

# Gated-tracking: Estimation of Respiratory Motion with Confidence

Valeria De Luca\*, Gábor Székely, and Christine Tanner

Computer Vision Laboratory, ETH Zürich, 8092 Zürich, Switzerland  
vdeluca@vision.ee.ethz.ch

**Abstract.** Image-guided radiation therapy during free-breathing requires estimation of the target position and compensation for its motion. Estimation of the observed motion during therapy needs to be reliable and accurate. In this paper we propose a novel, image sequence-specific confidence measure to predict the reliability of the tracking results. The sequence-specific statistical relationship between the image similarities and the feature displacements is learned from the first breathing cycles. A confidence measure is then assigned to the tracking results during the real-time application phase based on the relative closeness to the expected values. The proposed confidence was tested on the results of a learning-based tracking algorithm. The method was assessed on 9 2D B-mode ultrasound sequences of healthy volunteers under free-breathing. Results were evaluated on a total of 15 selected vessel centers in the liver, achieving a mean tracking accuracy of 0.9 mm. When considering only highly-confident results, the mean (95th percentile) tracking error on the test data was reduced by 12% (16%) while duty cycle remained sufficient (60%), achieving a 95% accuracy below 3 mm, which is clinically acceptable. A similar performance was obtained on 10 2D liver MR sequences, showing the applicability of the method to a different image modality.

**Keywords:** confidence, tracking, learning, respiratory motion, image guidance, ultrasound.

## 1 Introduction

Image-guided radiation therapy of abdominal organs during free-breathing requires estimation of the target position and compensation for its motion over the duration of the entire treatment session [9]. Examples of imaging techniques available for observing the internal motion are ultrasound (US) and magnetic resonance imaging (MRI). US is increasingly used, as it can image soft tissues in real time, and with higher temporal and spatial resolution than other modalities, is non-ionizing and inexpensive [4]. MRI is a popular choice for guidance of focused ultrasound surgery as it provides thermometry images, also used for tracking [15], yet the temporal resolution is about 5 times lower than US.

---

\* We thank the European Union's 7th Framework Program (FP7/2007-2013) under grant agreement no. 611889 (TRANS-FUSIMO) for funding.

The intrafraction motion estimation needs to be reliable and accurate, to reduce the size of the safety margins due to expected residual tumor motion in the planned dose delivery [19]. Gating is commonly used to reduce motion uncertainties. Yet it generally leads to relatively low duty cycles (20-50%) for maintaining acceptable accuracy [9,16]. Following the tumor motion by tracking keeps duty cycle at 100%, but can have high errors. Here we propose gated-tracking, where gating (i.e. accepting motion predictions) is based on the estimated accuracy (confidence) of the current tracking result and tracking is used to follow the tumor.

Confidence measures for motion estimation from images can generally be based on image appearance, motion field properties, or both. Estimates are likely more accurate for strong image features, high image similarity and realistic motion patterns. Usually maximization of these criteria has been incorporated in the tracking approach to increase accuracy. For example, methods based on optical flow [10,2,11] and probabilistic approaches [18,13] include already confidence values in their tracking formulation. Yet they are tailored for the specific tracking strategy. Learning-based methods have been employed for detecting the areas of the images which can be most reliably tracked [1,6,8]. A probabilistic confidence measure, based on the local image similarity, was provided for a block matching method [13]. Consistency of motion estimates across groups of image has been used in image registration to detect errors and improve accuracy [5]. However such a strategy is too time-consuming for real-time applications.

We propose a different approach for assigning confidence measures, which is independent from the motion estimation strategy and image modality, and is based on the relationship between image similarity and estimated displacement vectors. This method exploits the recurring similarities of both motion and image appearance in pseudo-repetitive scenarios, like breathing. The aim is to keep the tracking errors below 5 mm, to fulfill the clinical requirements for target margins in radiation therapies [9]. Further reductions will facilitate to keep the overall system error within this limit, when also having to incorporate spatio-temporal motion prediction [12]. A strategy was devised to automatically determine the required thresholds of the confidence values to stay within these limits.

