

# Keep breathing!

## Common motion helps multi-modal mapping

V. De Luca<sup>\*1</sup>, H. Grabner<sup>1</sup>, L. Petrusca<sup>2</sup>, R. Salomir<sup>2</sup>,  
G. Székely<sup>1</sup>, and C. Tanner<sup>1</sup>

<sup>1</sup> Computer Vision Laboratory, ETH Zürich, 8092 Zürich, Switzerland

<sup>2</sup> Radiology Department, Geneva University Hospital, Geneva, Switzerland

**Abstract.** We propose an unconventional approach for transferring of information between multi-modal images. It exploits the temporal commonality of multi-modal images acquired from the same organ during free-breathing. Strikingly there is no need for capturing the same region by the modalities. The method is based on extracting a low-dimensional description of the image sequences, selecting the common cause signal (breathing) for both modalities and finding the most similar subsequences for predicting image feature location. The approach was evaluated for 3 volunteers on sequences of 2D MRI and 2D ultrasound images of the liver acquired at different locations. Simultaneous acquisition of these images allowed for quantitative evaluation (predicted versus ground truth MRI feature locations). The best performance was achieved with signal extraction by slow feature analysis resulting in an average error of 2.6 mm (4.2 mm) for sequences acquired at the same (a different) time.

**Keywords:** motion prediction, multi-modal, ultrasound, magnetic resonance images, tracking, liver

## 1 Introduction

The attention paid to organ motion due to breathing during radiation therapy has risen in recent years [4]. As treatment devices become more sophisticated, their guidance requires higher accuracy. Yet motion prediction is complicated due to the large variability in respiratory patterns, invalidating the assumption of periodic motion [4]. Unfortunately, observation of the tumor motion in real-time during therapy is often impossible, making surrogate measures of this motion essential. Common surrogates include measurements from a breathing bellow, a spirometer, optical imaging of the chest wall and imaging of the diaphragm position. The latter could be achieved by fluoroscopy, ultrasound or cine MRI, with ultrasound being the preferred choice to avoid radiation and high costs.

The question remains how well the surrogates can be related to the tumor motion. Studies have shown that assuming a linear relationship between the abdominal and tumor displacements can be misleading (correlation range

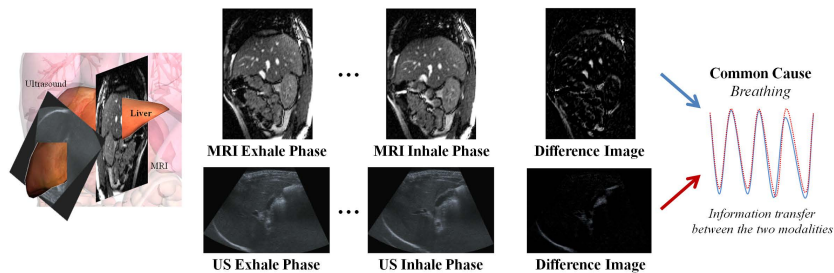
---

\* We acknowledge the Swiss National Science Foundation - Project CRSII2 127549.

[0.39 – 0.99]) [4]. Statistical motion models [8, 15] aim to overcome this uncertainty by learning the motion pattern for the whole organ. Tumor motion is then predicted from the motion of some landmarks in the organ. This requires that spatial correspondence between the surrogate and the organ motion model can be established, which is essentially a multi-modal image registration problem.

Relating the US structures back to the pre-therapeutic MR image is not an easy task, especially in 2D. A main challenge is to define an effective image similarity measure. For liver US images, approaches include gradient images [9], matching of extracted vessels [13, 11], simulations of US images from CT images [17, 5] and hybrids [6]. Assuming the same breathing phase, alignments are often optimized with respect to a rigid transformation. The reported registration errors were on average in the range of 3.7 mm. Additionally, tracking of US features are neither without error (e.g. 1.5 mm [2, 3] for 3D US). Processing time is another issue in real-time tracking.

