Chapter 8
Validation and Comparison of Approaches to Respiratory Motion Estimation

Sven Kabus, Tobias Klinder, Keelin Murphy, René Werner and David Sarrut

Abstract The accuracy of respiratory motion estimation has a direct impact on the success of clinical applications such as diagnosis, as well as planning, delivery, and assessment of therapy for lung or other thoracic diseases. While rigid registration is well suited to validation and has reached a mature state in clinical applications, for non-rigid registration no gold-standard exists. This chapter investigates the validation of non-rigid registration accuracy with a focus on lung motion. The central questions addressed in this chapter are (1) how to measure registration accuracy, (2) how to generate ground-truth for validation, and (3) how to interpret accuracy assessment results.

8.1 Lack of a Gold-Standard in Non-Rigid Image Registration

Respiratory motion estimation is a topic receiving much attention in medical imaging. For clinical applications such as diagnosis as well as better planning, delivery, and assessment of therapy for lung or liver diseases, estimation of and compensation for motion is indispensable and its accuracy has direct impact on the success of the clinical applications.

S. Kabus (✉) · T. Klinder
Philips Research Laboratories, Hamburg, Germany
e-mail: sven.kabus@philips.com

K. Murphy
Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

R. Werner
University Medical Center Hamburg-Eppendorf, Hamburg, Germany

D. Sarrut
Université de Lyon, CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon, Université Lyon 1,
Centre Léon Bérard, Lyon, France

J. Ehrhardt and C. Lorenz (eds.), *4D Modeling and Estimation of Respiratory Motion for Radiation Therapy*, Biological and Medical Physics, Biomedical Engineering,
DOI: 10.1007/978-3-642-36441-9_8, © Springer-Verlag Berlin Heidelberg 2013

As detailed in the previous chapters of this book, respiratory motion estimation and compensation require non-rigid registration of CT thorax data typically acquired in a dynamic protocol as for respiratory-gated 4D scans. In each case, a voxel-wise computation of respiratory motion between different respiratory states is needed. However, while rigid registration is well suited for validation [1] and has come to a mature state in clinical applications, for non-rigid registration no gold-standard exists. Moreover, a discrepancy between the maturity of non-rigid registration in the image processing community and its dissemination in clinical workstations can be observed, indicating a lack of acceptance of the technique that can only be overcome by establishing commonly accepted validation metrics and procedures. This chapter therefore investigates current approaches for the validation of non-rigid registration when applied to motion estimation, focusing particularly on lung motion.

A necessary criterion for a successful registration is the alignment of visible image structures, often converted into an inspection of the residuum (i.e., the subtraction of the aligned data) where mis-aligned image structures show up. However, the absence of any structure in the residuum image does not guarantee that the non-rigid registration was successful, since the residuum is invariant to the deformation of image regions with homogeneous intensities. Even for image regions containing structures such as lung vessels an increased similarity of the aligned data (determined by a correlation coefficient for example) does not always imply an increased registration accuracy [2]. Furthermore, a registration scheme allowing for a very flexible alignment tends to result in an almost perfect residuum but suffers from implausible deformations and consequently from decreased accuracy. Additional evaluation criteria are therefore indispensable.

The most obvious method is probably to identify corresponding positions of anatomical structures in the images to be compared. For example, in lung data such features are usually anatomical points located on lung structures with adequate image contrast like vessel bifurcations, fissures and pleura, or the boundary of a potential tumor. To provide additional information, point positions can also be extended to line structures such as the centerlines of the bronchial tree defined at different respiratory states, and surface structures and volumes define objects like the lung fissures or target regions (e.g. masses) and organs at risk in radiotherapy. In each case, after applying the registration result, a *validation metric* such as a landmark-based registration error, or a line or surface alignment error is computed to measure registration accuracy quantitatively.

The validation metrics described so far focus on *morphological* structures but not on the physical and physiological plausibility of the entire deformation. A physically implausible result such as local folding in the deformation can be detected by analysis of the local volume change [3] as a simple example for *functional* validation. Another validation metric based on the entire deformation vector field measures the sensitivity of the registration result to the order of the input data (consistency metric). For a clinical application, the computed structural correspondence is certainly expected to be the same when aligning a follow-up scan with a baseline scan or vice versa (i.e., a consistent mapping). Moving from conventional diagnostic CT data to respiratory-gated CT data (4D-CT), anatomical positions can be tracked over (respiratory) time leading to motion trajectories. Since 4D-CT data usually suffer from lower spatial resolution and higher noise levels due to reduced radiation dose, and, in addition, often contain motion-induced artifacts [4] due to irregular patient breathing during image acquisition, the use of motion trajectories in combination with a breathing model for validation is beneficial.

A reliable ground-truth is essential to make use of any type of validation metric. However, there are many different types of ground-truth. Whereas morphological validation metrics such as landmark-, centerline- and overlap-based metrics require corresponding lists of annotated voxel positions, functional validation metrics rest upon underlying models, e.g. a certain breathing model for trajectory analysis or the positive Jacobian map as minimum requirement for a physiologically plausible tissue deformation. While in the past published validation studies [3, 5–7] have often been based on a limited number of registration algorithms and/or on proprietary datasets (typically with different imaging parameters such as dose or voxel resolution) there is an increasing trend towards multi-institutional validation studies. Recently, there have been two examples in the field of respiratory motion estimation. The Multi-Institutional Deformable Registration Accuracy Study (MIDRAS) [8] was motivated by the use of registration schemes for improved radiation therapy planning and therefore selected CT and MRI data showing the lungs, the liver and the prostate, while the Evaluation of Methods for Pulmonary Image REgistration 2010 (EMPIRE10) challenge [9], organized in conjunction with the Grand Challenge workshop[1] at MICCAI 2010, provided a public platform for comparison of registration algorithms applied to thoracic CT data. Based on the selected datasets, the participants calculated deformation vector fields and submitted them to the organizational teams for independent evaluation. Evaluation was, dependent on the study, performed considering anatomical landmarks, lung boundary alignment, fissure alignment, and the presence of deformation field singularities. Furthermore, the DIR-lab of the University of Texas M. D. Anderson Cancer Center[2] and the Léon Bérard Cancer Center together with the CREATIS-LRMN CNRS Research lab (POPI-model)[3] provide a series of freely available 4D-CT and exhale-inhale CT image data along with landmark lists for the validation and comparison of non-linear registration algorithms [2, 10].

Common to these studies and databases is the potentially exhausting task of generating ground-truth. The manual selection of landmarks, for example, is time-consuming and landmark locations are prone to uncertainties due to intra- and inter-observer variability and approaches for (semi-)automation are therefore desirable; finally, an interpretation of validation metric results also needs to be addressed since each validation metric poses only a single necessary condition.

Taking into account the above mentioned two principal types of validation metrics, their ability to identify characteristics of individual registration schemes and the issues of ground-truth generation, this chapter is divided into the description of

[1] http://www.grand-challenge.org/index.php/MICCAI_2010_Workshop

[2] http://www.dir-lab.com

[3] http://www.creatis.insa-lyon.fr/rio/popi-model

morphological validation criteria and of functional validation criteria. Each section then discusses the specific validation criterion in detail, including metric definitions, available ground-truth and the interpretation of corresponding results.

8.2 Morphological Validation Criteria

Morphological validation is currently the state-of-the-art in estimating the accuracy of registration results. Validation metrics defined in zero-dimensional space (e.g. landmarks) or higher spaces (e.g. tree structures or tumor surfaces) are frequently used and well understood. This section describes the validation metrics and the considered anatomical structures in the order of their dimensionality. Each subsection then describes how registration accuracy can be measured, what possibilities exist to generate ground-truth, and how to interpret the results when comparing multiple registration schemes.

8.2.1 Landmarks

8.2.1.1 Validation Metrics

Landmarks are usually understood as being characteristic anatomical points, which can therefore be considered as being zero-dimensional features and leading to validation metrics defined in zero-dimensional space, respectively. Typical landmark candidates are, as described in Sect. 8.1, salient bifurcations of the bronchial tree (lungs), specific branches of vessel trees (lungs, liver), or calcified nodules [5, 8, 11, 12], cf. Fig. 8.1.