## 2 Materials

We collected nine 2D US sequences of the liver of healthy volunteers under free-breathing, from the same spatial location. In details, second harmonic B-mode US sequences were acquired over 3-10 min, using a Siemens Antares scanner (CH4-1 transducer, center frequency 1.82 - 2.22 MHz) [14]. These images (longitudinal or intercostal planes) are characterized by spatial and temporal resolution ranging in [0.28, 0.71] mm and [14, 25] Hz, and size in [82632, 279300] pixels. To explore the validity of the method independently of the image modality, we gathered 2D MR sequences from 10 healthy volunteers under free-breathing. These consisted of sagittal navigator slices, acquired at 2.6-3.4 Hz for 4D-MRI [17]. The MRIs were captured using a balanced Steady State Free Precession sequence (flip angle  $70^\circ$ , TR=3.1 ms) on a 1.5T Philips Achieva MR system. Image resolution and size ranged in [1.29, 1.37] mm and [50176, 65536] pixels.

Given a sequence of  $T$  2D images  $I(t)$ , with  $t \in [0, T - 1]$ , the tracking objective is to compute the position of anatomical landmarks inside the liver (vessel centers)  $P_j(t) \in \mathbb{R}^2$  in each image.  $ROI_{j,\tilde{t}}(t)$  and  $\mathbf{d}_j(t) = P_j(t) - P_j(0)$  denote a rectangular region of interest from  $I(\tilde{t})$  centered in  $P_j(t)$  and the displacement of the landmark with respect to the initial frame, respectively.

### 3 Method

We propose a novel, sequence-specific confidence measure to predict the reliability of tracking results. The sequence-specific statistical relationship between the image similarities and the feature displacements is learned from the first breathing cycles. A confidence measure is then assigned to the tracking results during the real-time application phase based on the relative closeness to the expected values. The proposed measure was tested on the results of two tracking methods and two image modalities: an automatic learning-based tracking algorithm for US images; and a deformable registration for MRIs.

#### 3.1 Tracking Algorithm

For US, we used the learning-based block matching algorithm (LB-BMA) proposed in [3], which exploits the pseudo-repetitive nature of respiration. During an initial training phase of 10 breathing cycles the images are registered and the relationship between the image appearance and the displacements is learned. Image registration is performed non-iteratively by optimizing a local affine transformation with respect to normalized cross-correlation (NCC). The transformation is defined for  $ROI_{j,0}(0)$ , which is automatically extracted by finding the blob-like feature centered at  $P_j(0)$  [3] and expanding it by 3 times its size. The  $ROI_j$  image appearance for the training set is efficiently stored by dimensionality reduction via principal component analysis (PCA). For each landmark in the real-time application phase, the most similar ROI from the training set is selected based on the distance in the PCA space and used for temporal alignment of the BMA. For previously unseen image variations (outliers), affine registration is performed to stay adaptive to non-periodic motions. For tracking the MR slices we used the deeds registration method [7], which can cope with sliding boundaries, is relatively fast ( $\approx 0.5$  s per slice) for a deformable registration and publicly available.

#### 3.2 Confidence Measure

Independently from the tracking algorithm, the confidence measure is learned from an initial set of  $T_{in}$  images  $I(t')$  in a sequence, with  $t' \in [0, T_{in} - 1]$ , where we extract the tracked position of  $J$  landmarks  $P_j(t')$ . For each landmark, we compute the magnitude of the Euclidean distances to the initial frame  $d_j(t') = \|\mathbf{d}_j(t')\|$  and normalize it to the range of the distances from the  $T_{in}$  frames,  $\bar{d}_j(t')$ .

We also compute NCC between the initial  $ROI_{j,0}(0)$  and the same sized and positioned, non-transformed ROI at  $t'$ , i.e.  $NCC_j(t') = NCC(ROI_{j,0}(0), ROI_{j,0}(t'))$ .

We collect all 2D data points  $N_j(t') = [\bar{d}_j(t'), NCC_j(t')]$  for the training data and fit a polynomial curve  $F(\mathbf{c}) = \sum_{i=0}^n \alpha_i \mathbf{c}^i$  of degree  $n \in \{1, 2, 3\}$  to the 5%-95% of this distribution, with  $\alpha_i$  the estimated parameters and  $\mathbf{c}$  the vector of  $\mathbf{c}(t') = \bar{d}_j(t')$ . Degree  $n$  is selected to minimize the sum of squared errors and maximize the coefficient of determination of the fit, while avoiding zero-crossing of the  $\alpha_i$  confidence bounds and hence overfitting. We also extract the 99% confidence bounds ( $F^{99}(\mathbf{c})$ ) of the fit via the Matlab `fit` function. An example of  $F(\mathbf{c})$  is shown in Fig. 1.