For a clinical setup, we propose to acquire a short pre-therapeutic 4D MRI, and record US images of the same organ during treatment. The organ under investigation is then the same before and during therapy, while the position of the image acquisition and modality have changed, see Fig. 1. Instead of multi-modal image registration of two static images, we propose a novel approach. It is based on the observation, that the appearance changes of the organ in time are mainly due to a common cause, i.e. the breathing. Inspired by recent investigation of learning invariants in time series [18], we look for low-dimensional embeddings of the image sequences in order to extract such invariant signals, which are associated to a common cause. Invariance in this context refers to independence of multi-modal image appearance and exact slice location. Summarizing, our hypothesis is that due to the presence of a global temporal pattern in the data, it is possible to relate multi-modal images of a moving organ without the need for explicit image registration.



**Fig. 1.** **Left:** MR and US images of the liver are acquired at different spatial locations (e.g. sagittal MRI, right subcostal oblique US). **Center:** typical liver images from simultaneously acquired 2D MR (top) US (bottom) in exhale and inhale position are depicted. **Right:** We aim for extracting, from the image sequences, a description of the breathing as common cause, invariant of the image modality and slice location, in order to transfer image information between the two modalities.

## 2 Material

Ultrasound and MR images of the liver were simultaneously acquired for 3 volunteers during free-breathing at the Radiology Department of Geneva University Hospital [12]. This was accomplished by modifying the ultrasound equipment to be MR compatible. The ultrasound transducer and cables were EM shielded, coupling material was added and the transducer finally mounted on a fixation frame. The sequence of 2D ultrasound images obtained from real time US 2nd-harmonic imaging (center frequency = 2.2 MHz) had a temporal resolution of 25 fps and a spatial resolution of 0.6 mm. The acquisition sequence of MRIs consisted of alternating between 2D navigator slices (at a fixed spatial position) and 2D data slices (at different locations to cover the liver) with a frequency of 2.45 Hz for the first two volunteers and 2.86 Hz for the third volunteer. From the aforementioned acquisition sequence, 4D MR images can be created by sorting the data slices based on the similarity of the enclosing navigator slices [14]. This allows us to process 2D navigators and extend our results to 3D MR volumes. MR navigator images had a spatial resolution of 2.34 mm, 2.42 mm and 1.72 mm for the first, second and third volunteer respectively.

## 3 Method

In this work we propose a new approach for exploiting the temporal commonality of multi-modal images (US, MRI) acquired from the liver during free-breathing. We assume that the breathing signal is statistically similar across different image modalities that represent the same organ. Hence, we extract the breathing signal from each modality, look for the similarity between these signals, and use the resulting match to predict the corresponding images. In particular, we predicted MR navigators from the observation of US images. Our method (illustrated in Fig. 2 and described in more details in Sects. 3.1-3.3) consists of 3 steps, namely

**Dimensionality Reduction.** Given an image sequence  $I^m$  from modality  $m \in \{MRI, US\}$ , we embed  $I^m$  into a low-dimensional data representation  $S^m$  using dimensionality reduction techniques.

**Selection of Common Cause Signals.** We select components from  $S^{MRI}$  and  $S^{US}$  characterized by a global repetitive pattern related to breathing.

