Simulation of Range Imaging-based Estimation of Respiratory Lung Motion

Influence of Noise, Signal Dimensionality and Sampling Patterns

M. Wilms¹; R. Werner²; M. Blendowski¹; J. Ortmüller¹; H. Handels¹

¹Institute of Medical Informatics, University of Lübeck, Lübeck, Germany;
²Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Keywords
Respiratory motion, motion estimation, correspondence modeling, regression, image registration

Summary
Objectives: A major problem associated with the irradiation of thoracic and abdominal tumors is respiratory motion. In clinical practice, motion compensation approaches are frequently steered by low-dimensional breathing signals (e.g., spirometry) and patient-specific correspondence models, which are used to estimate the sought internal motion given a signal measurement. Recently, the use of multidimensional signals derived from range images of the moving skin surface has been proposed to better account for complex motion patterns. In this work, a simulation study is carried out to investigate the motion estimation accuracy of such multidimensional signals and the influence of noise, the signal dimensionality, and different sampling patterns (points, lines, regions).

Methods: A diffeomorphic correspondence modeling framework is employed to relate multidimensional breathing signals derived from simulated range images to internal motion patterns represented by diffeomorphic non-linear transformations. Furthermore, an automatic approach for the selection of optimal signal combinations/patterns within this framework is presented.

Results: This simulation study focuses on lung motion estimation and is based on 28 4D CT data sets. The results show that the use of multidimensional signals instead of one-dimensional signals significantly improves the motion estimation accuracy, which is, however, highly affected by noise. Only small differences exist between different multidimensional sampling patterns (lines and regions). Automatically determined optimal combinations of points and lines do not lead to accuracy improvements compared to results obtained by using all points or lines.

Conclusions: Our results show the potential of multidimensional breathing signals derived from range images for the model-based estimation of respiratory motion in radiation therapy.

1. Introduction

Besides chemotherapy and surgery, radio-therapy is a cornerstone of curative and palliative treatment of tumors in the thoracic and abdominal regions. A major problem associated with the irradiation of these tumors is respiratory motion because many patients are not able to hold their breath sufficiently long during the treatment [1]. As a consequence, tumor motion amplitudes up to several centimeters [2] need to be compensated to ensure appropriate dose coverage of the moving tumor. One way to compensate for respiratory motion is the enlargement of safety margins, e.g., by using motion information from 4D (= 3D + t) image data sets [3]. This helps to provide the desired tumor dose coverage but also leads to large volumes of surrounding healthy tissue being irradiated. Therefore, technical approaches for effective motion compensation during the treatment have been developed in recent years, such as gating [4] and robotic motion compensation [5]. Both techniques have in common that they require a precise and real-time knowledge of the current tumor position. However, a direct and continuous monitoring of the tumor position throughout the treatment using, e.g., fluoroscopy is usually not possible due to the additional radiation exposure. In clinical practice, gating and robotic motion compensation approaches are, therefore, usually steered by low-dimensional breathing signals (so-called surrogates) that are easily acquirable during the treatment, like the expansion of the thorax/abdomen recorded with a belt system or spirometry-
based tidal volume measurements. Given a breathing signal measurement, a patient-specific correspondence model, which relates external signal measurements to internal motion patterns, is used to estimate the sought internal motion of structures of interest in real-time [6].

Patient-specific respiratory motion has been shown to be highly variable [7]. Observable effects include hysteresis (intracycle variability), variations in inspiration depth between cycles, and baseline shifts. As a consequence, correspondence models solely parameterized by one-dimensional surrogate signals (e.g., spirometry) are not able to sufficiently describe all of these effects and motivate the introduction of multidimensional surrogate signals [6]. With the advent of low-cost, high-resolution range imaging devices, like structured light or Time-of-Flight (ToF) cameras (e.g., Microsoft Kinect), in recent years, much attention has been drawn to the use of (multidimensional) signals derived from range images (depth maps) of the moving skin surface for motion estimation and patient positioning purposes in radiation therapy (see the overviews by McClelland et al. [6] and Bauer et al. [8]). In range images, pixel values represent the distance between the sensor and points on objects in the scene.