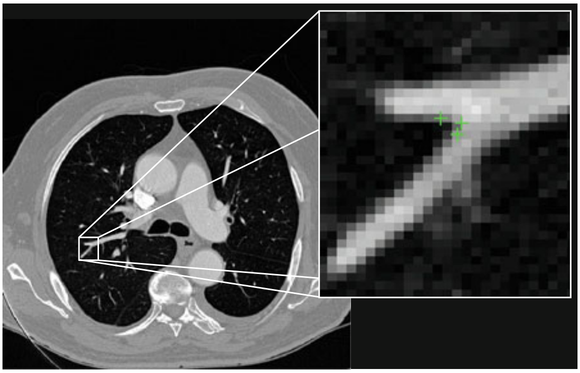


Fig. 8.1 Example of a landmark in a lung CT data set, selected by three medical experts. It can be seen that manual landmark selection and subsequent quantitative registration evaluation suffers from interobserver variability of landmark identification

For evaluation purposes, a (usually relatively sparse) set of landmarks is identified within each of the images A and B to register. For a single landmark $\boldsymbol{x}_A$ within the discrete domain of image A, Ω_A, and a spatial transformation $\boldsymbol{T}$ mapping Ω_A onto Ω_B, the standard validation metric is the Euclidean distance between the mapped landmark position $\boldsymbol{x}_B^{[pred]} = \boldsymbol{T}(\boldsymbol{x}_A)$ after registration and the position $\boldsymbol{x}_B^{[actual]}$ in Ω_B that is anatomically corresponding to landmark $\boldsymbol{x}_A$. In literature, this metric

$$\mathrm{TRE}(\boldsymbol{x}_A) = \left\| \boldsymbol{x}_B^{[pred]} - \boldsymbol{x}_B^{[actual]} \right\|_2 \tag{8.1}$$

is also often referred to as landmark-based, point, or target registration error [2, 13], and to summarize the error distribution for a set of landmarks, the mean error, the standard deviation, the maximum error and/or error quantiles are commonly considered.

Whilst landmark-based registration errors as defined above represent (in the sense of a metric) absolute, non-negative error values, over- or underestimation of respiratory motion along a certain direction can also be identified using landmark sets. Therefore let the direction of interest being represented by a unit vector $\boldsymbol{e_a}$ along a vector $\boldsymbol{a}$; then, directional errors and over/underestimation can be deduced from the projections of the misalignments $\left(\boldsymbol{x}_B^{[pred]} - \boldsymbol{x}_B^{[actual]}\right)$ onto $\boldsymbol{e}_a$.

8.2.1.2 Ground-Truth Generation

Ground-truth for evaluating the landmark-based registration error is commonly generated by annotating corresponding landmarks within the images to register manually and is usually carried out by ‘medical experts’ (radiologists, medical students, etc.). The manual selection of such points is, however, time-consuming and landmark locations are prone to uncertainties due to intra- and interobserver variability concerning exact point selection [5, 12] (cf. Fig. 8.1), for instance caused by low image resolution or partial volume effects. In the case of lung CT registration, landmark identification additionally suffers from low contrast in near-to-pleura regions, which often leads to very limited landmark sets grouped around the mediastinum [3]. To serve as reliable ground-truth, the landmarks should preferably consist of a well-distributed set of verifiable anatomical correspondences throughout the image region of interest (e.g. the lungs) and be large enough in number to enable meaningful statistical analysis. The required number of landmarks can be assessed by a posteriori statistical sample size calculation [2]. The number may vary for individual data sets and motion estimation approaches, but it has been reported to be even more than 1000 anatomical point pairs [2] and so efforts have been made to (partially) automate identification of landmark sets.

One such algorithm is described in [11, 14], here serving as an example. The algorithm starts with automatic detection of landmarks in an image A. A so-called distinctiveness term is defined to quantify the distinctiveness of a voxel within its local neighborhood. The distinctiveness term combines both differential properties (gradient magnitude) and intensity characteristics to quantify the suitability of the voxel as a landmark candidate. Good distribution throughout the region of interest is ensured by forcing a minimum Euclidean distance between the landmark candidates. In a second step, the interactive landmark transfer to the image B to be aligned with image A by registration is supported by computing and progressively refining a thin-plate-splines transformation based on user-annotated landmark correspondences. After manually transferring an adequate number of landmarks, the transformation can be applied to guide the user to find correspondences for the remaining landmarks or even be used to transfer the landmarks fully automatically.

The algorithm has recently been applied for a number of registration evaluation studies [3, 9, 15], but it is only one example for construction of a landmark-based ground-truth. For instance, reviving earlier works on landmark detection [16] it has been suggested to incorporate curvature-based operators for distinctiveness calculation [17] or to consider Shannon entropy instead [18]. Furthermore, template matching methods have been applied for landmark transfer in order to fully automate the evaluation process [17, 19]; however, especially automating the landmark transfer can be controversial [2]. The transfer represents, by definition, a (point-based) registration problem. Thus, a (semi-)automatic landmark transfer may lead to biased evaluation if the registration method to evaluate and the landmark transfer methods are similar in some sense (for example if they maximize the same similarity measure).

8.2.1.3 Validation in Practice

The usage of landmarks is the most popular method of validation of non-linear registration. As well as the validation studies mentioned above and the described publicly accessible 4D-CT databases, numerous research articles have used anatomical landmark sets denoted by experts.

However, the actual size of the landmark sets varies from small numbers of about 20 landmark positions to large sets of 1500 positions, cf., e.g. [2, 20]. Naturally, the larger the number of landmarks, the better the estimation of accuracy is likely to be (cf. requirements for ground-truth generation described above). For a simple illustration beyond pure statistical computations, consider a lung with a volume of 4 l together with a set of 20 landmark positions. In this example, each cube of lung parenchyma with edge length 58 mm contains one landmark on average. Increasing the set of landmarks to 100 or even to 1500, the edge length of this cube reduces to 34 mm and 14 mm, respectively. A landmark spacing of a size of 14 mm undoubtedly allows for registration accuracy estimation on a coarse scale but not on a finer scale when taking into account the distinctive inhomogeneity of lung parenchyma.

Another critical point is the requirement of the landmarks to represent a well-distributed set over the structure/region of interest. If landmarks are placed on the bifurcations of major blood vessels as shown exemplarily in Fig. 8.2, left, they are concentrated around the mediastinum which has a number of disadvantages: (1) this type of ground-truth does not allow for accurate estimation in regions near to the pleura or diaphragm where registration accuracy is typically worse; (2) the stiffness

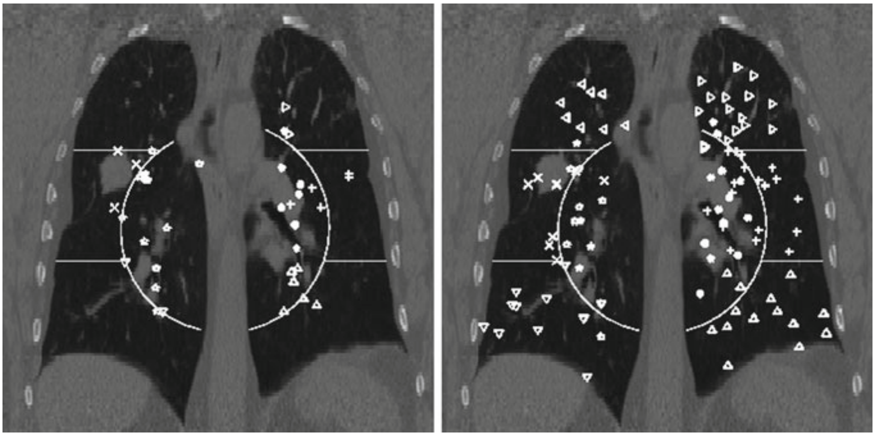


Fig. 8.2 Region definition and landmark positions shown in a coronal projection for a landmark set as commonly used in the literature (*left*) and for a well-dispersed landmark set (*right*). For improved visualization each region is given a unique landmark symbol

of the lungs is highest in the surroundings of the major bronchial tree—reducing the local motion and meaning that the registration task is most difficult elsewhere; (3) major bifurcations are clearly visible to a human observer but also to the majority of registration schemes—unlike the hardly detectable low-contrast structures near to the pleura. The importance of the requirement of a good distribution of the landmarks in the region of interest is demonstrated in [3]. The authors compared landmark-based registration error as obtained for two landmark sets, one set as commonly used in the literature (cf. POPI-model), the other set well-distributed throughout the lung volume (shown in Fig. 8.2). A collection of six popular methods including surface- and volume-based as well as parametric and non-parametric methods was investigated. From each of these algorithms, a deformation vector field was extracted and used to transform the landmarks from both sets. While the mean landmark-based registration error on both landmark sets differs only slightly, a region-based analysis reveals smaller errors in apical regions but also a significantly higher error in the lower right lung (Fig. 8.3) and therefore a dependency of the landmark-based registration error on the distribution of landmarks. This dependency is observed for each of the six registration schemes.

Conclusion: Landmarks are a popular and intuitive method of registration validation. Both a large number of corresponding positions and a good distribution of the points throughout the organ of interest are crucial for reliable registration accuracy assessment. However, landmarks estimate registration accuracy only at selected locations and additional validation metrics are beneficial to provide deeper insight.

