

# Assessing Accuracy of Non-linear Registration in 4D Image Data using Automatically Detected Landmark Correspondences

René Werner<sup>1</sup>, Christine Duscha<sup>2</sup>, Alexander Schmidt-Richberg<sup>2</sup>,  
Jan Ehrhardt<sup>2</sup>, and Heinz Handels<sup>2</sup>

<sup>1</sup> Institute of Computational Neuroscience, University Medical Center Hamburg-Eppendorf,  
Hamburg, Germany

<sup>2</sup> Institute of Medical Informatics, University of Lübeck, Lübeck, Germany

## ABSTRACT

4D imaging becomes increasingly important in clinical practice. Its use in diagnostics and therapy planning usually requires the application of non-linear registration techniques. The reliability of information derived from the computed transformations is directly dependent on the registration accuracy. Ideally, this accuracy should be evaluated on a patient- and data-specific level – which requires appropriate evaluation criteria and procedures. A standard approach for evaluation of non-linear registration accuracy is to compute a landmark- or point-based registration error by means of manually detected landmark correspondences in the images to register, with the landmarks being anatomically characteristic points. Manual detection of such points is, however, time-consuming and error-prone. In this contribution, different operators for automatic landmark detection and a block matching strategy for landmark propagation in 4D image sequences (here: 4D lung CT, 4D liver MRT) are proposed and evaluated. It turns out that the so-called Förstner-Rohr operators perform best for detection of anatomically characteristic points and that the proposed propagation strategy ensures a robust transfer of these landmarks between the images. The automatically detected landmark correspondences are then used to evaluate the accuracy of different registration approaches (in total 48 variants) applied for registering 4D lung CT data. The resulting registration error values are compared to errors obtained by manually detected landmark pairs. It is shown that derived statements concerning differences in accuracy of the registration approaches are identical for both the manually and the automatically detected landmark sets.

**Keywords:** 4D imaging, registration, evaluation, landmark detection

## 1. PURPOSE

4D imaging becomes increasingly important in clinical practice, for instance in radiation therapy when managing respiratory motion.<sup>1,2</sup> For this clinical scenario, a 4D image sequence is usually considered to be a series of 3D images representing the patients' anatomy at different breathing states. During treatment planning, the temporal image information is then used for dimensioning safety margins and analyzing motion effects on dose distributions by taking into account the patients' respiratory motion patterns. Such steps require the use of non-linear registration for estimation of the motion fields between the 3D (spatial) images of the 4D image sequence, which eventually form the basis of, e. g., 4D contouring<sup>3,4</sup> and dose accumulation techniques.<sup>5-7</sup>

It is obvious that the reliability of extracted information – in this case safety margins and 4D dose distributions – is directly tied to the accuracy of the registration applied;<sup>8</sup> therefore, an accuracy assessment of the registration is indispensable, ideally performed on a patient- and data-specific level. This, however, is difficult because usually no ground truth deformations are known that could be compared with the registration results.<sup>9</sup> As a widely used and accepted approach to solving this problem, a landmark- or point-based registration error, often also referred to as target registration error (TRE), is computed and considered as quantitative evaluation criterion,<sup>8,10-13</sup> with the landmarks being anatomically characteristic points in the images to be registered (e. g. bifurcations of vessel trees of organs of interest).

---

Further author information: (Send correspondence to René Werner; E-mail: r.werner@uke.de)

Such points are mostly manually identified, but this results in a time-consuming and error-prone process concerning variability/reproducibility of exact point selection,<sup>12</sup> and therefore efforts have been made to (partially) automate the identification of appropriate landmark sets. Important steps toward automatic landmark correspondence detection for evaluation of non-linear registration have been taken, e. g., by Murphy *et al.*,<sup>9,14</sup> exploiting earlier ideas of Likar and Pernus<sup>15</sup> for constructing a semi-automatic reference standard. However, the use of an entirely automatic technique for defining landmark correspondences for quantitative evaluation of non-linear registration is still discussed controversially.<sup>11</sup> Placed in this context, the contribution of this paper is twofold: First, we discuss and evaluate the use of different operators and approaches for automatic landmark detection in 3D lung CT and 3D liver MRT images, which could be used for either a semi-automatic or an automatic reference standard construction. Aiming at an entirely automatic approach, we then apply a block matching strategy for propagating the detected landmarks between images to be registered. As the second part of the contribution, automatically established landmark correspondences are used to evaluate the accuracy of different non-linear intensity-based registration approaches when applied for registering 4D lung CT data. Derived results/statements regarding differences in accuracy of the registration approaches are compared with corresponding numbers obtained by applying manually detected landmark sets for evaluation purposes.

## 2. METHODS

According to the main contributions of the paper, this section is structured as follows: In Sec. 2.1, the operators considered for landmark detection are defined (Sec. 2.1.1), and the block matching strategy used for propagating the landmarks between the images to be registered is explained (Sec. 2.1.2); the data sets and the strategy used for evaluation of the landmark detection and propagation are detailed in Sec. 2.1.3. The registration approaches applied for registering the 4D lung CT data and the strategy applied for comparison of the landmark-based registration errors based on automatically detected (TRE-a) and manually detected landmarks (TRE-m) are finally described in Sec. 2.2.