Fayad et al. [9] showed in a simulation study based on 4D CT data sets that correspondence models parameterized by multidimensional surrogate signals, representing the average motion of 10 different regions of the patients’ skin surface extracted from simulated range images, lead to better results in terms of motion estimation accuracy than models trained with 1D signals. These findings were experimentally verified in [10] by comparing surrogate-based motion estimation results obtained with real ToF data and an abdominal belt signal for one patient.

However, using multidimensional surrogate signals containing all points that are offered by a high-resolution range image increases processing times and can lead to computational problems arising from the presence of highly redundant information (e.g., multicollinearities in regression-based modeling). These problems can, for example, be circumvented by combining several neighboring points into a small number of regions [9], by pre-selecting a limited number of optimal points [11] or by applying other dimensionality reduction techniques [12].

Although some work has been done on the use of (multidimensional) signals derived from range images for motion estimation based on correspondence models in radiation therapy (see above), to our knowledge, no existing study provides an investigation of the influence of noise, signal dimensionality, and (optimally selected/combined) sampling patterns (e.g., points, lines, whole regions) on the estimation accuracy, which is, therefore, the purpose of this work. Our simulation study is based on 28 4D CT data set and focuses on surrogate-based lung motion estimation. We employ our recently proposed diffeomorphic correspondence modeling approach [13], to relate multidimensional surrogate signals to internal motion patterns represented by diffeomorphic non-linear transformations.

2. Methods and Materials

In this section, we briefly introduce our diffeomorphic correspondence modeling framework (section 2.1), describe the range imaging simulation method used (section 2.2), present an automatic approach for the selection of optimal surrogate signal combinations (section 2.3), and explain our study design (section 2.4).

2.1 Diffeomorphic Correspondence Modeling Framework

Let \( \{ f_j \}_{j \in \mathbb{N}} : \Omega \to \mathbb{R}^3 \) denote a 4D CT data set consisting of \( n_{ph} \) breathing phases \( j \). Furthermore, a set of corresponding \( n_{sur} \)-dimensional (multidimensional) surrogate signal measurements \( \{ \zeta_j \}_{j \in \mathbb{N}} \) with \( \zeta_j \in \mathbb{R}^{n_{sur}} \), is assumed to be given.

In our approach [13], the motion of internal structures between a reference breathing phase \( I_0 \) and each other phase \( I_j \) of the 4D CT data set is described by diffeomorphic non-linear transformations \( \phi_j : \Omega \to \Omega \). Following Arsigny et al. [14], we parameterize them by stationary velocity fields \( \psi_j : \Omega \to \mathbb{R}^3 \), which leads to \( \phi_j = \exp(\psi_j) \) with \( \exp \) being the group exponential map. Diffeomorphic transformations are topology preserving as they are bijective and differentiable mappings with a differentiable inverse, making them a natural choice for respiratory motion estimation and modeling [15]. In this work, the registration scheme proposed by Schmidt-Richberg et al. [16] is used to determine the diffeomorphic non-linear transformations \( \phi_j \).

Now, we are interested in learning a diffeomorphic correspondence model, which appropriately relates signal measurements to internal motion. Model training is based on the signal measurements \( \{ \psi_j \}_{j \in \mathbb{N}} : \Omega \to \mathbb{R}^3 \) and the estimated diffeomorphic transformations \( \{ \phi_j \}_{j \in \mathbb{N}} \). Furthermore, our approach relies on the Log-Euclidean framework [14] to ensure that transformations generated by the trained correspondence models are diffeomorphic. Roughly speaking, the use of the Log-Euclidean framework allows us to perform statistics on the diffeomorphic transformations \( \{ \phi_j \}_{j \in \mathbb{N}} \) by calculating simple Euclidean statistics on the corresponding stationary velocity fields \( \{ \psi_j \}_{j \in \mathbb{N}} \).