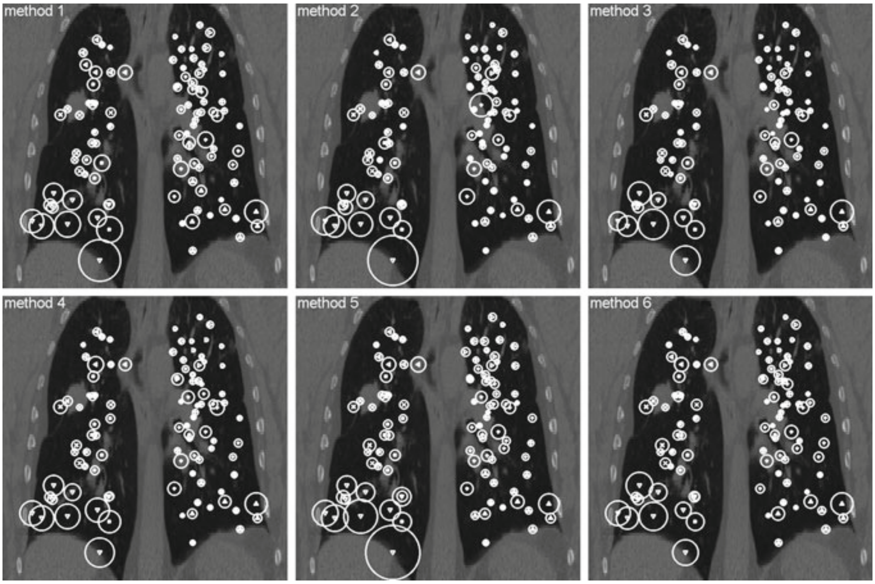


Fig. 8.3 Landmark-based registration error, shown proportional to the spheres' diameters after registration by six different algorithms. Note the different errors in apical regions and lower right lung indicating the need for a well-dispersed set of landmarks

8.2.2 Line-Like Anatomical Structures

8.2.2.1 Validation Metrics

Landmarks focus on prominent points which are in most cases bifurcations of airways or the vessel tree. In order to extend the evaluation, validation metrics for line-like anatomical structures analyze he registration accuracy not only at discrete positions, but take the properties of the structure (e.g. vessel or airway) into account. For example, significant changes in curvature, folding along a branch, or implausible changes in branch length are interesting properties that can not be captured by landmarks. In the following, it is assumed that for both images a segmentation of the airways and/or vessels is given. Furthermore, the individual branches are labelled, so that for each branch in one image, the corresponding part in the other image is known. From the segmentation, a centerline representation can be derived and the branching points can be extracted.

Now, consider a given set of centerline points defined between two branch points of consecutive branching generations. An adequate interpolation scheme can be chosen to represent the point set as a continuous curve with a parametrization as a function $\alpha(t)$ with $t \in [0, 1]$. The corresponding curve described by a function β can be derived in the other image as well. Having two continuous corresponding line segments allows the definition of a variety of distance measures.

As an example, the registration accuracy along the centerline can be evaluated as the difference between corresponding points that have the same value of the parametrization constant as

$$D_1^i(t) = ||T(\boldsymbol{c}_{\alpha(t)}^{i,A}) - \boldsymbol{c}_{\beta(t)}^{i,B}||_2\,, \tag{8.2}$$

where i defines the branch index and $T(\boldsymbol{c}_{\alpha(t)}^{i,A})$ is the transformed centerline belonging to the ith branch defined in image A. Note that this measure is sensitive to a change in the length of the centerline. As the structures considered in the lungs can be assumed to preserve their absolute length during respiration, this sensitivity might be desirable.

Alternatively, it is also possible to evaluate the distance by finding the closest point on the corresponding other centerline as

$$D_2^i(t_A) = \min_{t_B \in [0,1]} ||T(\boldsymbol{c}_{\alpha(t_A)}^{i,A}) - \boldsymbol{c}_{\beta(t_B)}^{i,B}||_2\,. \tag{8.3}$$

Based on the centerline representation, other measures can be evaluated that consider line properties of the transformed centerline, for example local curvature, to detect implausible deformations. However, so far metrics for line-like structures have rarely been applied, mainly because segmentation and labelling of the considered structures is difficult to achieve.

8.2.2.2 Ground-Truth Generation

Examples that can be considered for line-like structures are airways and vessels. In each case the ground-truth generation relies on a segmentation and a labelling. However, manual segmentation and labelling is very time-consuming and from a practical point of view not always possible. Although a variety of automatic segmentation algorithms exists, failure especially in the case of pathologies or low image resolution, which is the case in 4D-CT, is likely to occur. A reliable ground-truth may be most efficiently obtained by applying an automatic segmentation first which is then inspected and manually corrected.

Both airways and blood vessels form dense tubular structures, but typically differ in appearance in a CT image. Thus, most approaches that take into account the tubular characteristic of the structure to be segmented can be often applied for both airways and vessels by simply changing the appearance parameters.

Many existing algorithms for airway and vessel segmentation are based on region growing [21]. However, in areas where the contrast is low, for example due to resolution and noise, leakage is observed. One possibility to circumvent this problem is by means of explosion control, for example by introducing certain rules derived from anatomical knowledge [22], or using template tracking based methods [23, 24]. For a recent overview on vessel segmentation techniques see [25] as well as algorithms described within challenges on vessel segmentation in the lung (VESSEL12)[4] and on airway extraction (EXACT09, [26]).[5]

For matching of both airways and vessels, different approaches have been proposed [27, 28]. However, as the methods have been used on different data sets, it is not clear which is currently the method of choice.

8.2.2.3 Validation in Practice

Validation based on labelled airways or vessels as introduced in Sect. 8.2.2.1 has, to our knowledge, not been published so far. While a considerable amount of work has been done on automatic segmentation as well as on matching of both airways and vessels, little of this information has been used for registration evaluation. This might be due to the fact that even with the help of automatic segmentation algorithms, generating a reliable ground-truth would require the verification of the obtained segmentations which is still very difficult and time-consuming. Furthermore, automatic extraction of both airways and vessels is still difficult on 4D-CT with low resolution and severe pathologies.

Conclusion: Compared to the use of landmarks that evaluate the registration accuracy at distinct locations, metrics described here can be used to measure the registration accuracy along a line-like structure. These measures have the potential to provide additional valuable insight into the registration, for example by detecting folding of a branch or change in branch length, etc. Nevertheless, measurements based on line-like structures have not been used so far because of the difficulties of obtaining a reliable ground-truth.

8.2.3 Surface Structures and Volumes

8.2.3.1 Validation Metrics

Estimating the registration accuracy for anatomical structures like the lungs, the lobes and the fissures leads to validation metrics for surface structures and volumes. The segmentations of a corresponding anatomical surface area or volume in two images A and B are denoted as voxel sets S_A and S_B, respectively, with the transformed voxel set denoted as $\tilde{S}_A$ (cf. Fig. 8.4).

One common measure for evaluation of the registration accuracy using surface structures such as the outer lung boundaries or the fissures is the average surface distance. For each voxel $\boldsymbol{x}_B$ contained in S_B, the closest voxel $\boldsymbol{x}_A$ in $\tilde{S}_A$ is determined and the Euclidean distance between them is calculated as

[4] http://vessel12.grand-challenge.org

[5] http://image.diku.dk/exact

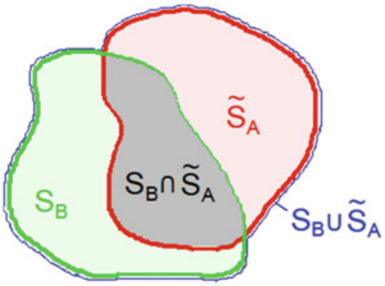


Fig. 8.4 Illustration of union and intersection between two volumes

$$d(\boldsymbol{x}_B, \tilde{S}_A) = \min_{\boldsymbol{x}_A \in \tilde{S}_A} \|\boldsymbol{x}_B - \boldsymbol{x}_A\|_2 \ . \tag{8.4}$$

To ensure symmetry, the error $d(\boldsymbol{x}_A, S_B)$ at each voxel in $\tilde{S}_A$ is also calculated and finally the average overall error is computed.

For the calculation of the volumetric overlap between two voxel sets several methods exist [29–32] where the most frequently used are

- *Dice coefficient* (DC) [33] which is also called volume overlap index [34]

$$\text{DC} = \frac{2(|S_B \cap \tilde{S}_A|)}{|S_B| + |\tilde{S}_A|} \ , \tag{8.5}$$

- *Jaccard coefficient* (JC) or *volumetric overlap* (VO) [35]

$$\text{JC} = \text{VO} = \frac{|S_B \cap \tilde{S}_A|}{|S_B \cup \tilde{S}_A|} \ , \tag{8.6}$$

which can also be alternatively calculated as [36]

$$\text{JC} = \frac{|S_B| + |\tilde{S}_A|}{|S_B \cup \tilde{S}_A|} - 1 \ , \tag{8.7}$$

- *target overlap* (TO)

$$\text{TO} = \frac{|S_B \cap \tilde{S}_A|}{|S_B|} \ . \tag{8.8}$$

8.2.3.2 Ground-Truth Generation

Ground-truth generation for surface structures and volumes requires the segmentation of the respective objects of interest. For the purpose of evaluation of lung motion estimation, the most relevant structures are (i) lungs, (ii) lung lobes and fissures and (iii) potential tumors. Manual segmentation of those structures is usually very time-consuming. Thus, ground-truth generation that results from semi-automatic or automatic segmentation algorithms which are inspected and manually corrected is much more feasible from a practical point of view.