For each  $P_j(\tau)$  during the application phase ( $\tau \in [T_{in}, T - 1]$ ), we calculated  $N_j(\tau) = [\bar{d}_j(\tau), NCC_j(\tau)]$  and find its projection to  $F(\mathbf{c})$  with minimum distance  $p_j(\tau) = \|N_j(\tau) - F(\mathbf{c}^*)\|$ , where  $\mathbf{c}^* = \operatorname{argmin}_{\mathbf{c}} \|p_j(\tau, \mathbf{c})\|$ . If the distance  $p_j(\tau)$  of the data point  $N_j(\tau)$  to  $F(\mathbf{c})$  is within the 99% confidence bounds and  $\bar{d}_j(\tau)$  does not exceed the observed  $\bar{d}_j(t')$  (to avoid extrapolation), the confidence measure  $C_j(\tau) \in [0, 1]$  is assigned according to the local 99% confidence as follows:

**If**  $p_j(\tau) < F^{99}(\mathbf{c}^*)$  **and**  $\bar{d}_j(\tau) \leq \max_{t'} \bar{d}_j(t')$  **do**  $C_j(\tau) = 1 - 2p_j(\tau)/F^{99}(\mathbf{c}^*)$   
**else do**  $C_j(\tau) = 0$

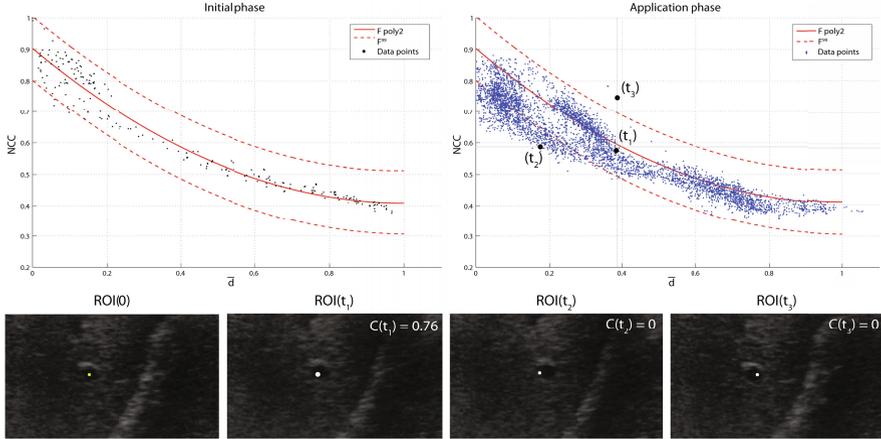
### 3.3 Evaluation

Quantitative evaluation of the tracking and the gated-tracking method was performed on a total of 15 selected vessel centers inside the liver from the US data and 20 from the MRIs. We randomly selected 10% of the application images in each sequence and annotated the landmark position  $\hat{P}_j(\hat{t})$  corresponding to  $P_j(0)$ , and computed the tracking error  $TE_j(\hat{t}) = \|P_j(\hat{t}) - \hat{P}_j(\hat{t})\|$ . Results were summarized by the mean (MTE), standard deviation (SD) and the 95th percentile ( $TE^{95}$ ) of the single distribution including all  $TE_j(\hat{t})$  belonging to the data subset. We also determined the motion magnitude  $M_j(\hat{t}) = \|P_j(0) - \hat{P}_j(\hat{t})\|$ .

The confidence values  $C_j(t) \in [0, 1]$  were divided into  $S = 20$  intervals  $CI_s = [s, 1]$ , with  $s \in \{0, 0.05, 0.10, \dots, 0.95\}$ . We computed  $S$  tracking errors based on accepting results with confidence  $C_j(t) \in CI_s$ . For each  $CI_s$  we define its duty cycle (DC) as the ratio between the number of accepted tracking results ( $T_{j,s}$ ) and the total number of frames in the application phase:  $DC_j = T_{j,s}/(T - T_{in})$ .

In order to select one confidence interval to apply during the application phase to all sequences, we split the US data into two similar sized subsets for training and testing. Subset  $\text{Set}_1^{US}$  included 7 landmarks from 5 sequences and subset  $\text{Set}_2^{US}$  consisted of the remaining 8 points from 4 sequences. Similarly, we annotated 2 vessel landmarks per MR sequence and selected alternately the sequences in order of acquisition for the training and test set, e.g. training on volunteers  $\{1, 3, \dots, 9\}$  ( $\text{Set}_1^{MR}$ ), testing on volunteers  $\{2, 4, \dots, 10\}$  ( $\text{Set}_2^{MR}$ ).