### 2.1 Automatic definition of landmark correspondences

#### 2.1.1 Operators applied for landmark detection

Within the widely used approach for semi-automatic landmark-based reference standard construction of Murphy *et al.*,<sup>9,14</sup> automatic landmark detection is based on a so-called distinctiveness term that is assumed to be high for voxels representing anatomically characteristic points. The distinctiveness term consists of two multiplicative parts: A normalized feature-based term and a dissimilarity term of the intensity values of a local neighborhood of voxels around the voxel considered and the neighborhood of the voxel itself. Let  $I_R : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$  denote the image considered for landmark detection and  $\nabla I_R : \Omega \rightarrow \mathbb{R}^3$  the corresponding gradient image; then, the proposed distinctiveness term of a voxel  $x \in \Omega$  reads as<sup>14</sup>

$$D(x) = \begin{cases} 0 & \text{if } \|\nabla I(x)\| < \theta_{\|\nabla I(x)\|} \\ \underbrace{\frac{\|\nabla I(x)\|}{\max_{x' \in \Omega} \|\nabla I(x')\|}}_{\text{feature-based term}} \cdot \underbrace{\sum_i \text{MSD}(I(N(x)), I(N(x_i)))}_{\text{intensity-based dissimilarity term}} & \text{else} \end{cases}$$

with the  $x_i \in \Omega$  being voxels in the proximity of  $x$  (here: on a surface of a sphere centered in  $x$ ) and  $N(x)$  denoting a neighborhood of  $x$  (here: a sphere around  $x$ ); the MSD term refers to the mean squared intensity differences of the voxels in  $N(x)$  and  $N(x_i)$ .

In the above definition, the image gradient magnitude is considered as the relevant image feature characterizing appropriate landmark candidates. However, potential anatomically characteristic points like bifurcations of, e. g., the vessel trees of lung and/or liver or the bronchial tree inside the lung feature specific curvature characteristics.<sup>16</sup> Reviving earlier works on landmark detection,<sup>17,18</sup> we suggested the so-called (3D-)Förstner operator instead.<sup>19,20</sup> In this paper, we extend our previous studies by evaluating and comparing the landmark

detection performance of the different Förstner-Rohr-type operators; the image gradient magnitude serves as the standard of comparison. Retaining the notation of Hartkens *et al.*,<sup>17</sup> these operators are defined as

$$\text{Op3} = \frac{\det \bar{C}_{I_R}}{\text{tr} \bar{C}_{I_R}}, \quad (1)$$

$$\text{Rohr3D} = \det \bar{C}_{I_R}, \quad (2)$$

$$\text{Förstner3D} = \frac{\det \bar{C}_{I_R}}{\text{tr} \bar{C}_{I_R}^{\text{adj}}} \quad (3)$$

with  $\bar{C}_{I_R} : \Omega \rightarrow \mathbb{R}^{3 \times 3}$  as a smoothed version of the structure tensor  $\nabla I_R \nabla I_R^T$ , computed by  $\bar{C}_{I_R} = K_\rho * (\nabla I_R \nabla I_R^T)$  with  $K_\rho$  being a Gaussian kernel with standard deviation  $\rho$ . To further suppress unintended high answers of the Förstner-Rohr-type operators at non-characteristic points, we additionally apply the operator

$$\text{OpR} = \frac{\prod_i \lambda_i}{\left(\frac{1}{3} \sum_i \lambda_i\right)^3} \quad (4)$$

( $\lambda_i$ : eigenvalues of  $\bar{C}_{I_R}$ ) in the sense of a rejection operator as originally proposed by Rohr,<sup>18</sup> i.e. we only consider answers of the operators (1)-(3) as denoting appropriate landmark candidates if the OpR values of the corresponding voxels lie above a pre-defined threshold. This allows, e.g., to avoid detection of points on edges (i.e. one eigenvalue  $\lambda_i \approx 0$ ).

For landmark detection, the Förstner-Rohr-type operators (filtered by OpR) and the image gradient magnitude  $\|I_R\|$  are evaluated in a specific region of interest (ROI) of  $I_R$ . Suitable landmark candidates are selected based on the values of the operator answers (high values = suitable candidates), starting with highest value as the first landmark candidate. Additional landmark candidates are added analogously to the candidate set, but a minimum Euclidean distance is postulated to be kept between the candidates to be selected<sup>9,14</sup> to provide an approximately equal distribution of the landmarks in the ROI as required for application to registration evaluation.<sup>11</sup>