For the segmentation of the lungs, automatic approaches have been presented ranging from voxel-based segmentation methods [37] to multi-atlas registration [38]. Voxel-based methods are based on the assumption that, for normal lung parenchyma, there is a large difference in attenuation between the lung parenchyma and the surrounding tissue. While those methods have low computational time, they fail especially in the case of pathological lungs or image artifacts. Other methods that involve prior knowledge give potentially better results on pathological cases but have a significant increase in runtime. Further algorithms are described within a challenge on lobe and lung analysis (LOLA11).[6]

Fissure segmentation has been recently described in [39, 40] and extensions have been presented to deal with incomplete fissures or cases where the fissures are hardly visible [41]. Interactive methods [42] allow for correction of a given automatic segmentation result or manual segmentation from scratch.

In the context of radiotherapy planning and treatment, evaluating the correctness of tumor motion estimation is of major importance. For ground-truth generation, a variety of methods for semi-automatic and automatic tumor segmentation exists including vessel removal and pleural surface removal [43].

8.2.3.3 Validation in Practice

Volumetric overlap measures are well established and often applied to evaluate the results of an automatic segmentation [38]. However, especially in the case of large volumes, surface distance metrics are probably more relevant as there can still be quite large errors near the boundaries even though large parts of the volumes are overlapping.

For the registration methods from Fig. 8.3, a careful inspection of the pleura revealed good alignment with no significant inter-method variation. This observation is supported by the EMPIRE10 challenge [9] where 12 (26) out of 34 methods matched more than 99.99 % (99.9 %) of pleura-adjacent voxels correctly to either the interior or the exterior of the lung boundary.

Unlike the lung boundaries, the fissures are of much lower contrast in CT and thus more challenging to align in particular for larger motion amplitudes. In the EMPIRE10 challenge 2 (20) out of 34 methods matched more than 99.9 % (99 %) of fissure-adjacent voxels to the correct lung lobe. Registration of fissures is examplarily shown in Fig. 8.5. For visual inspection, the fissure of the left lung (marked by green plus signs) as extracted from the reference image (shown top left) is overlayed onto the transformed template images from three of the six registration methods under consideration. Although none of the methods employs dedicated knowledge about

[6] http://www.lola11.com

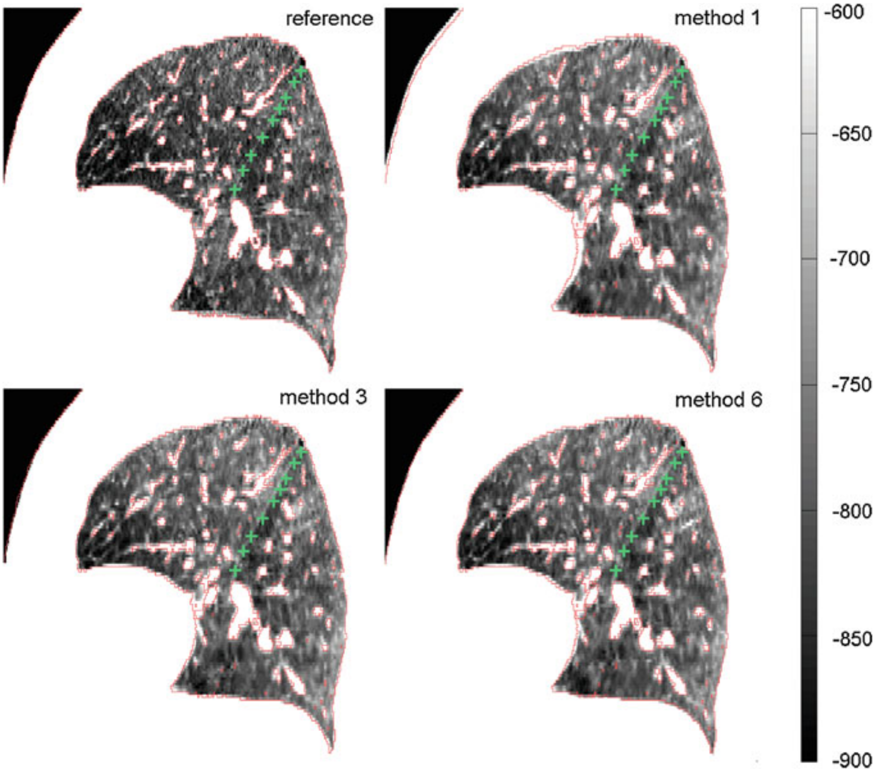


Fig. 8.5 Sagittal view of reference image (*top left*) and transformed template images after registration with different registration algorithms. *Green plus signs* and *red contours* indicate fissure and −650 HU iso-contour and are extracted from the reference image. Their overlay onto the transformed template images allow for visual inspection of fissure and vessel alignment

the fissures (e.g. by detecting them first), the fissures are roughly matched with a misalignment of only one to two voxels. Taking into account their low contrast, it can be assumed that matching of the fissures is assisted by high contrast surrounding vessel structures guiding the algorithm towards the desired deformation result (cf. Fig. 8.5 where the iso-contour is defined at −650 HU in the reference image and overlayed onto each transformed template image).

For both lung boundaries and fissures, it should be noted that surface-based metrics do not evaluate the motion in the tangential direction. At the lung boundary slipping occurs along the rib-lung interface, while sliding motion can also occur along the fissures which are built up of two tissue layers with lubricant fluid in-between. Evaluation of this sliding effect is not captured by the surface-based metrics.

While for both the lung boundary as well as the fissures, the individual methods seemed to have very similar performance results, inter-method differences exist that can be highlighted, for example, by comparing residual images. The residual image is

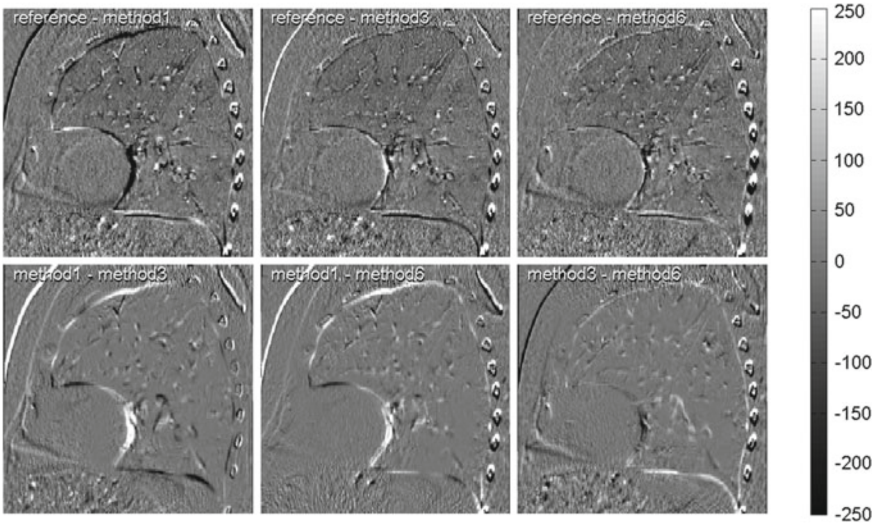


Fig. 8.6 Same view as in Fig. 8.5 but with residuals (*top row*) and subtraction of residuals (*bottom row*) depicted

defined as the subtraction of a transformed template image from the reference image (cf. Fig. 8.6, top row). Misaligned image structures show up whereas well-aligned structures cancel out after subtraction (given the same imaging protocol). It can therefore serve as a rough indicator of successful registration. Optimizing a registration result on the basis of the residual image will provide very little regularization of the deformation. Registration accuracy, however, does not necessarily improve with decreased regularization and is, moreover, often worsened since image noise starts to dominate the computed deformation.

Figure 8.6, bottom row, displays the residual images after subtraction from each other. These secondary residuals highlight image regions being differently deformed by the three registration methods under consideration. Moreover, secondary residuals are less impacted by differences in SNR or parenchymal densities between reference and template image since both the minuend and the subtrahend are interpolated from the template image. For the investigated case pleura and pleura-adjacent vessels are more similarly transformed by methods 3 and 6 than by 1 and 3 or 1 and 6. On the other hand, methods 1 and 6 better agree for vessel structures near to the fissure.