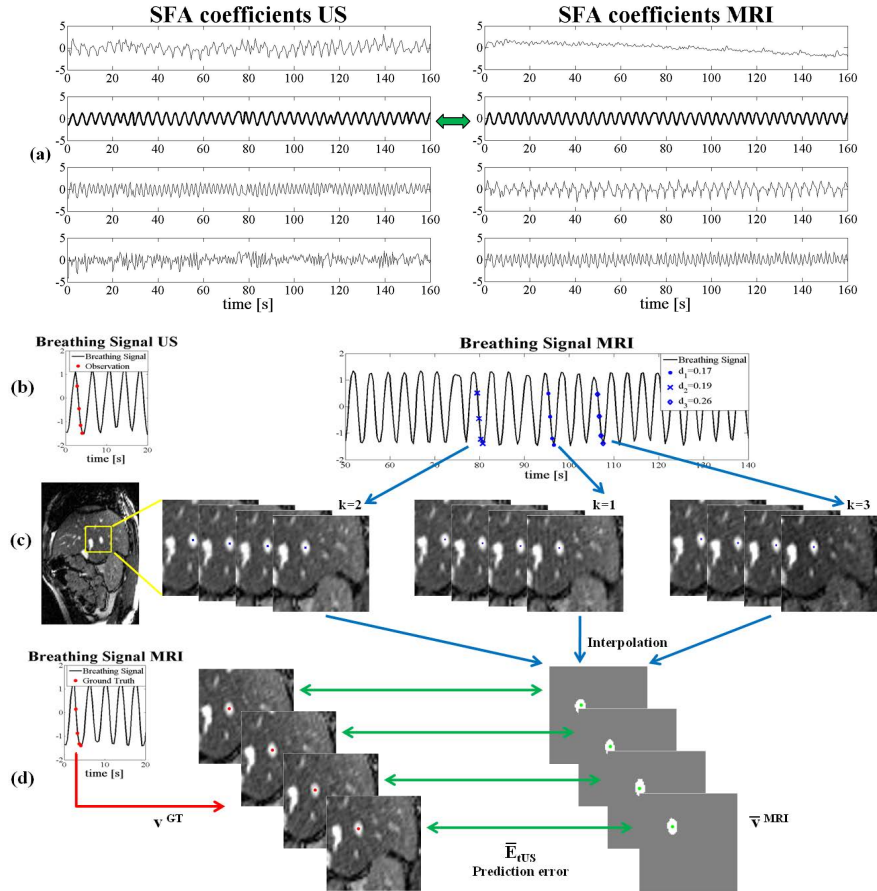
**Prediction of Feature Location.** The prediction of image feature locations in one modality is achieved by evaluating the similarity between the breathing signals and selecting the associated images.

### 3.1 Dimensionality Reduction

We aim to find an invariant description of the breathing signal with respect to the image modality and independently from an exact spatial correspondence. We assume that the breathing signal lies in a low-dimension manifold extracted from the data. Numerous methods are available for dimensionality reduction [7].

However it is often unclear which method is applicable for a specific problem. We explored the following dimensionality reduction methods: Principle Component Analysis (PCA), Slow Feature Analysis (SFA) [18], Isomaps [16] and Laplacian eigenmaps [1]. Especially, SFA was proposed to capture the invariant temporal structures in the data by extracting slowly to quickly changing components [18].

For each modality  $m \in \{MRI, US\}$ , we acquire images characterized by  $D^m$  pixels at a frequency rate of  $f^m$ . We consider a time sequence of  $L^m$  images  $\mathbf{I}^m(\tau_i^m)$ , with  $\tau_i^m = \tau_0^m + i/f^m$  for  $0 \leq i \leq L^m - 1$ . Dimensionality reduction methods transform the initial dataset  $\mathbf{I}^m(\tau_i^m) \in \mathbb{R}^{D^m}$  into a new dataset  $\mathbf{S}^m(\tau_i^m) = [\mathbf{s}_1^m(\tau_i^m); \dots; \mathbf{s}_d^m(\tau_i^m)] \in \mathbb{R}^d$ , with  $d \ll D^m$ . We used the dimension-



**Fig. 2.** (a) Low-dimensional representation of the images and selection of the breathing signals ( $\hat{\mathbf{s}}^{US}$ ,  $\hat{\mathbf{s}}^{MRI}$ ). (b) From the observation of a short sequence of  $\hat{\mathbf{s}}^{US}$ , we find the most similar sequences in  $\hat{\mathbf{s}}^{MRI}$ . (c) We select the associated MRIs and extract image information (vessel landmarks). (d) For evaluation, we interpolate the vessel landmarks ( $\bar{\mathbf{v}}^{MRI}$ ) and calculate the prediction error  $\bar{\mathbf{E}}_{tUS}$  w.r.t. ground truth ( $\mathbf{v}^{GT}$ ).

ality reduction toolbox from Laurens van der Maaten [7] for manifold learning of Isomaps and the Laplacian eigenmaps, and Matlab for PCA and SFA.