For the training data we selected  $s^*$ , which minimized the 95th percentile error ( $TE_{j,s^*}^{95}$ ) while keeping  $DC_{j,s^*} > 60\%$ . During gated-tracking we considered only the frames whose confidence values are within the selected interval  $CI_{s^*}$ .



**Fig. 1.** (Top) example of confidence curve  $F(\mathbf{c})$  and 99% bounds  $F^{99}(\mathbf{c})$  obtained from training data points (in black) for a representative sequence. (Bottom) example of tracking results in  $ROI_{j,0}$  for 3 points with different confidence values, showing  $P(t_1)$  with similar NCC as  $P(t_2)$ , and  $P(t_1)$  with similar distance  $\bar{d}$  as  $P(t_3)$ . The initial image  $ROI_{j,0}(0)$ , with point-landmark  $P(0)$  in yellow, is shown as a reference.

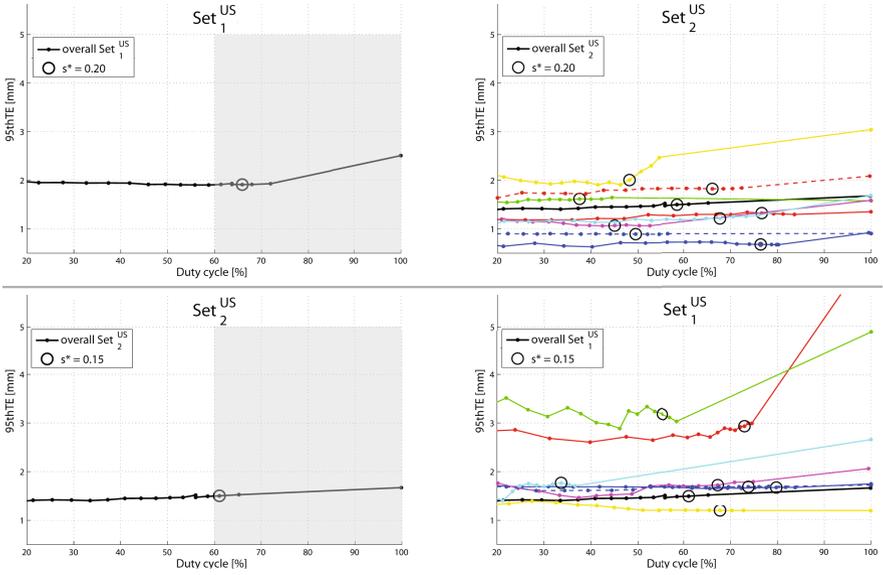
Two experiments were conducted, where the role of the subsets as training and test set was swapped to assess the sensitivity of the results to the data selection.

Finally, we compared the gated-tracking results to the ones of gating, where we accepted tracking results with small displacement, i.e.  $d_j(\tau) < 3$  mm.

## 4 Results

For US, we used 5 breathing cycles (260–453 frames) for  $T_{in}$ . The number of application frames ranged in [2233, 14211]. The  $MTE \pm SD$  ( $TE^{95}$ ) for all landmarks was  $0.90 \pm 0.77$  (2.04) mm. MTE ranged in [0.38, 2.34] mm, while the  $TE_j^{95}$  was in [0.90, 6.57] mm. For these landmarks the motion magnitude  $M$  was  $5.17 \pm 3.21$  mm, and the 95th percentile was 10.59 mm. Fig. 1 shows the relationship between motion and image similarity for (left) the initial phase and (right) the application phase. It can be observed that some of the application phase data falls outside the confidence region and will be discarded first. The example images illustrate outliers with respect to motion and image similarity.

The overall 95th percentile errors in comparison to duty cycle values can be seen for the training set on the left graphs in Fig. 2. The lowest  $TE^{95}$  in the objective region was 1.91 mm (-24%) and 1.50 mm (-11%) for  $Set_1^{US}$  and  $Set_2^{US}$  respectively, see Table 1. Using the resulting  $s^*$  for gated-tracking on the test set provided the same  $TE^{95}$  reduction, while keeping overall duty cycle at 59% and 68%. Gating results showed similar or worse  $TE^{95}$ , but  $DC < 47\%$ .



**Fig. 2.** 95th percentile of TE vs. duty cycle for (left) training set with determined optimum confidence interval  $CI_{s^*} = [s^*, 1]$  and (right) test set with sequence-specific curves (in colors) and indicated result for  $CI_{s^*}$  (circles).