Conclusion: Both surface-based as well as volume-based metrics are well established methods to evaluate the overlap of two structures. While the methods can be used to provide a first indication of the quality of a registration method, detailed analysis of the transformation is limited. Most prominently, in the case of lung boundaries and fissures surfaces, sliding motion that occurs in tangential direction along the boundary is not captured by surface-based metrics. Furthermore, as surface-based metrics evaluate the overall overlap (lung volume overlap) or registration accuracy at distinct structures (fissures), detailed information about particular regions or far away from the considered structures is difficult to derive as demonstrated in Fig. 8.6.

8.3 Functional Validation Criteria

In the previous section, a number of morphological validation criteria are described. Among these are anatomical landmarks measuring the correspondence of point positions and surface structures measuring the alignment of lung boundaries or lung fissures. All these criteria, however, focus on certain anatomical structures. As soon as the anatomy under investigation is aligned, from a morphological point of view it is irrelevant if and how other regions are transformed. Functional validation criteria can fill this gap by adding prior knowledge in terms of assumptions on the functional behaviour of the lung tissue.

The following subsection revisits landmarks but now considered in the context of (respiratory-gated) 4D datasets where they define a motion trajectory over time. Another class of validation metrics is based on the deformation vector field (DVF). Contrary to morphological validation criteria, ground-truth data does not exist for these metrics, thus careful interpretation of the results is required.

8.3.1 Trajectory Analysis

8.3.1.1 Validation Metrics

Landmark-based evaluation relies on a set of independent point positions, corresponding to anatomical features. When dealing with respiratory-gated 4D thoracic datasets, each point is expected to follow a cyclic trajectory and additional prior knowledge can be used on this trajectory. Such prior information is important for deriving robust and efficient algorithms but can serve also in the validation process.

The conventional approach for the registration of 4D-CT datasets is to compute a set of DVFs—either between a designated reference phase and all remaining phases or between any adjacent respiratory phases. An alternative is to register the designated reference phase with the entire 4D dataset (sometimes referred to as group-wise registration). This is equivalent to estimating motion trajectories of individual point positions and can be called spatio-temporal registration.

In this context, it can be interesting to generalize the landmark-based registration error to take into account the time spent at the main phases of a trajectory. For example, an error metric could take into account that more time is spent at the end-inspiration and end-expiration phases than between these extremes and therefore be designed to reflect the decreased likelihood for motion-induced image artifacts. In other words, estimation of errors at an intermediate phase of the cycle should have lower weight than errors at extreme phases. According to known prior 1D breathing models, such as the one proposed by [44], locally defined material points move along their trajectories at variable speeds (determined by the derivative of the globally defined lung volume curve). One such metric is called spatio-temporal error (STE) [12] and is defined by

$$STE_{t_a,t_b}(L_1, L_2) = \frac{1}{t_b - t_a} \int_{t_a}^{t_b} ||L_1(s(t)) - L_2(s(t))||_2 \qquad (8.9)$$

where L_1 and L_2 are two parametric trajectories (each of which defines the set of the different locations of a material point during its motion), $s(t)$ the normalized curvilinear abscissa of the trajectory according to a prior breathing model to take into account the relative breathing velocity. This abscissa is a function of time and denotes the curve length travelled between initial time t_a and time t. However, in practice, such methods have not been shown to be more effective or more robust than conventional landmark-based registration error in evaluating registration accuracy.

8.3.1.2 Validation in Practice

To our knowledge, few studies analyze respiratory motion estimation in terms of point trajectories. Sets of 4D landmark points sets were used to test the adequacy of some trajectory model. In [45] the authors proposed a 4D local trajectory model for thoracic 4D-CT, where trajectories were modeled with cubic polynomials through the expiratory phases (neither cyclic nor inspiratory phase was included). For the validation, experts were asked to select corresponding anatomical points in expiratory phases with the help of a dedicated GUI named Assisted Point Registration of Internal Landmarks (APRIL) [6].

In [46], projection sequences of cone-beam images were used to analyse craniocaudal positions of the diaphragm over time. A database of motion was obtained and used to assess the validity of several trajectory models. Validation of respiratory motion estimation on 4D-CT was then performed with sets of about 100 expert-selected corresponding points by temporal frame, using a semi-automatic software [11] (see Fig. 8.7 for an illustration of such trajectories). The mean distance between the experts' annotations was 0.5 mm (0.9 mm standard deviation).

In practice, two main issues are encountered. Firstly, the manual (even semi-automatic) definition of landmarks across the 4D dataset is very time-consuming. In addition, an alarmingly high number of acquisitions contain motion-induced artifacts, mainly due to irregular patient breathing during image acquisition. In the case of artifacts, the image information can be considered locally invalid, as it does not correspond to the patient anatomy. Clinical use of the estimated motion fields requires them to be as close to the unknown reality as possible. A patient-specific, spatio-temporal deformation model could assist in reducing sensitivity to local image irregularities and render the motion estimate more plausible and potentially more representative of the patient's breathing motion under these challenging circumstances. In [46], robustness of registration methods was illustrated and compared by

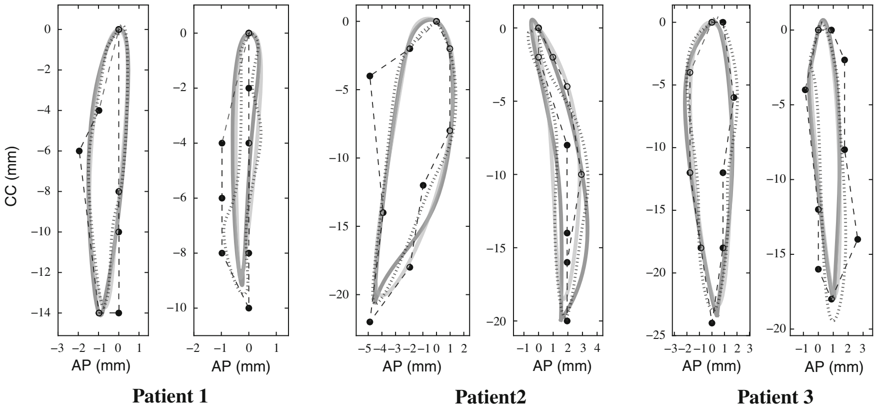


Fig. 8.7 (from [47]). Example landmarks with large displacements projected onto the sagittal plane: manually identified landmark positions are plotted together with trajectories obtained using three different trajectory models (shown using *black*, *gray* and *dashed lines*)

introducing artificial artifacts in 4D-CT. Qualitative evaluation was performed visually on images with real artifacts. From this very limited dataset, it seems that the addition of temporal information improves the robustness of the registration.

Conclusion: To validate the robustness of a registration method remains challenging, but taking into account the temporal dimension by means of point trajectories may be useful.

8.3.2 Deformation Vector Field Analysis

8.3.2.1 Validation Metrics

This paragraph describes validation metrics based on the deformation vector field (DVF) $\boldsymbol{T}$ mapping the discrete domain of an image A onto that of an image B, $\boldsymbol{T} : \Omega_A \rightarrow \Omega_B$.

One such validation metric employs the asymmetric nature of image registration[7]. A natural assumption is that independent of the order of the input images, the resulting transformations are inverse to each other. More specifically, defining $\boldsymbol{T}_{A \rightarrow B}$ as the spatial transformation after registration with A as the reference image and $\boldsymbol{T}_{B \rightarrow A}$ as that with B as the reference image, for an ideal registration $\tilde{\boldsymbol{x}}_A := \boldsymbol{T}_{B \rightarrow A}\left(\boldsymbol{T}_{A \rightarrow B}\left(\boldsymbol{x}_A\right)\right)$ is equal to $\boldsymbol{x}_A$ for all voxels of image A. More generally, in the case of n images with computed pairwise transformations $\boldsymbol{T}_{A_1 \rightarrow A_2}$, $\boldsymbol{T}_{A_2 \rightarrow A_3}$,

[7] Symmetric registration approaches are independent of the order of input images while asymmetric approaches are not. Naturally, a validation metric based on a specific criterion is not designed for approaches already relying on the same criterion.

$\dots \boldsymbol{T}_{A_{n-1}\to A_n}$, $\boldsymbol{T}_{A_n\to A_1}$ let $\tilde{\boldsymbol{T}}$ denote the concatenation of all transformations,

$$\tilde{\boldsymbol{T}} := \boldsymbol{T}_{A_n\to A_1} \circ \boldsymbol{T}_{A_1\to A_2} \circ \dots \circ \boldsymbol{T}_{A_{n-1}\to A_n} \; . \tag{8.10}$$

The n-consistency metric [48] is then defined as

$$C(\boldsymbol{x}_A) := \|\tilde{\boldsymbol{T}}(\boldsymbol{x}_A) - \boldsymbol{x}_A\|_2 \; , \quad \boldsymbol{x}_A \in \Omega_A \; . \tag{8.11}$$

It is common to consider mean and standard deviation as well as the maximum of the consistency map C.