### 2.1.2 Landmark propagation between images

After detecting a set  $\mathbb{L}_R$  of landmarks in the image  $I_R$ , which is subsequently assumed to be one of the 3D frames of a 4D image sequence, the corresponding landmark sets  $\mathbb{L}_{T_i}$  of the other images  $I_{T_i} : \Omega \rightarrow \mathbb{R}$  of the image sequence ( $i = 1, \dots, N$ : breathing phases represented by the 3D images) are identified by a cross correlation (CC)-based block matching strategy detailed in Werner *et al.*<sup>20</sup> and Ehrhardt *et al.*<sup>19</sup> Let  $I_T$  be a particular image to transfer  $\mathbb{L}_R$  to, two block matching runs are performed: a block matching using the original intensity information of  $I_R$  and  $I_T$ , and a block matching based on the answers of the specific operator applied for the original landmark detection in  $I_R$ . The results of both runs are evaluated and combined to improve robustness of landmark propagation: The propagation of a landmark candidate in  $\mathbb{L}_R$  is only accepted if the estimated target landmark positions of both block matching runs agree and if the correlation coefficient of the intensity values in a local neighborhood of the landmark positions in  $I_R$  and  $I_T$  is larger than a pre-defined correlation value; otherwise, the landmark propagation is rated as not reliable and the landmark candidate is not considered for subsequent evaluation purposes.

### 2.1.3 Data sets and experiments

For evaluation of the detection performance of the operators when applied to lung CT data, we used the POPI phantom<sup>21</sup> (image resolution:  $0.98 \times 0.98 \times 2 \text{ mm}^3$ ; mid-inspiration phase), one data set of the DIR-Lab data pool<sup>11</sup> ( $0.97 \times 0.97 \times 2.5 \text{ mm}^3$ ; mid-inspiration phase), and two diagnostic lung CT images of high spatial resolution ( $0.76 \times 0.76 \times 0.7 \text{ mm}^3$ ) of our own fund. Within each image and for each operator, 50 inner lung landmarks were automatically detected. The operator performance was then evaluated independently by three observers by classifying the individual landmarks into the categories “*definitely / potentially / not an anatomically characteristic point*”. A similar evaluation strategy has been used for landmark detection in a mid-inspiration liver MRT data set<sup>22</sup> with the liver as ROI, but only 20 landmarks were identified in this case due to low image resolution and size ( $1.37 \times 1.37 \times 4 \text{ mm}^3 / 166 \times 195 \times 25 \text{ voxel}$ ).

For the POPI- and DIR-lab data, the landmarks detected by the different operators were then propagated to the corresponding end-inspiration phase and the propagation was evaluated visually on a landmark-by-landmark basis (“*propagation successful / not successful*”).

## 2.2 Comparison of TRE-a and TRE-m as evaluation criteria: study design

With the optimal operator (according to the results of the experiments of Sec. 2.1.3) as the basis of the automatic landmark detection (here: the feature-based term of the distinctiveness value computation according to Murphy *et al.*<sup>9,14</sup>), we performed an in-depth comparison of the TRE-a and the TRE-m as evaluation criteria for registration in 4D lung CT data. Therefore, end-inspiration and end-expiration data of in total 23 4D CT data sets were registered (POPI, 10 DIR-Lab data sets, 12 4D CT sequences of our own fund) using different variants of a variational registration scheme. The POPI and the DIR-lab data come along with their own set of manually detected landmark correspondences, provided through the respective platforms. For our own data sets we identified 70 landmark pairs manually for each 4D image sequence to calculate the TRE-m values. For computation of the TRE-a values, we used approx. 150 automatically detected landmarks per data set (detected in the end-inspiration image, then propagated to the end-expiration image). TRE-a and TRE-m values obtained by the different registration variants were compared considering corresponding results for the individual registration approaches; statistical significance of differences between TRE-a and TRE-m values for particular approaches was tested by a paired t-test on the patient-specific mean TRE-m and TRE-a values (mean over all landmarks for the individual patient/4D CT data set).

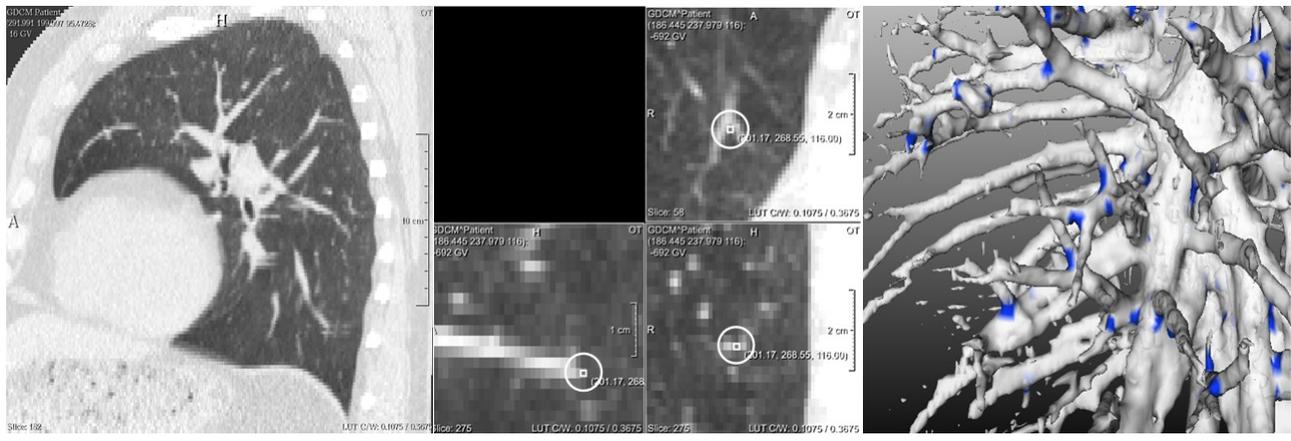
The variational framework used for registration of two CT frames  $I_R$  (reference image) and  $I_T$  (target image) reads as,