**Table 1.** Summary of the results for the two **US** subsets ( $Set_1^{US}$  and  $Set_2^{US}$ ) used interchanged as test and training set. Results of the TE are in mm, while DC is in %.

	Tracking	Gated-tracking trained on $Set_1^{US}$ CI = [0.20, 1]	Gated-tracking trained on $Set_2^{US}$ CI = [0.15, 1]	Gating	
$Set_1^{US}$	MTE $\pm$ SD	1.10 $\pm$ 0.93	0.97 $\pm$ 0.56	0.97 $\pm$ 0.57	1.23 $\pm$ 1.20
	TE <sup>95</sup>	2.51	1.91	1.91	3.64
	range MTE <sub>j</sub>	[0.65, 2.34]	[0.70, 1.55]	[0.69, 1.58]	[0.58, 3.14]
	range TE <sub>j</sub> <sup>95</sup>	[1.18, 6.57]	[1.19, 3.24]	[1.18, 3.18]	[1.07, 6.89]
	DC	100	65.90	67.96	46.15
	range DC <sub>j</sub>	[100, 100]	[31.82, 77.61]	[33.66, 79.75]	[15.30, 68.38]
$Set_2^{US}$	MTE $\pm$ SD	0.73 $\pm$ 0.54	0.62 $\pm$ 0.44	0.63 $\pm$ 0.45	0.62 $\pm$ 0.47
	TE <sup>95</sup>	1.68	1.49	1.50	1.52
	range MTE <sub>j</sub>	[0.38, 1.06]	[0.33, 1.02]	[0.33, 1.02]	[0.35, 1.03]
	range TE <sub>j</sub> <sup>95</sup>	[0.90, 3.04]	[0.67, 2.00]	[0.67, 2.18]	[0.88, 2.28]
	DC	100	58.57	61.06	38.02
	range DC <sub>j</sub>	[100, 100]	[37.53, 76.66]	[39.98, 79.21]	[16.12, 53.72]

Due to the lower temporal resolution of the MRIs, we considered 25 breathing cycles ( $T_{in} \in [141, 519]$  frames). The following 1160 frames (7 min) were used for testing. For a mean motion amplitude of  $5.63 \pm 5.12$  mm, tracking results

**Table 2.** Summary of MR results as in Table 1.

		Tracking	Gated-tracking trained on $\text{Set}_1^{MR}$ CI = [0.35, 1]	Gated-tracking trained on $\text{Set}_2^{MR}$ CI = [0.55, 1]	Gating
$\text{Set}_1^{MR}$	<b>MTE±SD</b>	1.14±1.75	1.03±1.40	1.05±1.42	1.66±0.95
	<b>TE<sup>95</sup></b>	2.99	2.31	2.38	3.24
	<b>DC</b>	100	81.12	64.56	41.24
$\text{Set}_2^{MR}$	<b>MTE±SD</b>	0.96±0.91	0.90±0.74	0.88±0.70	1.63±0.87
	<b>TE<sup>95</sup></b>	2.21	2.07	1.95	3.17
	<b>DC</b>	100	83.21	67.79	39.28

achieved an overall accuracy of  $1.05\pm 1.38$  mm ( $\text{TE}^{95}=15.69$  mm). Similarly to US, gated-tracking trained on the same subset resulted in an improvement of the  $\text{TE}^{95}$  by 23% and 12% for  $\text{Set}_1^{MR}$  and  $\text{Set}_2^{MR}$  respectively, see Table 2. On the test set, the 95th percentile error reduction was 20% and 6%, with overall DC of 65% and 83%. Gating resulted in higher errors and lower DCs ( $\approx 40\%$ ).

The mean run-time of the confidence estimation is 0.2 ms/frame, with Matlab implementation on a single PC with Intel®Core™i7 CPU (2.66 GHz, 8 GB).

## 5 Conclusion

We proposed a method for estimating the confidence of tracking results for pseudo-repetitive motion scenarios. It is based on learning the sequence-specific relationship between image similarity and motion vectors, based on the assumption that during the initial learning phase the tracking errors are low. The confidence prediction is independent from the tracking method and image modality. The actually applied confidence interval is decided based on training data. Tested on US sequences, the method reduced 95th percentile errors by between 11% and 24% (to below 1.7 mm on average) while keeping duty cycle at 59% and 68%. Similar results were obtained for the MR sequences, with a reduction of the 95th percentile errors between 6% and 20% while keeping duty cycle at 65% and 83%. For both scenarios, performance was improved by gated-tracking over gating.