A second DVF-based validation metric investigates the local volume change at every voxel position. In particular, it measures how much an infinitesimally small region around a voxel is contracting or expanding. For applications involving respiratory motion, large contractions of parenchymal tissue can occur. Taking a contraction beyond its physically possible limit, however, it can occur that two anatomical positions different from each other are mapped onto the same anatomical position, resulting in a loss of image information. This property is often described as the limit to folding, non-invertibility, non-bijectivity or non-diffeomorphism. For registration tasks containing pathologies, e.g. registration of a pre-operative planning image with an image taken after tumor resection, a locally folding DVF can be reasonable. However, respiratory motion estimation in this chapter assumes the absence of intervention related tissue loss. Therefore, any occurences of folding indicate a local registration failure. This is measured by calculation of the Jacobian map $\det(\nabla \boldsymbol{T}_{A\to B}(\boldsymbol{x}_A))$. Folding is defined as

$$F := \{\boldsymbol{x}_A \in \Omega_A \mid \det(\nabla \boldsymbol{T}_{A\to B}(\boldsymbol{x}_A)) \leq 0\} \; . \tag{8.12}$$

8.3.2.2 Validation in Practice

Since ground-truth is not available, necessary conditions for registration accuracy, so-called indications, can be considered. For the following discussion, the comparison study [3] cited in previous subsections is referenced again.

Firstly, for registration methods with both forward and backward DVFs available, the 2-consistency metric is computed. Since none of the registration methods under consideration is consistent by definition, measuring the consistency error is a suitable indicator of how independent the registration result is from the image input order. Figure 8.8, top row, as well as quantitative evaluation (details are described in [3]) indicate different consistency errors for methods 1, 3, 4 and 6.

Next, the Jacobian map is computed, a metric which is often used as a surrogate for local lung ventilation estimation. Since each DVF considered here is defined in the end-inspiration domain, a contraction is generally expected. A comparison of the various methods (cf. Fig. 8.8, bottom row) revealed large differences: whereas methods 1, 3 and 6 show a relatively homogeneous contraction, the remaining

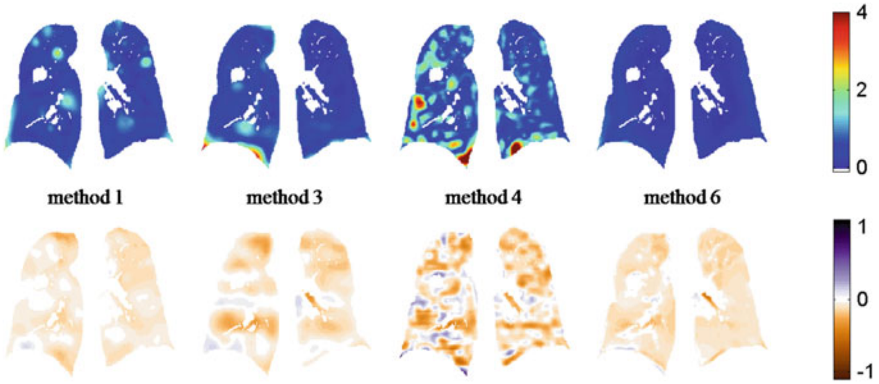


Fig. 8.8 2-consistency metric (*top*) and Jacobian map (*bottom*) for coronal views after registration with different registration algorithms. The consistency error as the amplitude of the geometrical discrepancy between forward and backward registration is given in mm. The Jacobian map estimates local lung ventilation: positive (negative) values indicate local expansion (contraction)

methods result in heterogeneous contraction-expansion patterns, methods 2 and 5 even show severe foldings [3]. Folding in the spatial domain is equivalent to partial loss of image information and should therefore be avoided. However, even without folding, heterogeneous contraction-expansion patterns are less plausible from a physiological point of view. Lungs in healthy condition are expected to be equally ventilated throughout with two exceptions: (1) gravity can impact both parenchymal density and ventilation resulting in a ventral-dorsal gradient [49, 50], (2) inertia can cause the lower lung regions to be more ventilated at intermediate respiratory phases when using dynamic acquisition protocols (as in respiratory-gated 4D-CT for example) rather than breath-hold imaging.

Lung diseases such as emphysema or fibrosis can cause ventilation of a certain lung region to be poor or even absent. Diseases showing (partially) obstructed airways can lead to an abnormal local level of synchrony since air flow entering a lung compartment with obstructed airways is slowed down and therefore continues to fill this compartment after the rest of the lung has stopped inhalation and switched to exhalation [51]. Such diseases can explain a heterogeneous level of contraction between end-inspiration state and end-expiration state but not a mix between contraction and expansion. On this basis, an experienced radiologist rated the result from method 6 (Fig. 8.8, last column) as the most plausible one, followed by method 1 (Fig. 8.8, first column).

Conclusion: Ground-truth in a strict sense does not exist for DVF-based validation metrics. Metrics such as consistency or the Jacobian map serve as necessary conditions for registration accuracy, thus they are indicators only. Registration methods with similar accuracy measured by morphological metrics can differ significantly for DVF-based metrics illustrating out the inadequacy of landmark- or surface-based validation alone. Use of the Jacobian map to define lung ventilation is an important measure of functional validation.

8.3.3 Beyond Pure Deformation

Functional criteria from the previous section such as the consistency metric and the Jacobian map were demonstrated to add further information to the validation space. This can increase plausibility of the registration result and highlight previously unseen variations among different registration methods. A fully-automated evaluation of the Jacobian map is, however, difficult to establish: regions undergoing folding can, without any doubt, be classified as local registration failure. But what about regions compressed to 30 % or even to 10 % of their original volume, for example? Without dedicated knowledge of parenchymal elasticity in general and the specific patient status in particular, it is certainly not possible to rate this level of compression. Likewise, the radiologist's preference for a homogenoeus contraction as shown in Fig. 8.8 can not be automatically achieved.

One possible way to deal with the limitations of the Jacobian map for automatic validation is to link it to the voxel-specific tissue density. From a functional perspective, the lungs are comprised of tissue structures ranging from stiff to elastic. In CT imaging, higher densities are usually associated with stiff structures (e.g., bronchial walls or larger vessels) whereas low densities result from lung parenchyma consisting of alveoli and capillary vessels. In fact, due to the partial volume effect almost all voxels represent a mixture of stiff and elastic structures. During breathing, air can inflate or deflate the lungs, thus leading to a change in volume of lung parenchyma indicated by a change in local density. Generally, the change is proportional to the

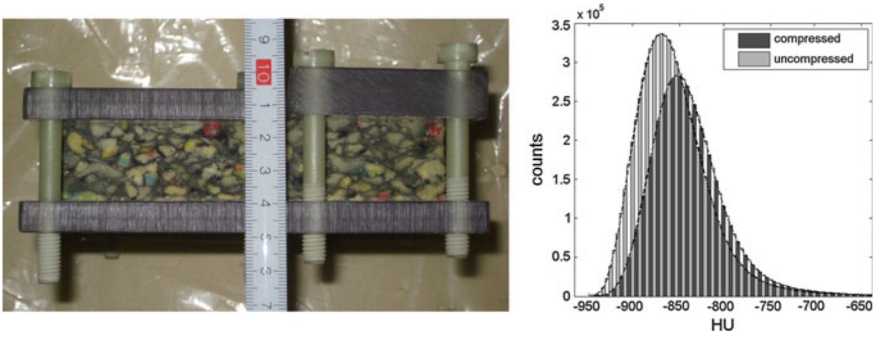


Fig. 8.9 Phantom composed of foam pieces with size ranging from about 1 mm to more than 10 mm. A sensory analysis reveals the single pieces to be of individual elasticity. The composite foam is mounted between two plates of acrylic glass with the upper plate splitted into two parts. Screws are used to fix the upper plate parts to the lower plate. On the *left*, the phantom is displayed in compressed state ("axial" view), on the *right*, intensity histograms after CT acquisition are shown