$$\mathcal{J}[\varphi] = \mathcal{D}[R, T, \varphi] + \alpha \mathcal{S}[\varphi] \stackrel{\varphi}{\rightarrow} \min, \quad (5)$$

with the transformation  $\varphi : \Omega \rightarrow \Omega$  being sought in such a way that it minimizes a dissimilarity measure  $\mathcal{D} : \text{Img}(\Omega) \times \text{Img}(\Omega) \times \mathcal{C}^2(\Omega, \Omega) \rightarrow \mathbb{R}_0^+$  between  $I_R$  and  $(I_T \circ \varphi)$ , but still features a desired smoothness as specified by the regularization term  $\mathcal{S} : \mathcal{C}^2(\Omega, \Omega) \rightarrow \mathbb{R}_0^+$ ;  $\text{Img}(\Omega)$  denotes the set of all images with domain  $\Omega$ . For this study, we applied the following building blocks of (5):

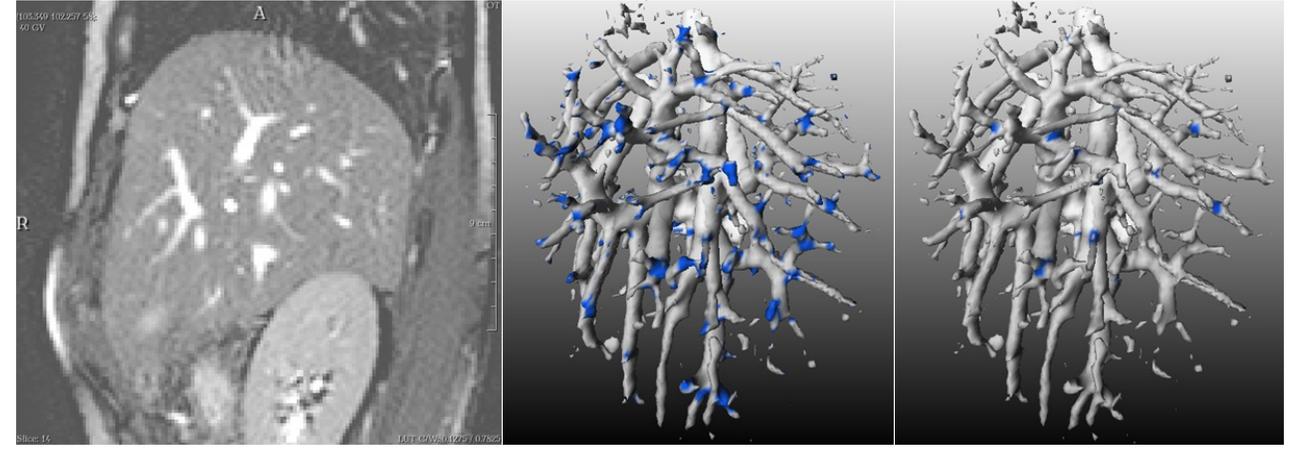
- **Transformation space:** We considered both an optimization over all standard  $\mathcal{C}^2$ -fields (see, e.g., Werner *et al.*<sup>13</sup>) and a minimization with the transformations  $\varphi$  being restricted to diffeomorphisms parameterized by stationary velocity fields (see Schmidt-Richberg *et al.*<sup>23</sup> for details). Corresponding registration schemes are subsequently denoted as *standard* and *diffeomorphic* registration schemes.
- **Distance measure / related force term:** As distance measures we applied the standard sum-of-squared intensity-differences measure (*SSD*) and three variants of *Thirion-type terms* – active, passive, dual terms; depending on the force domain considered; for details see, e. g., Schmidt-Richberg *et al.*<sup>24</sup> All variants are evaluated in both a masked and an unmasked manner (i. e. evaluation of intensity differences/forces only inside the lung vs. for the entire image domain). We further considered a symmetrization of force computation, i. e. the evaluation of  $\mathcal{D}_{\text{sym}}[R, T, \varphi] = 1/2 (\mathcal{D}[R, T, \varphi] + \mathcal{D}[T, R, \varphi^{-1}])$  instead of only  $\mathcal{D}[R, T, \varphi]$ . This, however, requires an efficient computation of the inverse transformation  $\varphi^{-1}$  – which is an advantage of the applied diffeomorphic registration scheme. We therefore evaluated the symmetrization of the force computation only within the diffeomorphic registration scheme and refer to the corresponding registration methods as *symmetric-diffeomorphic* registration scheme.
- **Regularization terms:** For regularization we applied a *diffusion* and a *linear-elastic* smoothing term.<sup>25</sup>

Thus, in total, 48 variants of (5) were considered for evaluation and comparison of the TRE-m and TRE-a values [(standard, diffeomorphic, symmetric-diffeomorphic) × (SSD-, 3 × Thirion-type forces) × (masked/unmasked force computation) × (diffusion/elastic regularization)].



