REAL-TIME RESPIRATORY SIGNAL EXTRACTION FROM X-RAY SEQUENCES USING INCREMENTAL MANIFOLD LEARNING

Peter Fischer[⋆] *Thomas Pohl*[†] *Joachim Hornegger*[⋆]

[⋆] Pattern Recognition Lab and Erlangen Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany
[†] Siemens AG, Healthcare Sector, Forchheim, Germany

ABSTRACT

In X-ray fluoroscopy-guided minimally invasive interventions, overlays of pre-procedurally acquired image data can be used to visualize soft-tissue. In the thoracic and abdominal regions, static overlays are inconsistent to the live X-ray images due to respiratory motion of the patient. This error can be reduced by dynamically adapting the overlay to the respiration. A first step in this direction is the real-time extraction of the respiratory state from the live X-ray images. The respiratory state can drive a motion model to compensate the breathing motion.

We present a method to extract respiratory signals from X-ray sequences in real-time. Respiratory signal extraction is viewed as a dimensionality reduction problem, which is performed for each X-ray image using incremental Isomap. The method has a correlation of 0.97 ± 0.02 with internal breathing motion and an average runtime of 42 ms per image. The method is accurate, robust, and can be used in a wide range of clinical applications and fields of view.

***Index Terms*—** respiratory signal extraction, motion compensation, X-ray fluoroscopy, dimensionality reduction, manifold learning

1. INTRODUCTION

Fluoroscopic guidance is well suited to visualize high density structures like interventional devices and bones. However, many minimally invasive interventions require the visualization of soft tissues, e.g. heart, liver, or vasculature, which have low contrast in X-ray images. Soft tissue can be visualized by overlays generated from pre-procedurally acquired 3D or 4D images like CT or MR. The usefulness of static overlays is limited due to cardiac and respiratory motion of the patient. The motion causes an inconsistency between the overlay and the live X-ray images. Dynamically adapted overlays are potentially much more accurate. This work focuses on respiratory motion only.

There are many approaches to compensate for patient motion. Tracking of well-visualized objects that are coupled with the breathing motion, e.g. certain interventional devices [1] or body parts [2], provides a direct estimate of the desired motion, but is restricted to specific applications or fields of view. A widely applicable method are respiratory motion models. A model of the tissue motion due to respiration is generated pre-procedurally. The model allows to infer the motion using a surrogate respiratory signal. During the intervention, only the surrogate signal has to be acquired. A comprehensive introduction to respiratory motion models can be found in [3].

The respiratory signal can be acquired in different ways. A straight-forward solution is to measure the respiration using an external device. External devices create additional effort for patient setup and synchronization with the imaging system. These drawbacks are avoided if the respiratory signal is extracted directly from the X-ray images. Berbeco et al. achieve this by measuring the intensity changes in a region of interest (ROI) [4]. Vergalasova et al. analyze the effects of respiratory motion on the frequency transform of X-ray images [5]. Zijp et al. enhance diaphragm-like features and track their movement along the cranio-caudal axis [6]. In recent years, dimensionality reduction techniques have been adopted frequently to extract the respiratory signal. Manifold learning is a popular nonlinear approach to dimensionality reduction. Wachinger et al. use manifold learning to generate a respiratory signal for ultrasound and MR gating [7]. Yan et al. approximate the breathing manifold locally using foreground extraction and local principal component analysis (PCA) for cone beam CT reconstruction [8].

So far, manifold learning has been applied to respiratory signal extraction only retrospectively after all the images were acquired, which is not suitable for dynamic overlays. This work presents a method to extract a respiratory signal using manifold learning from live X-ray sequences in real-time. In addition, we experimentally compare it to the state of the art.

This work was supported by Siemens AG, Healthcare Sector, Forchheim, Germany. The concepts and information presented in this paper are based on research and are not commercially available. The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative.