In future work we will verify the method’s general applicability to other tracking algorithms and larger datasets. For slow varying drift of the target, we envision to update the learning phase. Integration of the estimated confidence values in an overall motion prediction system is the long term plan to facilitate robust and reliable therapy guidance.

## References

1. Avidan, S.: Ensemble tracking. IEEE Trans. Pattern Anal. Mach. Intell. 29(2), 261 (2007)

2. Brox, T., Bregler, C., Malik, J.: Large displacement optical flow. In: Proc. IEEE Int. Conf. Comp. Vis. Pattern Recognit., p. 41 (2009)
3. De Luca, V., Tschannen, M., Székely, G., Tanner, C.: Learning-based approach for fast and robust vessel tracking in long ultrasound sequences. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part I. LNCS, vol. 8149, pp. 518–525. Springer, Heidelberg (2013)
4. Fontanarosa, D., van der Meer, S., Bamber, J., Harris, E., O’Shea, T., Verhaegen, F.: Review of ultrasound image guidance in external beam radiotherapy. *Phys. Med. Biol.* 60, R77 (2015)
5. Gass, T., Szekely, G., Goksel, O.: Detection and correction of inconsistency-based errors in non-rigid registration. In: SPIE Medical Imaging, p. 90341B (2014)
6. Grabner, H., Leistner, C., Bischof, H.: Semi-supervised on-line boosting for robust tracking. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 234–247. Springer, Heidelberg (2008)
7. Heinrich, M., Jenkinson, M., Brady, M., Schnabel, J.: MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Trans. Med. Imag.* 32, 1239 (2013)
8. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 34(7), 1409 (2012)
9. Keall, P.J., Mageras, G.S., Balter, J.M., Emery, R.S., Forster, K.M., Jiang, S.B., Kapatoes, J.M., Low, D.A., Murphy, M.J., Murray, B.R., Ramsey, C.R., Van Herk, M.B., Vedam, S.S., Wong, J.W., Yorke, E.: The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med. Phys.* 33, 3874 (2006)
10. Kondermann, C., Mester, R., Garbe, C.S.: A statistical confidence measure for optical flows. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 290–301. Springer, Heidelberg (2008)
11. Kybic, J., Nieuwenhuis, C.: Bootstrap optical flow confidence and uncertainty measure. *Comput. Vis. Image. Und.* 115, 1449 (2011)
12. McClelland, J., Hawkes, D.J., Schaeffter, T., King, A.P.: Respiratory motion models: a review. *Med. Image Anal.* 17, 19 (2013)
13. Patras, I., Hendriks, E.A., Lagendijk, R.L.: Probabilistic confidence measures for block matching motion estimation. *IEEE Trans. Circuits Syst. Video Technol.* 17(8), 988 (2007)
14. Petrusca, L., Cattin, P., De Luca, V., Preiswerk, F., Celicanin, Z., Auboiroux, V., Viallon, M., Arnold, P., Santini, F., Terraz, S., Scheffler, K., Becker, C.D., Salomir, R.: Hybrid Ultrasound/Magnetic Resonance Simultaneous Acquisition and Image Fusion for Motion Monitoring in the Upper Abdomen. *Invest. Radiol.* 48, 333 (2013)
15. Roujol, S., Moonen, C., De Senneville, B.D.: Motion correction techniques for MR-guided HIFU ablation of abdominal organs. *Frontiers of Medical Imaging*, 355 (2014)
16. Shirato, H., Shimizu, S., Kitamura, K., Onimaru, R.: Organ motion in image-guided radiotherapy: lessons from real-time tumor-tracking radiotherapy. *Int. J. Clin. Oncol.* 12, 8 (2007)
17. von Siebenthal, M., Szekely, G., Gamper, U., Boesiger, P., Lomax, A., Cattin, P.: 4D MR imaging of respiratory organ motion and its variability. *Phys. Med. Biol.* 52, 1547 (2007)
18. Simoncelli, E.P., Adelson, E.H., Heeger, D.J.: Probability distributions of optical flow. In: Proc. IEEE Int. Conf. Comp. Vis Pattern Recognit., p. 310 (1991)
19. Van Herk, M.: Errors and margins in radiotherapy. *Seminars in Radiation Oncology* 14, 52 (2004)