From here on, both the surrogate measurements and the velocity fields are interpreted as random variables \( \mathbf{Z}_j \) and \( \mathbf{V}_j \in \mathbb{R}^{3n_{sur}} \) \( (n = \text{number of voxels of image } I_j) \), respectively. With mean vectors \( \mathbf{V} = 1/n_{ph} \sum_{j=1}^{n_{ph}} \mathbf{V}_j \) and \( \mathbf{Z} = 1/n_{ph} \sum_{j=1}^{n_{ph}} \mathbf{Z}_j \), our diffeomorphic correspondence model is defined as

\[
\tilde{\mathbf{V}} = \mathbf{V} + \mathbf{B} (\mathbf{Z} - \bar{\mathbf{Z}})
\]

Using this model, a velocity field vector \( \mathbf{V} \) can be estimated given a corresponding surrogate signal measurement \( \mathbf{Z} \). The sought transformation \( \hat{\phi} = \exp(\hat{\mathbf{v}}) \) is then obtained by interpreting the entries of \( \hat{\mathbf{V}} \) as components of a velocity field \( \hat{\mathbf{v}} \) and subsequently applying the exponential map. Based on the available training data, we estimate the coefficient matrix

B ∈ R^{2m} × R^{n} \text{by performing multivariate multiple linear regression analysis, or more precisely, by solving the system of linear equations}

V = BZ,

with V := (V_{0}^e, ..., V_{n_y}^e) and Z := (Z_1^e, ..., Z_{n}^e) being matrices that contain the mean centered random variables V_{i}^e = V_{i} - \bar{V} and Z_{i}^e = Z_{i} - \bar{Z}. This linear system is solved by

B = VZ^{*},

where Z^{*} denotes the Moore-Penrose pseudoinverse of Z (i.e., Z^{*} = Z^{T}ZZ^{T})^{-1} or Z^{*} = (Z^{T}Z)^{-1}Z^{T}, cf. [17]). In this work, multidimensional surrogate signals with highly or even perfectly correlated dimensions (multicollinearities) are used. Therefore, Z will always be (at least nearly) singular. We try to circumvent this problem by approximating the matrices ZZ^{T} and Z^{T}Z with ZZ^{T} + \lambda I and Z^{T}Z + \lambda I, (\lambda > 0), respectively (Tikhonov regularization). An optimal \lambda is determined via leave-one-out cross validation on the model’s training data.

2.2 Simulation of Range Imaging-based Surrogate Signals

Spatio-temporal image data sets with corresponding range images, as in principle needed for this study, are only rarely acquired in clinical practice. We, therefore, simulate range images covering the motion of the chest wall/abdominal skin surface from available 4D CT data sets.

Given a 4D CT data set with n_{ph} phases (cf. section 2.1), range images of the skin surface are simulated by placing a virtual range imaging sensor with a sensor size of x_{res} × y_{res} pixels directly above a region-of-interest (ROI) of the patient’s body in each 3D CT image I_{j}. Noise present in the CT images is reduced by employing a bilateral filter. Then, rays starting at the center of each pixel of the sensor plane are traced in anterior-posterior direction until they intersect with the skin surface according to a predefined grey-value threshold.

The realism of the proposed simulation approach is further enhanced by integration of pixel-wise uncertainties related to the depth measurements. This is done by utilizing an experimentally derived Gaussian noise model of the Microsoft Kinect sensor presented by Olesen et al. [18]. The variance \sigma_{noise,d} of the Gaussian noise added to a simulated depth measurement depends on the depth d (in mm) and is given by

\sigma_{noise,d} = p_{0} + p_{1}d + p_{2}d^{2},

with p_{0} = 2.344, p_{1} = -2.734 × 10^{-3}, p_{2} = 1.233 × 10^{-6}. In contrast to the model proposed in [18], we neglect the dependence of \sigma_{noise,d} from the radial distance between a particular sensor pixel and the sensor center for simplicity. Figure 1 shows simulated range images with and without artificially added noise.