(a) POPI phantom, sagittal view (b) Example of an “anatomically characteristic point” in the lung (c) Landmark candidates, detected by the Op3 operator

Figure 1. Example of an anatomically characteristic point (middle) and illustration of inner lung landmark candidates detected by the Op3 operator within the POPI phantom (right figure; landmark candidates = centers of the blue areas; figures in color in electronic version).



(a) Sagittal view of a liver MRT data set (b) Operator answer, filtered by the OpR operator (c) Positions of the landmarks finally detected

Figure 2. Illustration of landmark detection (here: Op3 as detection operator) in a liver MRT data set (final landmarks = centers of the blue areas in the right figure; figures in color only in electronic version).

### 3. RESULTS

#### 3.1 Comparison of operators used for landmark detection

An example of a landmark rated as an “anatomically characteristic point” of the lung can be found in Fig. 1; the areas of appropriate landmark candidates as detected using the operator *Op3* are also visualized. It can be seen that the landmarks detected in lung CT data are mainly bifurcations of the bronchial tree – and therefore represent a natural choice of anatomically characteristic points. This also holds for the operators *Rohr3D* and *Förstner3D*; in total  $84 \pm 10\%$  of the lung points detected by these operators were rated as being “definitely anatomically characteristic” by the three observers (“potentially anatomically characteristic”:  $10 \pm 7\% / 12 \pm 10\% / 10 \pm 7\%$  for *Op3/Rohr3D/Förstner3D*; “not anatomically characteristic”:  $6 \pm 4\% / 5 \pm 3\% / 6 \pm 5\%$ ). In contrast, only  $50 \pm 8\%$  ( $29 \pm 8\%$ ;  $21 \pm 6\%$ ) of the points were classified as “definitely (potentially; not) anatomically characteristic” if using the intensity magnitude as detection operator.

Similar statements hold for the landmark detection in the liver MRT data set (cf. Fig. 2), but with a lower

number of detected points featuring obvious anatomical characteristics. Interestingly, for this data set the 20 landmarks detected by the Förstner-Rohr-type operators were identical, with the fraction of points rated as “definitely anatomically characteristic” being  $73\pm 10\%$  (“potentially anatomically characteristic”:  $25\pm 13\%$ ; “not anatomically characteristic”:  $2\pm 3\%$ ); in contrast, the detection by the image gradient magnitude resulted in a fraction of only  $35\pm 13\%$  points being considered as definitely anatomically characteristic.

Thus, no obvious differences are observed between the different Förstner-Rohr-type operators when applied for landmark detection, but all operators (1)-(3) yield superior results when compared to the image gradient magnitude as detection operator. For the subsequent comparison of TRE-m and TRE-a values we therefore applied the Op3 operator for landmark detection (here: for computation of the feature-based term of a distinctiveness value according to Murphy *et al.*; cf. Sec. 2.1.1).

### 3.2 Evaluation of the landmark propagation

Similar to our earlier studies,<sup>20</sup> the propagation of the detected landmarks by the proposed block matching strategy turned out to be very robust; when using the consistency checks as described in Sec. 2.1.2 and a correlation coefficient threshold of 0.9, all automatically established landmark correspondences were evaluated as being successful. However, high thresholds led to a rejection of a larger number of detected landmark candidates, and so – in order to obtain a final landmark set of fair size and almost equal distribution in the ROI – the correlation threshold was decreased to 0.8 for subsequent experiments, accepting that some landmark correspondences used for calculation of the TRE-a were established erroneously.

### 3.3 Comparison of TRE-a and TRE-m

The TRE-m and TRE-a values (Op3 as basis of automatic landmark detection; cf. Sec. 3.1) as obtained for the different variants of the registration scheme (5) are summarized in Tab. 1. No significant differences were observed between TRE-m and TRE-a values of all variants of the applied registration scheme (mean difference between TRE-m and TRE-a values:  $0.02\pm 0.10$  mm;  $p=0.12$ ) – except for the variants with standard SSD as distance measure. For these registration schemes, the TRE-a is significantly lower than the TRE-m (mean difference:  $0.6\pm 0.2$  mm;  $p<0.001$ ), which may be due to the close relationship of CC and SSD.<sup>26</sup>

However, using TRE-m and TRE-a values for comparison of different registration strategies, general statements as well as the assessment of statistical significance of differences in registration accuracy due to different choices of the individual building blocks of the registration scheme turned out to be identical. For the selected parameters and data sets, TRE values were on average almost identical for the elastic registration variants when compared to corresponding approaches with diffusion smoothing (mean differences  $0.1\pm 0.2$  mm for both TRE-a and TRE-m); differences between the three Thirion-type force variants were in the same order for both TRE-m and TRE-a ( $p$ -values  $> 0.6$  for pairwise comparison of the force terms). Further, TRE values were observed to be slightly, but significantly worse for the symmetric-diffeomorphic schemes than for the standard or the diffeomorphic schemes (differences 0.1 mm,  $p$ -values  $< 0.05$  for both manually and automatically detected landmarks). Finally, statements that are obvious when considering Tab. 1 (Thirion-type measures better than standard SSD, masked registration more accurate than unmasked registration) also agreed when considering TRE-m and TRE-a as evaluation criteria.

