Phys Med Biol*, 51(5):1077–1099.

[Meyer et al., 2015] Meyer, A., Lasso, A., Ungi, T., and Fichtinger, G. (2015). Live ultrasound volume reconstruction using scout scanning. 9415:94152A–94152A–9.

- [Moore et al., 2004] Moore, C. S., Liney, G. P., and Beavis, A. W. (2004). Quality assurance of registration of ct and mri data sets for treatment planning of radiotherapy for head and neck cancers. *J Appl Clin Med Phys*, 5(1):25–35.
- [Mutic et al., 2001] Mutic, S., Dempsey, J. F., Bosch, W. R., Low, D. A., Drzymala, R. E., Chao, K. S., Goddu, S. M., Cutler, P. D., and Purdy, J. A. (2001). Multimodality image registration quality assurance for conformal three-dimensional treatment planning. *Int J Radiat Oncol Biol Phys*, 51(1):255–260.
- [Naini et al., 2010] Naini, A. S., Patel, R. V., and Samani, A. (2010). Ct-enhanced ultrasound image of a totally deflated lung for image-guided minimally invasive tumor ablative procedures. *IEEE Transactions on Biomedical Engineering*, 57(10):2627–2630.
- [Nithiananthan et al., 2011] Nithiananthan, S., Schafer, S., Uneri, A., Mirota, D. J., Stayman, J. W., Zbijewski, W., Brock, K. K., Daly, M. J., Chan, H., Irish, J. C., and Siewerdsen, J. H. (2011). Demons deformable registration of ct and cone-beam ct using an iterative intensity matching approach. *Medical Physics*, 38(4):1785–1798.
- [Pinter et al., 2012] Pinter, C., Lasso, A., Wang, A., Jaffray, D., and Fichtinger, G. (2012). Slicerrt – radiation therapy research toolkit for 3d slicer. *Med. Phys.*, 39(10):6332–6338.
- [Plastimatch, 2016] Plastimatch (2016). <http://plastimatch.org>.
- [Pujol et al., 2016] Pujol, S., Baldwin, M., Nassiri, J., Kikinis, R., and Shaffer, K. (2016). Using 3d modeling techniques to enhance teaching of difficult anatomical concepts. *Acad Radiol*, 23(4):507–16.
- [R et al., 2010] R, C., H, C., SA, C., W, D., Y, I., J, K., YH, K., G, K., A, N., Packer, D., Skanes, A., F, A., and E, B. (2010). *Circulation. Arrhythmia and electrophysiology.*, volume 3, chapter Updated Worldwide Survey on the Methods, Efficacy, and Safety of Catheter Ablation for Human Atrial Fibrillation, pages 32–38.
- [Rey et al., 2002] Rey, D., Subsol, G., Delingette, H., and Ayache, N. (2002). Automatic detection and segmentation of evolving processes in 3d medical images: application to multiple sclerosis. *Med Image Anal*, 6(2):163–179.
- [Richter, 2012] Richter, D. (2012). *Treatment planning for tumors with residual motion in scanned ion beam therapies.* Thesis/dissertation, TU Darmstadt.
- [Rueckert et al., 1999] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721.

- [Sarrut, 2006] Sarrut, D. (2006). Deformable registration for image-guided radiation therapy. *Z Med Phys*, 16(4):285–297.
- [Seppenwoolde et al., 2002] Seppenwoolde, Y., Shirato, H., Kitamura, K., Shimizu, S., Lebesque, J. V., and Miyasaka, K. (2002). Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 53(4):822–834.
- [Shackleford et al., 2010] Shackleford, J. A., Kandasamy, N., and Sharp, G. C. (2010). On developing b-spline registration algorithms for multi-core processors. *Physics in Medicine and Biology*, 55(21):6329–6351.
- [Slicer, 2016a] Slicer (2016a). <https://www.slicer.org/>.
- [Slicer, 2016b] Slicer (2016b). <http://www.slicer.org/publications>.
- [Thirion, 1998] Thirion, J. (1998). Image matching as a diffusion process: an analogy with maxwell's demon. *Med Image Anal*, 2(3):243 – 260.
- [Trofimov et al., 2005] Trofimov, A., Rietzel, E., Lu, H. M., Martin, B., Jiang, S., Chen, G. T., and Bortfeld, T. (2005). Temporo-spatial imrt optimization: concepts, implementation and initial results. *Phys Med Biol*, 50(12):2779–2798.
- [Varadhan et al., 2013] Varadhan, R., Karangelis, G., K., K., and Hui, S. (2013). A framework for deformable image registration validation in radiotherapy clinical applications. *J Appl Clin Med Phys*, 14(1).
- [Venugopal et al., 2005] Venugopal, N., McCurdy, B., Hnatov, A., and Dubey, A. (2005). A feasibility study to investigate the use of thin-plate splines to account for prostate deformation. *Physics in Medicine and Biology*, 50(12):2871.
- [Wu et al., 2008] Wu, Z., Rietzel, E., Boldea, V., Sarrut, D., and Sharp, G. C. (2008). Evaluation of deformable registration of patient lung 4dct with subanatomical region segmentations. *Medical Physics*, 35(2):775–781.
- [Yan,] Yan, D. Adaptive radiotherapy: Merging principle into clinical practice. *Seminars in Radiation Oncology*, 20(2):79 – 83.
- [Yan et al., 1997] Yan, D., Vicini, F., Wong, J., and Martinez, A. (1997). Adaptive radiation therapy. 42(1):123–132.
- [Yoo et al., 2002] Yoo, T. S., Ackerman, M. J., Lorensen, W. E., Schroeder, W., Chalana, V., Aylward, S., Metaxas, D., and Whitaker, R. (2002). Engineering and algorithm design for an image processing api: a technical report on itk—the insight toolkit. *Stud Health Technol Inform*, 85:586–592.

[Zhong et al., 2007] Zhong, H., Peters, T., and Siebers, J. V. (2007). Fem-based evaluation of deformable image registration for radiation therapy. *Physics in Medicine and Biology*, 52(16):4721.