Surrogate signals of different dimensionality are derived from the simulated range images by varying the number of points selected (= sensor pixels finally considered) and by using different sampling patterns. Here, we distinguish three types of sampling patterns: 1) single points with n_{sur} = 1, 2) lines of n_{sur} neighboring points, and 3) whole (rectangular) regions of n_{sur} = x_{sur}y_{sur} neighboring points.

2.3 Automatic Selection of Optimal Surrogate Signal Combinations

Using all points of a high-resolution range image of the complete skin surface as a surrogate signal (regions in section 2.2) can lead to high processing times and computational problems due to the linear relations of the different signal dimensions, as mentioned in sections 1 and 2.1. Therefore, our goal is to generate multidimensional surrogate signals with as little redundant information as possible by automatically determining an optimal subset of possible surrogate signals (here: points or lines).

Given a 4D CT data set with n_{ph} phases (cf. section 2.1) and a set of n_{sig} surrogate signals (e.g., n_{sig} different points/lines), an optimal subset U_{opt} \subseteq S is sought that is best suited for model-based estimation of the internal motion (cf. section 2.1). A leave-out cross validation strategy is applied to determine the suitability of a subset U. More specifically, n_{ph} models are
trained on \( n_{ph} = 1 \) breathing phases by leaving out a different phase \( j \in L \subseteq [1, \ldots, n_{ph}] \) each time. Subsequently, each model is utilized to estimate the velocity field vector \( V_j^U \), which characterizes the motion between the reference phase and the left out phase \( j \). Squared Euclidean distances \( \| V_j - V_j^U \|_2 \) between estimated vectors \( V_j \) and corresponding ‘ground truth’ vectors \( V_j^U \) obtained via registration (cf. section 2.1) are calculated and summed to determine the suitability of a particular subset \( U \). More formally, the selection of an optimal surrogate signal combination \( U_{opt} \subseteq S \) can be expressed as the optimization problem

\[
U_{opt} = \arg \min_{U \in \mathcal{P}(S)} \sum_{j \in L} \| V_j - V_j^U \|_2^2,
\]

where \( \mathcal{P}(S) \) denotes the power set of \( S \). In our implementation, sequential forward selection is used as the (heuristic) optimization algorithm because we assume small optimal subsets.

### 2.4 Study Design

This simulation study aims at investigating the influence of noise, signal dimensionality, and sampling patterns on the accuracy of surrogate-based estimation of respiratory motion using (multi-dimensional) surrogate signals derived from range images. We focus on respiratory lung motion estimation and use 28 proprietary and publicly available thoracic 4D CT data sets of 28 different patients:

- twelve data sets consisting of 10–14 breathing phases with an average spatial resolution of \( 1 \times 1 \times 1.5 \) mm taken from our in-house database [2].
- ten publicly available 4D CT data sets from the DIR-Lab at the University of Texas M. D. Anderson Cancer Center, USA, consisting of 10 breathing phases with a spatial resolution of \( 0.97 \times 1.16 \times 0.97 \times 1.16 \times 2.5 \) mm [19].
- six 4D CT data sets publicly available from the Léon Bérard Cancer Center & CREATIS lab, Lyon, France. These data sets consist of 10 breathing phases with a spatial resolution of \( 0.897 \times 1.172 \times 0.879 \times 1.172 \times 2 \) mm [20].