## 4. DISCUSSION AND CONCLUSIONS

Reviving earlier work on landmark detection we showed that the Förstner-Rohr-type operators are well suited for automatic landmark detection in lung CT and liver MRT data – especially when compared with only the image gradient magnitude as detection operator or even the combination of both the normalized image gradient magnitude and an intensity-based local dissimilarity measure as proposed by Murphy *et al.*<sup>9,14</sup> (Werner *et al.*<sup>20</sup> 60% of the points detected by the combined terms rated as “anatomically characteristic”; study based on solely lung CT images). Some open questions still remain: Applying the Förstner-Rohr-type operators, bifurcations of large bronchi or vessels are not detected as landmark candidates, and localization problems well known for these operators<sup>27</sup> were also observed. Studying more sophisticated versions/implementations of the structure tensor and the operators<sup>28</sup> and/or integrating, e. g., multi-scale detection strategies will be part of our future work.

Table 1. TRE-m and TRE-a values for the registration approaches considered (Th = Thirion-type term; A = active, P = passive, D = dual). Listed are mean values and standard deviations of the patient-specific mean values; numbers in mm.

| regularizer +<br>dissimilarity term                              | Standard<br>reg. scheme |               | diffeomorphic<br>reg. scheme |               | sym.-diffeomorphic<br>reg. scheme |               |
|--|-------------------------|---------------|------------------------------|---------------|-----------------------------------|---------------|
|  | TRE-m                   | TRE-a         | TRE-m                        | TRE-a         | TRE-m                             | TRE-a         |
| — masked registration (i.e., evaluation only inside lung mask) — |                         |               |                              |               |                                   |               |
| $\mathcal{S}^{\text{diff}} + \mathcal{D}^{\text{SSD}}$           | $2.0 \pm 1.0$           | $1.7 \pm 0.8$ | $2.0 \pm 1.0$                | $1.6 \pm 0.7$ | $2.1 \pm 1.2$                     | $1.7 \pm 0.7$ |
| $\mathcal{S}^{\text{diff}} + \mathcal{D}^{\text{Th,P}}$          | $1.2 \pm 0.2$           | $1.4 \pm 0.4$ | $1.3 \pm 0.2$                | $1.4 \pm 0.3$ | $1.6 \pm 0.7$                     | $1.5 \pm 0.6$ |
| $\mathcal{S}^{\text{diff}} + \mathcal{D}^{\text{Th,A}}$          | $1.3 \pm 0.2$           | $1.3 \pm 0.3$ | $1.3 \pm 0.3$                | $1.3 \pm 0.3$ | $1.6 \pm 0.7$                     | $1.6 \pm 0.7$ |
| $\mathcal{S}^{\text{diff}} + \mathcal{D}^{\text{Th,D}}$          | $1.2 \pm 0.2$           | $1.3 \pm 0.3$ | $1.2 \pm 0.2$                | $1.3 \pm 0.3$ | $1.6 \pm 0.7$                     | $1.5 \pm 0.6$ |
| $\mathcal{S}^{\text{elas}} + \mathcal{D}^{\text{SSD}}$           | $2.7 \pm 1.2$           | $2.4 \pm 1.0$ | $2.8 \pm 1.3$                | $2.4 \pm 1.0$ | $2.8 \pm 1.3$                     | $2.4 \pm 1.0$ |
| $\mathcal{S}^{\text{elas}} + \mathcal{D}^{\text{Th,P}}$          | $1.2 \pm 0.2$           | $1.4 \pm 0.4$ | $1.3 \pm 0.2$                | $1.4 \pm 0.4$ | $1.4 \pm 0.3$                     | $1.5 \pm 0.5$ |
| $\mathcal{S}^{\text{elas}} + \mathcal{D}^{\text{Th,A}}$          | $1.2 \pm 0.1$           | $1.3 \pm 0.3$ | $1.3 \pm 0.2$                | $1.3 \pm 0.3$ | $1.4 \pm 0.4$                     | $1.4 \pm 0.4$ |
| $\mathcal{S}^{\text{elas}} + \mathcal{D}^{\text{Th,D}}$          | $1.2 \pm 0.1$           | $1.3 \pm 0.3$ | $1.2 \pm 0.2$                | $1.3 \pm 0.3$ | $1.3 \pm 0.2$                     | $1.4 \pm 0.4$ |
| — unmasked registration —  |                         |               |                              |               |                                   |               |
| $\mathcal{S}^{\text{diff}} + \mathcal{D}^{\text{SSD}}$           | $3.6 \pm 1.9$           | $2.7 \pm 1.4$ | $3.1 \pm 1.5$                | $2.3 \pm 1.0$ | $3.6 \pm 1.6$                     | $2.6 \pm 1.2$ |
| $\mathcal{S}^{\text{diff}} + \mathcal{D}^{\text{Th,P}}$          | $2.0 \pm 1.0$           | $1.9 \pm 1.0$ | $2.1 \pm 1.0$                | $2.0 \pm 1.0$ | $2.1 \pm 1.0$                     | $1.9 \pm 0.9$ |
| $\mathcal{S}^{\text{diff}} + \mathcal{D}^{\text{Th,A}}$          | $2.0 \pm 1.0$           | $1.9 \pm 0.9$ | $2.0 \pm 1.0$                | $1.9 \pm 0.9$ | $2.0 \pm 1.0$                     | $1.9 \pm 1.0$ |
| $\mathcal{S}^{\text{diff}} + \mathcal{D}^{\text{Th,D}}$          | $2.0 \pm 1.0$           | $1.8 \pm 0.9$ | $1.9 \pm 1.0$                | $1.8 \pm 0.9$ | $2.0 \pm 1.0$                     | $1.9 \pm 1.0$ |
| $\mathcal{S}^{\text{elas}} + \mathcal{D}^{\text{SSD}}$           | $3.8 \pm 1.7$           | $3.0 \pm 1.3$ | $3.8 \pm 1.7$                | $3.0 \pm 1.4$ | $3.8 \pm 1.7$                     | $3.1 \pm 1.4$ |
| $\mathcal{S}^{\text{elas}} + \mathcal{D}^{\text{Th,P}}$          | $2.0 \pm 1.0$           | $1.9 \pm 1.0$ | $2.1 \pm 1.1$                | $1.9 \pm 1.0$ | $2.1 \pm 1.0$                     | $2.0 \pm 1.0$ |
| $\mathcal{S}^{\text{elas}} + \mathcal{D}^{\text{Th,A}}$          | $1.8 \pm 0.9$           | $1.8 \pm 0.9$ | $1.9 \pm 1.0$                | $1.9 \pm 0.9$ | $2.0 \pm 0.9$                     | $1.9 \pm 0.9$ |
| $\mathcal{S}^{\text{elas}} + \mathcal{D}^{\text{Th,D}}$          | $1.8 \pm 0.8$           | $1.7 \pm 0.8$ | $1.8 \pm 0.8$                | $1.8 \pm 0.8$ | $1.9 \pm 0.8$                     | $1.8 \pm 0.9$ |