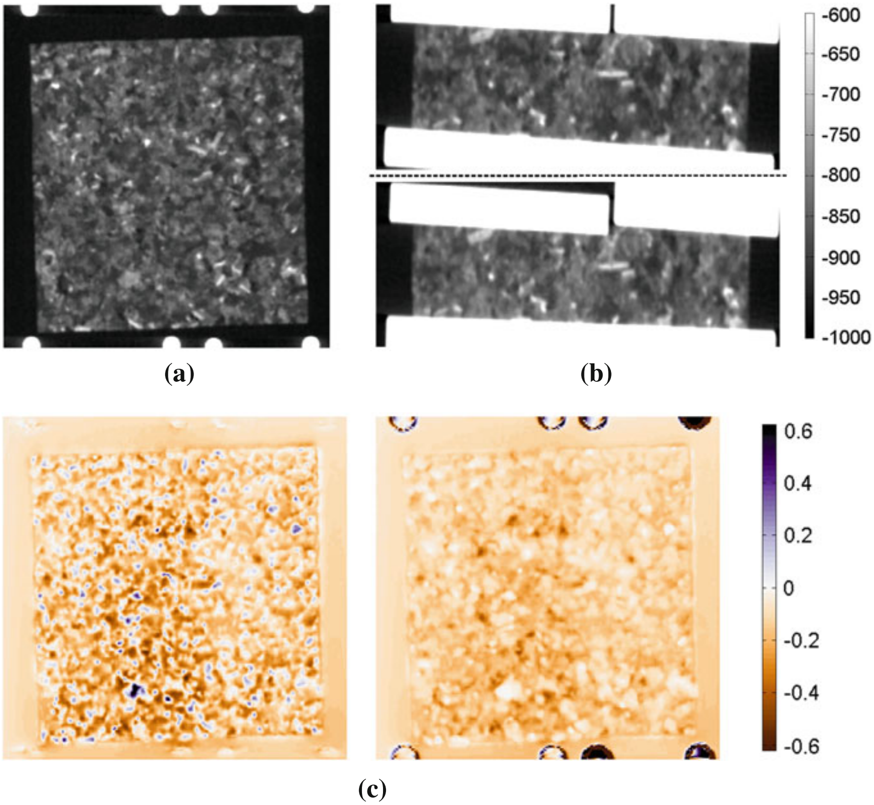
fraction of air. A change in volume of a certain structure, however, requires the structure to be elastic. Thus, elasticity is proportional to the fraction of air as well.

A recent study [52] investigated the voxel-wise calculated correlation between the Jacobian map and tissue density based entities such as the SAI measure (Eq. 13.17). However, the reported resulting correlation is not sufficient for use as a validation metric yet. Possible sources of errors include image noise, image artifacts, registration errors and inhomogeneous changes in the blood distribution during respiration. It should also be noted that lung diseases may change elasticity locally with possible impact on the relationship between expected parenchymal elasticity and air fraction. For instance, lung regions affected by emphysema or fibrosis are characterized by decreased compliance of the lung tissue [53] resulting in reduced ventilation and loss of elasticity. To rule out any possibility of impact from unknown pathologies, a compressible CT phantom with spatially varying elasticity (see Fig. 8.9) has been chosen [54] to investigate the relation between parenchymal elasticity and air fraction for functional validation.

The phantom is scanned twice ($0.33 \times 0.33 \times 0.45\,\text{mm}^3$, further details are described in [54]), once in the uncompressed state and once in a compressed state (see Fig. 8.10a, b for exemplary slices). A histogram analysis (Fig. 8.9, right) reveals that, as for the lungs, the density in the compressed state is higher than in the uncompressed state.

For illustration and comparison, two registration methods are chosen: (1) one with a spatially constant elasticity constraint and (2) one with a spatially varying elasticity constraint. The two methods result in a similar landmark accuracy (50 well-dispersed landmarks; TRE given as mean ± std (max) [mm]) of $0.28 \pm 0.13(0.72)$ and $0.33 \pm 0.14(0.76)$, respectively. The Jacobian maps (cf. Sect. 8.3.2.1) shown in Fig. 8.10c reveal mostly contracting regions but also expanding regions occuring with a spatially constant elasticity constraint as depicted by regions color-coded in blue. Recalling that the phantom was exposed to overall contraction, expanding regions would be physically unrealistic.

Registration accuracy is now functionally analysed by relating the Jacobian map voxel-wise with the HU densities. Relating these two entitites is achieved by means of joint histograms. Since the relationship depends on the applied compression level, for each level of relative compression (relC) a normalized joint histogram is computed. Finally, to combine the information from the different compression levels, from each histogram a median graph is extracted. The collection of graphs shown in an ensemble plot (Fig. 8.11) now describes the reaction of a material with a certain HU value to individually applied compression forces. The differences between the spatially constant elasticity constraint (shown left) and the spatially varying constraint (shown right) underline the impact of the elasticity constraint on the registration result: the constant elasticity setting results in a graph ensemble with implausible positive Jacobian values for voxels with higher intensity. The positive values are directly linked to the expanding regions visible in Fig. 8.10c. On the contrary, for the spatially varying constraint no part of the phantom has been expanded.



(a) **(b)**

(c)

Fig. 8.10 **a** Mid coronal slice of CT phantom in uncompressed state. **b** Mid axial slice in uncompressed (*top*) and compressed state (*bottom*). **c** Jacobian maps (same slice as for (**a**)) for spatially constant elasticity constraint (*left*) and for spatially varying elasticity constraint (*right*)

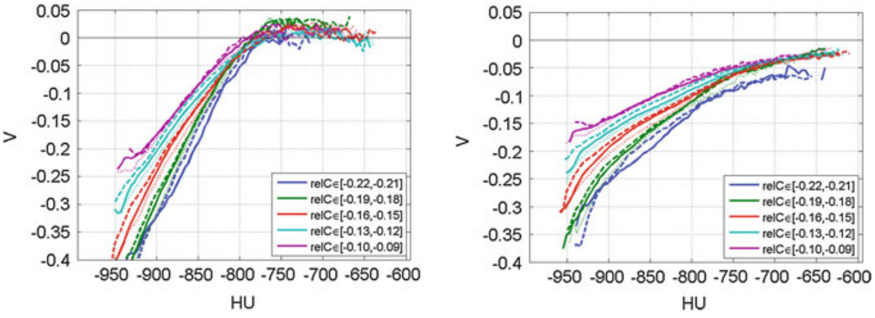


Fig. 8.11 Ensemble plots derived from normalized joint histograms (see text for explanation) for spatially constant elasticity constraint (*left*) and for spatially varying elasticity constraint (*right*)

Conclusion: It has been shown that landmark-based registration error is not sufficient to validate the deformations of an elastic body under compression. Moreover, validation based solely on this error may lead to a tuning of registration methods towards high flexibility but less physiologically plausible deformations. This experiment, which was carried out under laboratory conditions (no patient induced artifacts, high image dose) supports another recent study [3] where different registration schemes showed partially implausible contraction-expansion patterns but resulted in similar landmark-based registration errors.

Functional validation is exemplarily demonstrated by relating the Jacobian map voxel-wise with HU densities. Validation is no longer restricted to analysis of high contrast anatomical structures, but has been extended to include assessment of image regions with homogeneous intensities.

References

1. West, J., Fitzpatrick, J., Wang, M., Dawant, B., et al.: Comparison and evaluation of retrospective intermodality brain image registration techniques. J. Comput. Assist. Tomogr. **21**(4), 554–566 (1997)
2. Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys. Med. Biol. **54**, 1849–1870 (2009)
3. Kabus, S., Klinder, T., Murphy, K., van Ginneken, B., Lorenz, C., Pluim, J.P.W.: Evaluation of 4D-CT lung registration. In: MICCAI, vol. 5761, pp. 747–54 (2009)
4. Yamamoto, T., Langner, U., Loo, B.W., Shen, J., Keall, P.J.: Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients. Int. J. Radiat. Oncol. Biol. Phys. **72**(4), 1250–1258 (2008)
5. Werner, R., Ehrhardt, J., Schmidt-Richberg, A., Handels, H.: Validation and comparison of a biophysical modeling approach and non-linear registration for estimation of lung motion fields in thoracic 4D CT data. In: Reinhardt, J., Pluim, J. (eds.) Proceedings of SPIE Medical Imaging, vol. 7259, pp. 0U1–0U8 (2009)
6. Castillo, E., Castillo, R., Zhang, Y., Guerrero, T.: Compressible image registration for thoracic computed tomography images. J. Med. Biol. Eng. **29**, 222–233 (2009)
7. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Med. Image Anal. **12**, 731–741 (2008)
8. Brock, K.: Deformable registration accuracy consortium: results of a multi-institution deformable registration accuracy study (MIDRAS). Int. J. Radiat. Oncol. Biol. Phys. **76**(2), 583–596 (2010)
9. Murphy, K., van Ginneken, B., Reinhardt, J.M., et al.: Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. IEEE Trans. Med. Imaging **30**(11), 1901–1920 (2011)
10. Vandemeulebroucke, J., Sarrut, D., Clarysse, P.: The POPI-model, a point-validated pixel-based breathing thorax model. In: ICCR (2007)
11. Murphy, K., Pluim, J., van Rikxoort, E., de Jong, P., de Hoop, B., Gietema, H., Mets, O., de Bruijne, M., Lo, P., Prokop, M., van Ginneken, B.: Toward automatic regional analysis of pulmonary function using inspiration and expiration thoracic CT. Med. Phys. **39**(3), 1650–1662 (2012)
12. Sarrut, D., Delhay, B., Villard, P.F., Boldea, V., Beuve, M., Clarysse, P.: A comparison framework for breathing motion estimation methods from 4-D imaging. IEEE Trans. Med. Imaging **26**(12), 1636–1648 (2007)