In total, 18 different range imaging-based surrogate signals are generated for each patient: First, patient-specific rectangular regions of interest (ROIs) of the chest wall/abdominal skin surface are defined to exclude parts of the images that contain abdominal belts or the neck area (Figure 2). A realistic distance of \( d = 1500 \) mm is assumed between the closest part of the skin surface and the sensor for noise generation. The automatic selection approach (cf. section 2.3) is used to generate a one-dimensional signal containing the motion of the optimal point of the 25 pixel range image and a multi-dimensional signal containing the motion of the optimal combination of points of the 25 pixel range image. From the 10,000 pixel range image, 8 different lines of 100 points length are sampled (3 equally spread horizontal/vertical lines, 2 diagonals) [19].

In total, 18 different range imaging-based surrogate signals are generated for each patient: First, patient-specific rectangular regions of interest (ROIs) of the chest wall/abdominal skin surface are defined to exclude parts of the images that contain abdominal belts or the neck area (Figure 2). A realistic distance of \( d = 1500 \) mm is assumed between the closest part of the skin surface and the sensor for noise generation. The automatic selection approach (cf. section 2.3) is used to generate a one-dimensional signal containing the motion of the optimal point of the 25 pixel range image and a multi-dimensional signal containing the motion of the optimal combination of points of the 25 pixel range image. From the 10,000 pixel range image, 8 different lines of 100 points length are sampled (3 equally spread horizontal/vertical lines, 2 diagonals) [19].

### 3. Results

The results of the conducted experiments are summarized in Table 1. Paired t-tests (\( p < 0.05 \), pairing of patient-specific mean TREs) are performed to determine whether...
significant differences exist between the overall mean values. First of all, our results show that realistic noise present in the surrogate signals negatively affects the estimation accuracy. All differences between results obtained for the same surrogate signal/experiment with and without noise are statistically significant with the only exception being the optimal combination of points for the extrapolation scenario (2.45 ± 1.28 mm vs. 3.28 ± 1.77 mm, p = 0.005). Regarding the influence of the signal dimensionality, our results reveal that in general the use of multidimensional surrogate signals significantly improves the estimation accuracy compared to one-dimensional signals (optimal points). Some exceptions are, for example, the differences reported between the optimally selected point and the optimal combination of points. For the signals without noise, only small and mostly non-significant differences are observed between the various multidimensional line and region samplings. However, when noise is added to the range images, the motion estimation accuracy obtained for the 25 points region sampling signals in the interpolation scenario is significantly worse than the results of all other high-dimensional signals (w/o optimal combination of points).

On average, ≈ 5 points and ≈ 2 lines were selected by the automatic selection approach. For both extrapolation scenarios, the accuracy significantly decreases when the optimal point combination is used instead of the single best point. However, motion estimation with all 25 points generally leads to significantly better results than using the determined optimal combination or the optimal point. Mostly different points/ combinations were selected for different patients (Figure 4). On a patient-specific level, the selections are also different between the scenarios (inter-/extrapolation, without noise/with noise). The same is observed for the line selections. However, for them, no significant difference exists between the accuracy achieved with the optimal single line, the optimal combination, and the combination of all 8 lines (exception: extrapolation with noise, opt. line/opt. comb of lines significantly better than all 8 lines).