Moreover, based on a robust block matching strategy for landmark propagation between images of 4D datasets, it was shown that a target registration error computed using automatically detected landmark correspondences is appropriate for (a first) assessment of non-linear registration accuracy – at least in the case of registration of lung CT data considered in this contribution. Further studies on the interdependency of the proposed propagation strategy and dissimilarity measures considered during registration are, however, required.

Eventually, in this study only 4D image sequences were considered that represent anatomical differences due to breathing motion. It would also be interesting to evaluate the appropriateness of the proposed methods for landmark correspondence detection and evaluation of non-linear registration when applied to, e.g., baseline/follow-up image pairs.

## REFERENCES

- [1] Hugo, G. D. and Rosu, M., “Advances in 4D radiation therapy for managing respiration: Part I - 4D imaging.,” *Z Med Phys* **22**, 258–271 (2012).
- [2] Li, G., Citrin, D., Camphausen, K., et al., “Advances in 4D medical imaging and 4D radiation therapy.,” *Tech Canc Res Treat* **7**(1), 67–81 (2008).
- [3] Wang, H., Dong, L., O’Daniel, J., et al., “Validation of an accelerated ‘demons’ algorithm for deformable image registration in radiation therapy.,” *Phys Med Biol* **50**(12), 2887–2905 (2005).
- [4] Zhang, T., Orton, N. P., and Tom, W. A., “On the automated definition of mobile target volumes from 4D-CT images for stereotactic body radiotherapy.,” *Med Phys* **32**(11), 3493–3502 (2005).
- [5] Rosu, M. and Hugo, G. D., “Advances in 4D radiation therapy for managing respiration: Part II - 4D treatment planning.,” *Z Med Phys* **22**, 272–280 (2012).
- [6] Starkschall, G., Gibbons, J. P., and Orton, C. G., “Point/counterpoint. to ensure that target volumes are not underirradiated when respiratory motion may affect the dose distribution, 4D dose calculations should be performed.,” *Med Phys* **36**(1), 1–3 (2009).