**PCA.** Each image  $\mathbf{I}^m(\tau_i^m)$  is reshaped into a single column vector  $\mathbf{x}^m(\tau_i^m) \in \mathbb{R}^{D^m}$ . We calculate the mean with respect to time ( $\bar{\mathbf{x}}^m$ ) and the covariance matrix  $\mathbf{C}^{\hat{\mathbf{X}}} = \hat{\mathbf{X}}^m \hat{\mathbf{X}}^{mT}$ , where the  $i$ -th column of  $\hat{\mathbf{X}}$  is equal to  $\mathbf{x}^m(\tau_i^m) - \bar{\mathbf{x}}$ . We then solve the eigenproblem  $\mathbf{C}^{\hat{\mathbf{X}}} \mathbf{w}_j^{PCA} = \lambda_j^{PCA} \mathbf{C}^{\hat{\mathbf{X}}} \mathbf{w}_j^{PCA}$ ,  $\forall j \in [1, \dots, L^m]$ .  $\lambda_j^{PCA}$  is the  $i$ -th largest eigenvalue and  $\mathbf{w}_j^{PCA}$  the corresponding eigenvector. In each  $j$ -th eigendirection we calculate the data projection  $\mathbf{p}_j^{PCA} = (\mathbf{w}_j^{PCA})^T \hat{\mathbf{x}}^m$ . We consider a set of  $d$  projections so that  $\mathbf{s}_j^m(\tau_i^m) = \mathbf{p}_j^{PCA}(\tau_i^m)$  for  $j \in [1, \dots, d]$ .

**SFA.** We perform SFA on the PCA projections in order to extract signals with increasing temporal frequency from  $\mathbf{P} = [\mathbf{p}_1^{PCA}; \dots; \mathbf{p}_J^{PCA}] \in \mathbb{R}^{J \times L^m}$ , with  $J \ll L^m$ . Let  $\hat{\mathbf{P}}$  be the zero-mean data matrix. We solve the generalized eigenproblem  $\dot{\mathbf{C}}^{\hat{\mathbf{P}}} \mathbf{w}_j^{SFA} = \lambda_j^{SFA} \mathbf{C}^{\hat{\mathbf{P}}} \mathbf{w}_j^{SFA}$ , where  $\mathbf{C}^{\hat{\mathbf{P}}}$  and  $\dot{\mathbf{C}}^{\hat{\mathbf{P}}}$  are the covariance matrix of the dataset and the temporal difference, respectively. The slowest components in  $\hat{\mathbf{P}}$  are the projections onto the eigenvectors  $\mathbf{w}_j^{SFA}$  associated with the smallest  $\lambda_j^{SFA}$  and  $\mathbf{s}_j^m(\tau_i^m) = (\mathbf{w}_j^{SFA})^T \hat{\mathbf{P}}^m$ ,  $\forall j \in [1, \dots, d]$ .

The individual components  $\mathbf{s}_i^m(\tau_i^m)$  are then normalized to zero mean and a standard deviation of one. The signals are further normalized to a common sampling frequency  $f = \min\{f^{MRI}, f^{US}\}$  (in our case  $f = f^{MRI}$ ). We denote the normalized low-dimensional descriptions as  $\hat{\mathbf{S}}^m(t) = [\hat{\mathbf{s}}_1^m(t); \dots; \hat{\mathbf{s}}_d^m(t)]$ , with  $t = \tau_0^m + i/f$  and  $0 \leq i \leq L^m - 1$ . Fig. 2(a) shows examples of  $\hat{\mathbf{S}}^{US}$  and  $\hat{\mathbf{S}}^{MRI}$ .

### 3.2 Selection of Common Cause Signals

We aim to select corresponding individual components in  $\hat{\mathbf{S}}^{US}$  and  $\hat{\mathbf{S}}^{MRI}$ . Relying on the statistical assumption that a similar breathing pattern is observed in the two modalities, we want to find the components which indicate the breathing as common cause. Hence, we employ frequency analysis (i.e. Fast Fourier Transform) to the 1D eigenmodes of both modalities. Then we choose the signals characterized by a power spectral density maximum at a frequency in the range of 0.15-0.40 Hz (common breathing frequency window). We denote the selected breathing signals as  $\hat{\mathbf{s}}^m(t)$  for modality  $m \in \{US, MRI\}$ .