2. MATERIALS AND METHODS

2.1. Manifold Learning

Dimensionality reduction is the projection of N points $\mathbf{X}$ in high-dimensional space $\mathbb{R}^D$ to N points $\mathbf{x}$ in low-dimensional space $\mathbb{R}^d$. Manifold learning describes dimensionality reduction methods that assume that the data in high dimensional space actually lies on a low-dimensional, nonlinear manifold. The goal is to remove the unnecessary dimensions and unwarp the nonlinearity in order to embed the data in a low-dimensional Euclidean space.

There are a of number manifold learning algorithms, each trying to preserve different properties of the high-dimensional space. For our application, Isomap yields good results [9]. In Isomap, the mapping to low-dimensional space is designed to preserve the geodesic distances between the high-dimensional points. The geodesic distance between two points is measured along the manifold. This is realized by a discrete approximation of the manifold with a k-nearest-neighbor (kNN) graph. The nearest neighbors are determined by the Euclidean distance $\|\mathbf{X}_i - \mathbf{X}_j\|_2$. The geodesic distance between points is computed using shortest paths on the kNN-graph. The result is a matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$ of geodesic distances between points, which can be interpreted as a kernel matrix. The low-dimensional embedding is determined by eigendecomposition of the centered kernel matrix $\tilde{\mathbf{K}}$. Given the eigenvector $\mathbf{v}_j$ corresponding to the j-th largest eigenvalue of $\tilde{\mathbf{K}}$, the embedding of the i-th point is $\mathbf{x}_i(j) = \sqrt{N}\mathbf{v}_j(i)$, where $\mathbf{v}_j(i)$ is the i-th component of $\mathbf{v}_j$. As the relative importance of the eigenvectors is not important for this application, the embedding is not weighted by the eigenvalues as in standard Isomap. Instead, the eigenvectors are normalized to length N to avoid shrinkage of the eigenvector components due to newly arriving samples.

2.2. Incremental Isomap

Real-time computation of Isomap is not feasible for each live image. With the standard implementation, nearest neighbor search grows with $\mathcal{O}\left(N^2\right)$, while all pair shortest path and eigendecomposition each grow with $\mathcal{O}\left(N^3\right)$. An option is to compute Isomap in a training sequence and use out of sample extension to embed live images. However, this option does not use the live images to improve the mapping and is not adaptive. A better option is incremental Isomap, which updates the embedding without recomputing unchanged information [10]. This is achieved by only computing the Euclidean distances between the new point $\mathbf{X}_{N+1}$ and the old points. The changes in the kNN-graph are tracked and only the affected shortest paths are recomputed. In addition, the eigendecomposition is not computed from scratch, but updated from the previous solution, assuming that the new point does not change the manifold and thus the eigenvectors and eigenvalues drastically.

2.3. X-Ray Sequence Processing

For X-ray processing, the input of the dimensionality reduction is a sequence of images of $M \times M$ pixels, each corresponding to a vector of dimensionality $D = M \cdot M$. The manifold assumption is well justified for medical X-ray sequences, as the imaging system is a controlled environment. All the variations in the images stem from motion of the patient, e.g. cardiac and respiratory motion, or interactions of the physician, e.g. device motion and contrast agent injection. We are interested in respiratory motion, which is assumed to be the most dominant cause of variation. If this is not true, the images are embedded in more than one dimension. The dimension with the largest eigenvalue that has a dominant frequency in the known range of human breathing is then identified as the respiratory signal. However, the assumption is valid in all our experiments. Consequently, the algorithm maps each image to a one-dimensional respiratory state ($d = 1$).

The mapping is learned for each sequence in an unsupervised manner. Therefore, no manual interaction is required and the algorithm does not need to deal with inter-patient variation. As the breathing motion is not parametrized, complex motion patterns can be handled automatically [7]. This approach also implies constraints. Respiratory motion is assumed to be quasi-periodic, i.e. images from the same respiratory state have a similar appearance. The reason is that the embedding is based only on image similarities. This also implies that at least one full breathing cycle must be observed in a training phase of T images in the beginning of the procedure until the mapping can give correct results. Another drawback of this method is that a change of the C-arm angulation or the table position invalidates the mapping. Therefore, it must be relearned after each system movement. However, if image acquisition is paused and the C-arm position is not changed, the mapping is still valid. In addition, if only a few fixed C-arm positions are used during a clinical procedure, a separate mapping can be maintained for each position.