13. Werner, R., Ehrhardt, J., Schmidt-Richberg, A., Heiss, A., Handels, H.: Estimation of motion fields by non-linear registration for local lung motion analysis in 4D CT image data. Int. J. Comput. Assist. Radiol. Surg. **5**(6), 595–605 (2010)
14. Murphy, K., van Ginneken, B., Klein, S., Staring, M., de Hoop, B.J., Viergever, M.A., Pluim, J.P.W.: Semi-automatic construction of reference standards for evaluation of image registration. Med. Image Anal. **15**(1), 71–84 (2011)
15. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.: Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging **29**(1), 196–205 (2010)
16. Hartkens, T., Rohr, K., Stiehl, H.S.: Evaluation of 3D operators for the detection of anatomical point landmarks in MR and CT images. Comput. Vis. Image Underst. **86**, 118–136 (2002)
17. Werner, R., Wolf, J.C., Ehrhardt, J., Schmidt-Richberg, A., Handels, H.: Automatische Landmarkendetektion und -übertragung zur Evaluation der Registrierung von thorakalen CT-Daten. In: Deserno, T., Handels, H., Meinzer, H., Tolxdorff, T. (eds.) Bildverarbeitung für die Medizin 2010, Informatik aktuell, pp. 31–35. Springer, Heidelberg (2010)
18. Berlinger, K., Roth, M., Sauer, O., Vences, L., Schweikard, A.: Fully automatic detection of corresponding anatomical landmarks in volume scans of different respiratory state. Med. Phys. **33**(6), 1569–1572 (2006)
19. Likar, B., Pernus, F.: Automatic extraction of corresponding points for the registration of medical images. Med. Phys. **26**(8), 1678–1686 (1999)
20. Vik, T., Kabus, S., von Berg, J., Ens, K., Dries, S., Klinder, T., Lorenz, C.: Validation and comparison of registration methods for free-breathing 4D lung CT. In: Proceedings of SPIE Medical Imaging, vol. 6914, pp. 69,142P–1–69,142P–10 (2008)
21. Wan, S.Y., Higgins, W.: Symmetric region growing. IEEE Trans. Image Process. **12**(9), 1007–1015 (2003)
22. Schlathölter, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. In: Proceedings of SPIE Medical Imaging, vol. 2, pp. 103–113 (2002)
23. Schaap, M., Smal, I., Metz, C., van Walsum, T., Niessen, W.: Bayesian tracking of elongated structures in 3d images. In: Proceedings of Information Processing in Medical Imaging (IPMI), vol. 4584, pp. 74–85 (2007)
24. Friman, O., Hindennach, M., Kühnel, C., Peitgen, H.O.: Multiple hypothesis template tracking of small 3D vessel structures. Med. Image Anal. **14**(2), 160–171 (2010)
25. Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. Med. Image Anal. **13**, 819–845 (2009)
26. Lo, P., van Ginneken, B., Reinhardt, J., de Bruijne, M.: Extraction of airways from CT (EXACT'09). IEEE Trans. Med. Imaging **31**(11), 2093–2107 (2012)
27. Tschirren, J., McLennan, G., Palagyi, K., Hoffman, E., Sonka, M.: Matching and anatomical labeling of human airway tree. IEEE Trans. Med. Imaging **24**(12), 1540–1547 (2005)
28. Bülow, T., Lorenz, C., Wiemker, R., Honko, J.: Point based methods for automatic bronchial tree matching and labeling. In: Proceedings of SPIE Medical Imaging, vol. 6143 (2006)
29. Bouix, S., Martin-Fernandez, M., Ungar, L., Nakamura, M., Koo, M.S., McCarley, R.W., Shenton, M.E.: On evaluating brain tissue classifiers without a ground truth. NeuroImage **36**(4), 1207–1224 (2007)
30. Harris, E.J., Donovan, E.M., Yarnold, J.R., Coles, C.E., Evans, P.M.: Characterization of target volume changes during breast radiotherapy using implanted fiducial markers and portal imaging. Int. J. Radiat. Oncol. Biol. Phys. **73**(3), 958–966 (2009)
31. Hanna, G.G.: Hounsell, a.R., O'Sullivan, J.M.: Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods. Clin. Oncol. (R. Coll. Radiol.) **22**(7), 515–525 (2010)
32. Crum, W., Camara, O., Hill, D.: Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans. Med. Imaging **25**(11), 1451–1461 (2006)
33. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297 (1945)

34. Louie, A.V., Rodrigues, G., Olsthoorn, J., Palma, D., Yu, E., Yaremko, B., Ahmad, B., Aivas, I., Gaede, S.: Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era. Radiother. Oncol. **95**(2), 166–171 (2010)
35. Jaccard, P.: Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin del la Société Vaudoise des Sciences Naturelles **37**, 547–579 (1901)
36. Tralins, K.S., Douglas, J.G., Stelzer, K.J., Mankoff, D.A., Silbergeld, D.L., Rostomily, R.C., Hummel, S., Scharnhorst, J., Krohn, K.A., Spence, A.M., Rostomilly, R.: Volumetric analysis of 18F-FDG PET in glioblastoma multiforme: prognostic information and possible role in definition of target volumes in radiation dose escalation. J. Nucl. Med. **43**(12), 1667–1673 (2002)
37. Hu, S., Hoffman, E., Reinhardt, J.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans. Med. Imaging **20**(6), 490–499 (2001)
38. van Rikxoort, E., de Hoop, B., Viergever, M., Prokop, M., van Ginneken, B.: Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. Med. Phys. **36**(7), 2934–2947 (2009)
39. van Rikxoort, E., van Ginneken, B., Klik, M., Prokop, M.: Supervised enhancement filters: application to fissure detection in chest CT scans. IEEE Trans. Med. Imaging **27**(1), 1–10 (2008)
40. Ukil, S., Reinhardt, J.: Anatomy-guided lung lobe segmentation in X-ray CT images. IEEE Trans. Med. Imaging **28**(2), 202–214 (2009)
41. van Rikxoort, E., Prokop, M., de Hoop, B., Viergever, M., Pluim, J.P.W., van Ginneken, B.: Automatic segmentation of pulmonary lobes robust against incomplete fissures. IEEE Trans. Med. Imaging **29**(6), 1286–1296 (2010)
42. Lassen, B., Kuhnigk, J.M., van Rikxoort, E., Peitgen, H.O.: Interactive lung lobe segmentation and correction in tomographic images. In: Proceedings of SPIE Medical Imaging, vol. 79631S, pp. 79, 631S–1–11 (2011)
43. Reeves, A., Chan, A., Yankelevitz, D., Henschke, C., Kressler, B., Kostis, W.: On measuring the change in size of pulmonary nodules. IEEE Trans. Med. Imaging **25**(4), 435–450 (2006)
44. Lujan, A., Larsen, E., Balter, J., Ten Haken, R.: A method for incorporating organ motion due to breathing into 3D dose calculations. Med. Phys. **26**(5), 715–720 (1999)
45. Castillo, E., Castillo, R., Martinez, J., Shenoy, M., Guerrero, T.: Four-dimensional deformable image registration using trajectory modeling. Phys. Med. Biol. **55**(1), 305–327 (2010)
46. Vandemeulebroucke, J., Rit, S., Kybic, J., Clarysse, P., Sarrut, D.: Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. Med. Phys. **38**(1), 166 (2011)
47. Vandemeulebroucke, J., Bernard, O., Rit, S., Kybic, J., Clarysse, P., Sarrut, D.: Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT. Med. Phys. **39**(2), 1006 (2012)
48. Christensen, G., Johnson, H.: Consistent image registration. IEEE Trans. Med. Imaging **20**(7), 568–582 (2001)
49. Kabus, S., Lorenz, C.: Fast elastic image registration. In: Grand Challenges in Medical Image Analysis, pp. 81–89 (2010)
50. Ding, K., Cao, K., Amelon, R., Christensen, G., Raghavan, M., Reinhardt, J.: Comparison of intensity- and jacobian-based estimates of lung regional ventilation. In: Proceedings of Third International Workshop on Pulmonary Image Analysis, pp. 49–60 (2010)
51. West, J.: Respiratory Physiology: The Essentials. Lippincott Williams & Wilkins, Philadelphia (2008)
52. Yamamoto, T., Kabus, S., Klinder, T., Lorenz, C., Loo, B., Keall, P.: Four-dimensional computed tomography pulmonary ventilation images vary with deformable image registration algorithms and metrics. Med. Phys. **38**(3), 1348–1358 (2011)
53. Pratt, P.: Pulmonary Pathology, chap. Emphysema and Chronic Airways Disease, pp. 651–69. Springer, New York (1988)
54. Kabus, S., Klinder, T., von Berg, J., Lorenz, C.: Functional non-rigid registration validation: a CT phantom study. In: Fischer, B., Dawant, B., Lorenz, C. (eds.) WBIR, Springer, vol. 6204, pp. 116–127 (2010)