### 3.3 Prediction of Feature Location

After normalization and selection of the breathing signals, we look for their similarity. We select the most similar short sequences in  $\hat{\mathbf{s}}^{MRI}$  from a finite number of observations in  $\hat{\mathbf{s}}^{US}$ , in order to predict MR navigators. Our observation consists of a short sequence  $[\hat{s}^{US}(t - (N - 1)/f), \dots, \hat{s}^{US}(t/f)]$  from the selected signal  $\hat{\mathbf{s}}^{US}$ , in order to take into account the breathing history (e.g. exhalation, inhalation). We calculate the  $K$ -nearest neighbors ( $K$ -NN) in  $\hat{\mathbf{s}}_{MRI}$  to this sequence.

In detail, we find  $t_k^{MRI}$  so that, for a given  $t^{US}$ , we minimize the distance

$$d_k = \sqrt{\sum_{n=0}^{N-1} [\hat{s}^{US}(t^{US} - n/f) - \hat{s}^{MRI}(t_k^{MRI} - n/f)]^2}, \quad (1)$$

$\forall 1 \leq k \leq K$ , see Fig. 2(b). We considered  $N = 4$  and  $K = 3$ . Having found the  $K$ -NN MR short sequences from the US sequence, we selected the associated MR navigator images, see Fig. 2(c).

We evaluated this approach on vessel center locations on the MR navigators. Specifically, vessels with cross-sectional cuts were semi-automatically segmented from MR navigator images<sup>3</sup> and their center locations  $\mathbf{v}(t)$  extracted. For each of the  $N$  frames, we linearly interpolate the  $K$  center locations, to get the prediction of the vessel center position (see Fig 2(d)):

$$\bar{\mathbf{v}}^{MRI}(t^{US} - n/f) = \sum_{k=1}^K \left(1 - \frac{d_k}{\sum_k d_k}\right) \mathbf{v}^{MRI}(t_k^{MRI} - n/f). \quad (2)$$

**Prediction Evaluation.** We used data from 3 volunteers (Vol.1 - Vol.3). For each volunteer we considered one US sequence and two MRI sequences and differentiate between two setups. For *Seq1* the US and MR sequences were acquired simultaneously. The MRIs of this sequence define our ground truth (GT) data, allowing us to validate the method. For *Seq2* we use MRIs acquired minutes later and the ultrasounds from *Seq1*. This setup was used to evaluate the performance for an independent MR set, similarly to the targeted clinical application.

We also assessed the effect of replacing  $\hat{\mathbf{s}}^{US}$  with the main diaphragm motion (first PCA eigenmode of a point displacement at the diaphragm). The diaphragm displacement was obtained from intensity-based tracking of US liver images. The tracking method was optimizing the parameters of an affine transformation of a region enclosing the diaphragm with respect to normalized cross correlation.

We quantified the prediction error by computing the mean of the Euclidean distance between the ground truth vessel center location  $\mathbf{v}^{GT}$  and the corresponding location on the predicted MRIs  $\bar{\mathbf{v}}^{MRI}$ , see Fig. 2(d):

$$\bar{E}_{t^{US}} = \frac{1}{N} \sum_{n=0}^{N-1} \|\mathbf{v}^{GT}(t^{US} - n/f) - \bar{\mathbf{v}}^{MRI}(t^{US} - n/f)\|. \quad (3)$$

We summarized the results by the mean and standard deviation of  $\bar{E}_{t^{US}} \forall t^{US}$ .

## 4 Results

Table 1 lists the mean prediction error  $\bar{E}$  obtained from each volunteer and tested method. It can be observed that SFA achieved on average lowest errors closely

<sup>3</sup> Sagittal navigator images were acquired. Hence out-of-plane motion is expected to be minor (2 mm) compared to the captured in-plane motion (15 and 8 mm) [10].