Apart from real-time computation, incremental Isomap allows for adaptively changing the manifold [10]. This is useful to adapt the embedding to gradual changes of the image appearance. This can be implemented easily by removing the oldest image from the problem each time a new image arrives [10]. Additionally, this has the benefit that memory consumption and runtime, which grow approximately linearly with the number of images, are bounded. However, the experiments of this paper keep all the images of the respective sequence.

3. EVALUATION AND RESULTS

The data set consists of 48 X-ray sequences from animal studies containing moving catheters and contrast agent. The field of view includes images of the thorax and abdomen. The sequence length N varies between 74 and 588. The images have

a size of $M = 1024$ pixels in each dimension, at a sampling rate of 8 to 15 images per second. The C-arm position is not changing during a single sequence.

The proposed incremental Isomap method (INCISO) is compared to the following algorithms:

- IA: intensity analysis in a ROI [4]
- AS: tracking of diaphragm-like features in the Amsterdam Shroud [6]
- LPCA: foreground extraction and local PCA [8]
- BATCH: manifold learning with standard Isomap

These algorithms cover a wide range of approaches to respiratory signal extraction. The BATCH and LPCA methods require the full sequence for processing and are thus not applicable to real-time respiratory signal extraction. The ROI for IA is fixed to the lower two thirds of the image. The parameters for INCISO are $K = 20$ neighbors for the graph and a training phase of $T = 40$ images. The same K is used for BATCH. The algorithms are not sensitive to changes in K, the results for 5, 10, 20 and 30 are similar. As fine image details are not crucial for BATCH and INCISO, the images are downsampled by a factor of 4 for these algorithm.

The quantitative comparison of algorithms is difficult as a ground truth respiratory signal is in general not available. In the 13 sequences where the diaphragm is visible throughout, a diaphragm tracking algorithm with additional manual supervision is used as ground truth [2]. The movement of the diaphragm in cranio-caudal direction is assumed to correspond to respiratory motion. The benefit of the above methods is that they are applicable even if the diaphragm is not visible. The algorithm outputs are compared to the diaphragm signal using normalized cross-correlation ($NCC_{diaphragm}$). To investigate how consistent the methods are over different views, the respiratory signals extracted from 6 simultaneous biplane sequences are compared to each other using normalized cross-correlation ($NCC_{biplane}$). Ideally, the respiratory signals should be the same irrespective of the view. To be able to evaluate all 48 sequences, the full inhale peaks were identified manually in the images. Full exhalation is not used as it is a longer state where the start and end are hard to define and measure exactly. The offset (in number of images) of the manually extracted peaks to the nearest peaks of the algorithm signals is used as a measure (Δ_{peak}). A peak is detected in the respiratory signals automatically if the signal value is higher than the 3 previous and subsequent samples.

The results of the experiments are summarized in Table 1. Either BATCH or LPCA is always the best method. BATCH has a high correlation of 0.98 to the diaphragm tracking signal, which shows that breathing motion is captured accurately. The LPCA signals from both detectors in the biplane sequence correlate almost perfectly. The results for INCISO are only slightly worse than for BATCH. Therefore, the incremental approximation is not harmful to the results. This is remarkable, as the incremental method has much less information available to compute the embedding, especially for the first images. IA and AS perform badly using both correlation measures, as the signals are noisy and influenced by many effects not related to breathing. Especially the low $NCC_{biplane}$ of 0.37 and 0.13 show that the breathing signals are not reliable. The peak errors show no clearly superior method, only AS has a substantially higher error.