- [7] Werner, R., Ehrhardt, J., Schmidt-Richberg, A., et al., “Towards accurate dose accumulation for step-&-shoot IMRT: Impact of weighting schemes and temporal image resolution on the estimation of dosimetric motion effects.,” *Z Med Phys* **22**, 109–122 (2012).
- [8] Murphy, K., van Ginneken, B., Reinhardt, J., et al., “Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge,” *IEEE Trans Med Imaging* **30**(11), 1901–1920 (2011).
- [9] Murphy, K., van Ginneken, B., Klein, S., et al., “Semi-automatic construction of reference standards for evaluation of image registration.,” *Med Image Anal* **15**(1), 71–84 (2011).
- [10] Brock, K. K. and Deformable Registration Accuracy Consortium, “Results of a multi-institution deformable registration accuracy study (MIDRAS).,” *Int J Radiat Oncol* **76**(2), 583–596 (2010).
- [11] Castillo, R., Castillo, E., Guerra, R., et al., “A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets.,” *Phys Med Biol* **54**(7), 1849–1870 (2009).
- [12] Sarrut, D., Delhay, B., Villard, P.-F., et al., “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., et al., “Estimation of motion fields by non-linear registration for local lung motion analysis in 4D CT image data.,” *Int J CARS* **5**(6), 595–605 (2010).
- [14] Murphy, K., van Ginneken, B., Pluim, J. P. W., et al., “Semi-automatic reference standard construction for quantitative evaluation of lung CT registration.,” in [*Int Conf Med Image Comput Comput Assist Interv*], **11**(Pt 2), 1006–1013 (2008).
- [15] Likar, B. and Pernus, F., “Automatic extraction of corresponding points for the registration of medical images.,” *Med Phys* **26**(8), 1678–1686 (1999).
- [16] Färber, M., Gawenda, B., Bohn, C.-A., and Handels, H., “Haptic landmark positioning and automatic landmark transfer in 4D lung CT data,” in [*Bildverarbeitung für die Medizin 2008*], Tolxdorff, T., Braun, J., Deserno, T., et al., eds., *Informatik aktuell*, 313–317, Springer (2008).
- [17] Hartkens, T., Rohr, K., and Stiehl, H. S., “Evaluation of 3D operators for the detection of anatomical point landmarks in MR and CT images,” *Computer Vision and Image Understanding* **86**, 118–136 (2002).
- [18] Rohr, K., [*Landmark-Based Image Analysis*], Kluwer Academic Publishers (2001).
- [19] Ehrhardt, J., Werner, R., Schmidt-Richberg, A., and Handels, H., “Automatic landmark detection and non-linear landmark- and surface-based registration of lung CT images,” in [*Medical Image Analysis for the Clinic: A Grand Challenge*], van Ginneken, B., Murphy, K., Heimann, T., et al., eds., *Workshop Proceedings from the 13th International Conference on Medical Image Computing and Computer Assisted Intervention*, 165–174 (2010).
- [20] Werner, R., Wolf, J. C., Ehrhardt, J., et al., “Automatische Landmarkendetektion und -übertragung zur Evaluation der Registrierung von thorakalen CT-Daten,” in [*Bildverarbeitung für die Medizin 2010*], Deserno, T., Handels, H., Meinzer, H., and Tolxdorff, T., eds., *Informatik aktuell*, 31–35, Springer (2010).
- [21] Vandemeulebroucke, J., Sarrut, D., and Clarysse, P., “Point-validated pixel-based breathing thorax model,” in [*XV<sup>th</sup> International Conference on the Use of Computers in Radiation Therapy (ICCR)*], (2007).
- [22] von Siebenthal, M., Szkely, G., Gamper, U., et al., “4D MR imaging of respiratory organ motion and its variability.,” *Phys Med Biol* **52**(6), 1547–1564 (2007).
- [23] Schmidt-Richberg, A., Ehrhardt, J., Werner, R., and Handels, H., “Diffeomorphic diffusion registration of lung CT images,” in [*Medical Image Analysis for the Clinic: A Grand Challenge*], van Ginneken, B., Murphy, K., Heimann, T., et al., eds., *Workshop Proceedings from the 13th International Conference on Medical Image Computing and Computer Assisted Intervention*, 55–62 (2010).
- [24] Schmidt-Richberg, A., Ehrhardt, J., Werner, R., and Handels, H., “Evaluation and comparison of force terms for the estimation of lung motion by non-linear registration of 4D-CT image data,” in [*World Congress on Medical Physics and Biomedical Engineering (WC 2009)*], Dössel, O. and Schlegel, W. C., eds., *IFBME Proceedings* **25**(IV), 2128–2131, Springer (2009).
- [25] Modersitzki, J., [*Numerical Methods for Image Registration*], Oxford University Press (2003).
- [26] Ramus, L. and Malandain, G., “Assessing selection methods in the context of multi-atlas based segmentation,” in [*Proc. IEEE Int Biomedical Imaging: From Nano to Macro Symp*], 1321–1324 (2010).

- [27] Frantz, S., Rohr, K., and Stiehl, H. S., “Refined localization of three-dimensional anatomical point landmarks using multistep differential approaches,” in [*SPIE Medical Imaging*], Hanson, K. M., ed., *Proc. SPIE* **3338**, 28–38 (1998).
- [28] Köthe, U., “Edge and junction detection with an improved structure tensor,” in [*Pattern Recognition*], Michaelis, B. and Krell, G., eds., *Lecture Notes in Computer Science* **2781**, 25–32, Springer (2003).