**Table 1.** Mean  $\pm$  SD of the prediction error (in mm), see Equation (3). Best results are in bold face and second best are underlined.

|             |             | common approach                   | our approach                      |                                   |                                   |                                   |
|-------------|-------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
|             |             | Tracking                          | Isomap                            | Lap.Eigenmap                      | PCA                               | SFA                               |
| <i>Seq1</i> | Vol. 1      | 3.91 $\pm$ 1.94                   | 2.86 $\pm$ 1.66                   | 3.73 $\pm$ 1.13                   | 1.89 $\pm$ 0.98                   | <b>1.72 <math>\pm</math> 0.93</b> |
|             | Vol. 2      | 3.00 $\pm$ 1.97                   | 4.36 $\pm$ 2.52                   | 3.08 $\pm$ 1.76                   | <u>2.98 <math>\pm</math> 1.72</u> | <b>2.89 <math>\pm</math> 1.63</b> |
|             | Vol. 3      | 4.02 $\pm$ 2.85                   | 4.64 $\pm$ 3.45                   | <u>3.39 <math>\pm</math> 2.45</u> | 3.57 $\pm$ 2.45                   | <b>3.35 <math>\pm</math> 2.37</b> |
|             | <b>Mean</b> | 3.65 $\pm$ 2.25                   | 3.96 $\pm$ 2.54                   | 3.40 $\pm$ 1.78                   | <u>2.81 <math>\pm</math> 1.72</u> | <b>2.65 <math>\pm</math> 1.65</b> |
| <i>Seq2</i> | Vol. 1      | 5.87 $\pm$ 3.60                   | 5.14 $\pm$ 2.38                   | 6.05 $\pm$ 2.59                   | <u>3.58 <math>\pm</math> 1.55</u> | <b>3.36 <math>\pm</math> 1.50</b> |
|             | Vol. 2      | <u>3.99 <math>\pm</math> 1.33</u> | <b>3.31 <math>\pm</math> 1.73</b> | 6.40 $\pm$ 3.41                   | 4.84 $\pm$ 2.64                   | 4.60 $\pm$ 2.57                   |
|             | Vol. 3      | <u>4.55 <math>\pm</math> 2.69</u> | 6.98 $\pm$ 5.07                   | 10.83 $\pm$ 6.77                  | <b>4.32 <math>\pm</math> 2.64</b> | 4.67 $\pm$ 2.88                   |
|             | <b>Mean</b> | 4.80 $\pm$ 2.54                   | 5.14 $\pm$ 2.73                   | 7.76 $\pm$ 4.26                   | <u>4.25 <math>\pm</math> 2.28</u> | <b>4.21 <math>\pm</math> 2.31</b> |

**Table 2.** Correlation between extracted signals and  $v^{GT}$ .

|                              |  | Vol. 1 | Vol. 2 | Vol. 3 | Mean  |
|------------------------------|--|--------|--------|--------|-------|
| Tracking - $\mathbf{v}^{GT}$ |  | 30.7%  | 84.4%  | 85.1%  | 66.7% |
| SFA                          | $\hat{\mathbf{s}}^{US} - \hat{\mathbf{s}}^{MRI}$ | 92.3%  | 79.6%  | 91.7%  | 87.9% |
|                              | $\hat{\mathbf{s}}^{MRI} - \mathbf{v}^{GT}$       | 87.6%  | 94.8%  | 87.6%  | 90.0% |

followed by PCA. Average errors increased by 1 mm (0.6 mm) for diaphragm tracking for *Seq1* (*Seq2*). When assuming no respiratory motion (using the mean position of  $\mathbf{v}^{GT}$  over *Seq1* as prediction), we obtained an average prediction error over all 3 volunteers of  $5.94 \pm 2.11$  mm ( $7.49 \pm 3.55$  mm) for *Seq1* (*Seq2*). For the best approach (SFA), we evaluated the linear dependence between the selected breathing signals  $\hat{\mathbf{s}}^{US}$  and  $\hat{\mathbf{s}}^{MRI}$ , and between  $\hat{\mathbf{s}}^{MRI}$  and  $\mathbf{v}^{GT}$ . We compared the correlation with that of the common prediction approach, i.e. diaphragm tracking to  $\mathbf{v}^{GT}$ , see Tab. 2. SFA provides on average higher correlations, supporting the better performance of SFA over diaphragm tracking for *Seq1* (Tab. 1).