Method	$NCC_{diaphragm}$	Δ_{peak}	$NCC_{biplane}$
IA	0.85 ± 0.16	1.2 ± 1.2	0.37 ± 0.25
AS	0.88 ± 0.10	1.9 ± 2.0	0.13 ± 0.13
LPCA	0.96 ± 0.01	1.3 ± 2.6	0.99 ± 0.003
BATCH	0.98 ± 0.01	1.0 ± 1.8	0.97 ± 0.01
INCISO	0.97 ± 0.02	1.1 ± 1.4	0.97 ± 0.01

Table 1. Comparison of respiratory signal extraction methods using several error metrics (mean $\pm$ standard deviation).

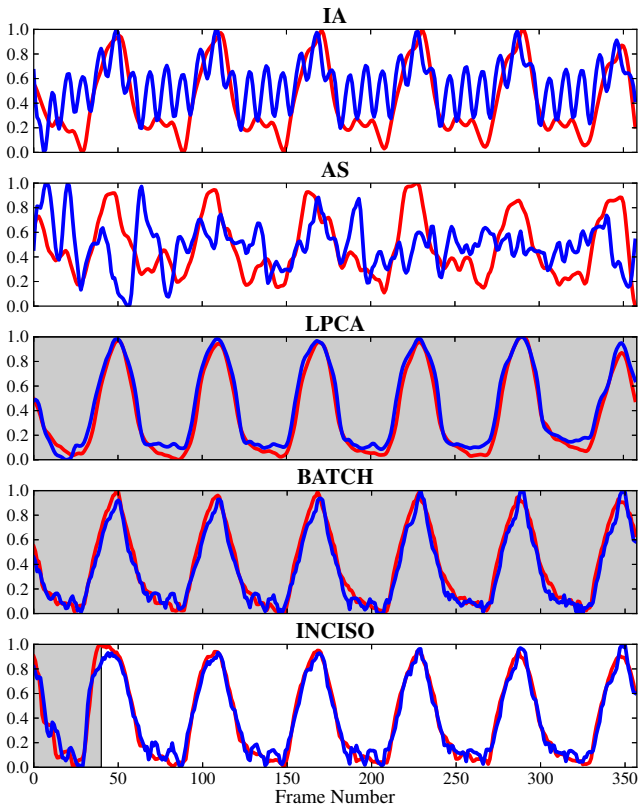

Fig. 1. Respiratory signals extracted from a biplane sequence using the investigated methods. The *red* lines are with anterior-posterior view, the *blue* lines with lateral view. *Grey* background indicates the length of the training phase.

In Figure 1, a qualitative impression of the extracted signals using the various methods is given. For all methods, 1 corresponds to full inhale and 0 to full exhale if the respiratory signal is extracted correctly. The signals originate from a biplane sequence, for which the first images are shown in Figure 2. The sequences of the two planes have different mo-

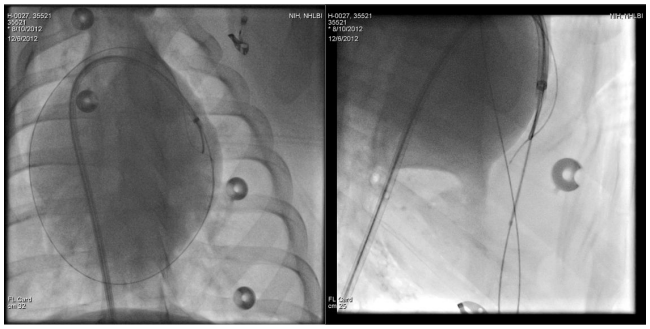

Fig. 2. The initial images of the biplane sequence used in Figure 1. The *left* image is acquired with anterior-posterior view, the *right* with lateral view.

tion directions, fields of view, and intensity distributions. IA and AS react sensitively to these changes. Especially the blue signals in Figure 1 from the lateral view do not correspond to breathing motion. The three other methods are based on dimensionality reduction and seem to be robust to different fields of view. Note that the difference between incremental and standard Isomap is getting smaller at the end of the sequence, because the incorporated data is more similar.