### Table 1

<table>
<thead>
<tr>
<th>Surrogate Signal</th>
<th>Without Noise</th>
<th>With Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Extrapolation</td>
<td>Interpolation</td>
</tr>
<tr>
<td>No motion estimation</td>
<td>7.59 ± 2.72</td>
<td>3.59 ± 1.46</td>
</tr>
<tr>
<td>Intra-patient registration</td>
<td>1.40 ± 0.49</td>
<td>1.52 ± 0.15</td>
</tr>
<tr>
<td>Points</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal point</td>
<td>2.28 ± 0.93</td>
<td>1.92 ± 0.30</td>
</tr>
<tr>
<td>Optimal comb. of points</td>
<td>2.45 ± 1.28</td>
<td>1.81 ± 0.33</td>
</tr>
<tr>
<td>Lines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal line</td>
<td>1.90 ± 0.70</td>
<td>1.72 ± 0.21</td>
</tr>
<tr>
<td>Optimal comb. of lines</td>
<td>1.87 ± 0.68</td>
<td>1.70 ± 0.21</td>
</tr>
<tr>
<td>All 8 lines</td>
<td>1.81 ± 0.49</td>
<td>1.71 ± 0.20</td>
</tr>
<tr>
<td>Regions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 points</td>
<td>1.92 ± 0.69</td>
<td>1.71 ± 0.20</td>
</tr>
<tr>
<td>100 points</td>
<td>1.91 ± 0.70</td>
<td>1.68 ± 0.20</td>
</tr>
<tr>
<td>1000 points</td>
<td>1.82 ± 0.55</td>
<td>1.69 ± 0.20</td>
</tr>
<tr>
<td>10000 points</td>
<td>1.82 ± 0.54</td>
<td>1.69 ± 0.20</td>
</tr>
</tbody>
</table>
Computation times for model training when a set of velocity fields and corresponding signal measurements is given range from ≈1 minute (25 points) to ≈75 minutes (optimal combination of points) on a standard PC (Intel W3520 2.67 GHz CPU and 24GB RAM). The model-based generation of a new velocity field for a signal measurement takes <1 second. It has to be noted that the current implementation is not runtime-optimized.

4. Discussion and Conclusion

In this work, we investigated the influence of signal dimensionality, the effects of different sampling patterns and the influence of noise on range imaging-based estimation of respiratory lung motion in the context of radiotherapy. The study was based on image-based simulations of range images using 28 4D CT data sets and a diffeomorphic correspondence modeling framework.

The study results showed that, in general, the use of multidimensional signals over one-dimensional signals leads to significantly improved estimation accuracies. However, independent of the sampling pattern, increasing the dimensionality of the signal beyond 25 (w/o noise)/100 (with noise) dimensions did not further significantly improve the accuracy. Interestingly, only small differences were found between the different sampling patterns used to generate multidimensional signals (lines, regions). Moreover, the results show that our correspondence modeling approach is highly affected by realistic noise on the surrogate signal. This finding suggests that in clinical practice noise needs to be accounted for, for example, by preprocessing the noisy ranges images (e.g., by temporally averaging several frames) or by applying appropriate dimensionality reduction approaches (e.g., principal component analysis [12]) to the extracted surrogate signal.

The automatically selected optimal subsets of points and lines were not found to be superior in terms of estimation accuracy when compared to results obtained by using all points or lines, respectively. In some cases, results of selected point combinations were even significantly worse than those of all 25 points. This indicates two things: 1) The time-consuming automatic selection approach sometimes stops in local minima. 2) Effects of problems theoretically arising from the use of high-dimensional surrogate signals (multicollinearities, overfitting) are mostly avoided by the presented training scheme (Tikhonov regularization with leave-out cross validation). Regarding the selected subsets, only minimal overlaps were found between sets of different patients and, interestingly, also between sets of different evaluation scenarios of the same patient. The first finding is in agreement with results reported in [11], whereas the observed differences between subsets of the same patient indicate that the subset selection is unstable, which is a well-known problem when using feature selection methods [21].

A reason for this instability might be the small number of training samples used here (7–13 training phases), but further research is required to confirm this assumption.

Two limitations, which may affect the generalizability of the findings of this study to clinical practice, are the relatively small number of data sets (12–28 depending on the experiment) it is based on and the use of simulated range images instead of real depth measurements. Future work will therefore include the incorporation of real range imaging data and an evaluation on additional data sets. We plan not only to add more patients but also to use more data sets per patient (e.g., follow-up 4D CT scans or 4D MRI data) to assess the estimation accuracy in presence of further clinically relevant types of breathing variability (e.g., intra-/inter-fraction variability).

Acknowledgments

This work is supported by the German Research Foundation DFG (HH 2355/9–2).

References