## 5 Conclusion

We proposed an unconventional method for relating multi-modal images sequences. It is based on representing the acquired data in a low-dimensional embedding, extracting common cause signals (e.g. breathing) from both image sequences, finding the most similar sub-sequences of these signals, and using the associated images in order to predict the location of image features. The method generates the prediction in a completely unsupervised manner.

Using this approach, we predicted the location of anatomical landmarks by relying on a temporal pre-therapeutical MRI sequence based on the observation of US images acquired during therapy with an average accuracy of 4.2 mm. Our

performance is comparable to state-of-the-art methods (3.7 mm for multi-modal registration [11, 17, 6], 1.5 mm for US tracking [2, 3]) while being less complex. Lowest mean errors were achieved when employing SFA. This demonstrates the advantage of explicitly using the temporal information stored in the data and supports previous investigations [18]. After these encouraging results, we plan to apply the method to 3D landmark prediction by using 4D MR techniques [14], to acquire longer image sequences and to investigate potential improvements when combining the presented approach with statistical breathing models [15].

## References

1. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15(6), 1373–1396 (2003)
2. Foroughi, P., et al.: Intra-subject elastic registration of 3D ultrasound images. *Med. Image Anal.* 10(5), 713–725 (2006)
3. Harris, E., et al.: Speckle tracking in a phantom and feature-based tracking in liver in the presence of respiratory motion using 4D ultrasound. *PMB* 55, 3363 (2010)
4. Keall, P., et al.: The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med. Phys.* 33, 3874 (2006)
5. King, A., et al.: Registering preprocedure volumetric images with intraprocedure 3-d ultrasound using an ultrasound imaging model. *Med. Imaging, IEEE Trans. on* 29(3), 924–937 (2010)
6. Lange, T., et al.: 3D ultrasound-CT registration of the liver using combined landmark-intensity information. *IJCARS* 4(1), 79–88 (2009)
7. Van der Maaten, L., et al.: Dimensionality reduction: A comparative review. *TiCC TR* (2009-005)
8. McClelland, J., et al.: Inter-fraction variations in respiratory motion models. *PMB* 56, 251 (2011)
9. Milko, S., et al.: Evaluation of bivariate correlation ratio similarity metric for rigid registration of US/MR images of the liver. *IJCARS* 4(2), 147–155 (2009)
10. Nguyen, T., et al.: Adapting population liver motion models for individualized online image-guided therapy. *Proc EMBS* pp. 3945–3948 (2008)
11. Penney, G., et al.: Registration of freehand 3D ultrasound and magnetic resonance liver images. *Med. Image Anal.* 8(1), 81–91 (2004)
12. Petrusca, L., et al.: Ultrasound/MR hybrid imaging: truly simultaneous motion monitoring in the abdomen and image co-registration. In: *ISMRM* (2011)
13. Porter, B., et al.: Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial markers. *Med. Imaging, IEEE Trans. on* 20(4), 354–359 (2002)
14. Siebenthal, M., et al.: 4D MR imaging of respiratory organ motion and its variability. *PMB* 52, 1547 (2007)
15. Siebenthal, M., et al.: Inter-subject modelling of liver deformation during radiation therapy. In: *Proc. MICCAI 2007* (2007)
16. Tenenbaum, J., et al.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290
17. Wein, W., et al.: Integrating diagnostic B-mode ultrasonography into CT-based radiation treatment planning. *Med. Imaging, IEEE Trans. on* 26(6), 866–879 (2007)
18. Wiskott, L., Sejnowski, T.: Slow feature analysis: Unsupervised learning of invariances 14(4), 715–770 (2002)