The average runtime of INCISO on a consumer laptop using a Python/Cython implementation is 42 ± 43 ms per image. The high standard deviation is a result of the different sequence lengths, which can be limited by forgetting old images, see Section 2.3. The computationally most intensive task is updating the shortest paths, followed by computing the image distances. For comparison, BATCH needs on average 6.9 ± 9.6 s to process a whole sequence.

4. CONCLUSIONS

We presented an accurate, real-time capable respiratory signal extraction algorithm based on incremental dimensionality reduction. It can be used in clinical applications that require an adaptive respiratory signal or real-time performance and infrequent changes of the field of view. The method is not restricted to C-arm X-ray imaging, but can be transferred to many other modalities like range imaging, MR, or CT. In future work, the algorithm could be improved by introducing more robust image similarity measurements, for example by detecting and removing the influence of catheters or collimators.

5. ACKNOWLEDGMENT

The authors thank Hao Yan for providing software to compute the IA, AS, and LPCA methods. We thank Anthony Faranesh and Robert Lederman from the National Institutes of Health, Bethesda, Maryland, USA for providing the X-ray images.

6. REFERENCES

[1] Alexander Brost, Andreas Wimmer, Rui Liao, Felix Bourier, Martin Koch, Norbert Strobel, Klaus Kurzidim, and Joachim Hornegger, "Constrained Registration for Motion Compensation in Atrial Fibrillation Ablation Procedures," *IEEE Trans. Med. Imag.*, vol. 31, no. 4, pp. 870–881, 2012.

[2] Marco Bögel, Andreas Maier, Hannes G. Hofmann, Joachim Hornegger, and Rebecca Fahrig, "Diaphragm Tracking in Cardiac C-Arm Projection Data," in *Bildverarbeitung für die Medizin*, 2012, pp. 33–38.

[3] Jamie R. McClelland, David J. Hawkes, Tobias Schaeffter, and Andrew P. King, "Respiratory Motion Models: A Review," *Med. Image Anal.*, vol. 17, no. 1, pp. 19–42, 2013.

[4] Ross I. Berbeco, Hassan Mostafavi, Gregory C. Sharp, and Steve B. Jiang, "Towards Fluoroscopic Respiratory Gating for Lung Tumours without Radiopaque Markers," *Phys. Med. Biol.*, vol. 50, no. 19, pp. 4481–4490, 2005.

[5] Irina Vergalasova, Jing Cai, and Fang-Fang Yin, "A Novel Technique for Markerless, Self-Sorted 4D-CBCT: Feasibility Study," *Med. Phys.*, vol. 39, no. 3, pp. 1442–51, 2012.

[6] Lambert Zijp, Jan-Jakob Sonke, and Marcel van Herk, "Extraction of the Respiratory Signal from Sequential Thorax Cone-Beam X-Ray Images," in *Int. Conf. on the Use of Computers in Radiation Therapy*, 2004, pp. 507–509.

[7] Christian Wachinger, Mehmet Yigitsoy, Erik-Jan Rijkhorst, and Nassir Navab, "Manifold Learning for Image-Based Breathing Gating in Ultrasound and MRI," *Med. Image Anal.*, vol. 16, no. 4, pp. 806–818, 2012.

[8] Hao Yan, Xiaoyu Wang, Wotao Yin, Tinsu Pan, Moiz Ahmad, Xuanqin Mou, Laura Cerviño, Xun Jia, and Steve B. Jiang, "Extracting Respiratory Signals from Thoracic Cone Beam CT Projections," *Phys. Med. Biol.*, vol. 58, no. 5, pp. 1447–1464, 2013.

[9] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction.," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[10] Martin H. C. Law and Anil K. Jain, "Incremental Nonlinear Dimensionality Reduction by Manifold Learning.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 377–391, 2006.