

A Learning-Based Approach for Fast and Robust Vessel Tracking in Long Ultrasound Sequences

Valeria De Luca^{*}, Michael Tschannen, Gábor Székely, and Christine Tanner

Computer Vision Laboratory, ETH Zürich, 8092 Zürich, Switzerland

Abstract. We propose a learning-based method for robust tracking in long ultrasound sequences for image guidance applications. The framework is based on a scale-adaptive block-matching and temporal realignment driven by the image appearance learned from an initial training phase. The latter is introduced to avoid error accumulation over long sequences. The vessel tracking performance is assessed on long 2D ultrasound sequences of the liver of 9 volunteers under free breathing. We achieve a mean tracking accuracy of 0.96 mm. Without learning, the error increases significantly (2.19 mm, $p < 0.001$).

Keywords: tracking, block-matching, learning, real-time, ultrasound.

1 Introduction

During conformal radiation therapies, motion in the treatment region needs to be compensated to ensure accuracy of the dose delivery. For the thorax and abdomen, motion due to respiration is substantial and can not be neglected [11,21]. Image-guided radiation therapies use image information gathered during therapy for adjusting the treatment plan. Tracking the respiratory motion on such images requires an accuracy in the millimeter range and real-time feedback. Potential imaging techniques for guidance include CT, MRI and ultrasound (US). The latter represents the only modality that is real-time, non-ionizing and cheap.

Structures on US sequences of the abdomen have been tracked using optical flow [5], speckle tracking algorithms [8], intensity-based registration [22], active contours [23], hybrid methods [6,13,3] and US imaging models [12]. Block-matching algorithms (BMAs) compute the local displacements from interpolating the translations that provide the best match of image regions in two consecutive frames. Many BMAs have been proposed (e.g. [1,19,14,2,8]), yet their performance has so far only been assessed on relative short sequences (< 1 min). Therapy guidance requires the tracking of long sequences, which poses a special challenge for BM due to its iterative nature. Moreover, our sequences suffer from noise, interferences, low SNR and frame dropouts. To create a robust framework for feature tracking in long sequences, we propose an algorithm, which combines several BM components and includes a novel adaptation of the block size to

^{*} We thank the Swiss National Science Foundation (CRSII2 127549) for funding.

the feature scale. In addition, we exploit the approximate periodicity of breathing motion for learning image appearance and corresponding motion behavior (extracted by accurate but slow image registration) to allow frequent temporal realignment of the BMA for drift-free real-time tracking.

2 Material

US liver sequences of 9 volunteers during free breathing were acquired at the Geneva University Hospital [17]. To evaluate US tracking performance for hybrid US and MR guided treatments [17], an Acuson clinical scanner (Antares; Siemens Medical Solutions, Mountain View, CA) was modified to be MR compatible, and US and MR images were simultaneously acquired. The US images (real-time second harmonic images with 1.8-2.2 MHz center frequency) were exported on-the-fly using a frame grabber device. 2D US images were acquired at a fixed location (longitudinal or intercostal plane) over 5:21, 5:28 and 10:08 min for 1, 7 and 1 volunteer(s), respectively. The resulting 2650 to 14516 frames had a temporal and spatial resolution of 14-25 Hz and 0.3-0.7 mm, respectively.

3 Method

3.1 Scale-adaptive Block-matching

The key components of our proposed scale-adaptive BMA (SA-BMA) are a novel adaptation of the block size to the feature scale and the new combination of the interpolation function from [14] with the temporal realignment from [19].

Block Configuration. Traditionally the size of the blocks is chosen empirically [16,7] or equal to the US speckle size [10]. We adapt the block size to the feature size to ensure that every block contains a part of the feature, which limits the aperture problem and avoids ambiguities due to homogeneous blocks.

The position of features to track, e.g. $P_j(t_0)$ for vessel j , are manually selected in $I(t_0)$, see Fig. 1. BM is performed for a region of interest ($ROI_j(t_0)$) around feature j , which covers a $M \times N$ grid of equally sized squares (called blocks) $B_{i,j}$ of size Δb_j with center points $G_{i,j}$, $i \in [1, \dots, MN]$, defined at t_0 . Δb_j is determined from the feature size. In detail, as vessel cross sections have elliptic shape, we search for blob-like features centered at P_j . A scale-space approach (local maxima of a Difference-of-Gaussian (DoG)) [15,20] is used to detect the most likely blob in $ROI_j(t_0)$. The resulting scale s is related to the minor semi-axis r_j of an ellipse fitted to the vessel section by $r_j = \sqrt{2s}$ and $\Delta b_j = \lceil r_j \rceil$.

Displacement Calculation. We compute the motion field in each ROI_j by determining the displacement at $G_{i,j}$ via BM, and use weighted interpolation [14] to obtain the displacement of P_j . At time step t^* the displacement of $G_{i,j}(t^{ref})$ in the reference frame t^{ref} to $G_{i,j}(t^*)$, denoted as $\mathbf{d}_{G_{i,j}}(t^*)$, is determined by the displacement \mathbf{v} which maximized the normalized cross-correlation (NCC) between $B_{i,j}(t^{ref})$ and the block from $I(t^*)$ centered at $G_{i,j}(t^{ref}) + \mathbf{v}$. The values

of \mathbf{v} are restricted to cover only a certain search region. The reference frame is generally the previous frame ($t^* - 1$). Other strategies for t^{ref} are described in the next paragraph. The displacement of the tracked point from t^{ref} to t^* ($\mathbf{d}_j(t^*)$) is deduced from the block displacements $\mathbf{d}_{G_{i,j}}(t^*)$ by weighted interpolation:

$$\mathbf{d}_j(t^*) = \sum_{\hat{i}} w_{\hat{i}} \mathbf{d}_{G_{\hat{i},j}}(t^*), \quad (1)$$

where $w_{\hat{i}}$ are the weights and $\hat{i} = \{i | Q(i, t^*) = 1\}$. $Q(i, t^*)$ is the filtering mask for ROI_j at time t^* , which is defined by $Q(i, t^*) = 1$ for the 9 $G_{i,j}(t^{ref})$ closest to $P_j(t^{ref})$, and $Q(i, t^*) = 0$ otherwise. We consider the weights $w_{\hat{i}}$ [14]:

$$w_{\hat{i}} = 0.5 \frac{1}{D_{\hat{i}}^2 + 1} \frac{1}{\sum_{\hat{i}} \frac{1}{D_{\hat{i}}^2 + 1}} + 0.5 \frac{\alpha_{\hat{i}}}{\sum_{\hat{i}} \alpha_{\hat{i}}}, \quad (2)$$

with $D_{\hat{i}}$ the Euclidean distance from $G_{\hat{i},j}(t^{ref})$ to $P_j(t^{ref})$, and $\alpha_{\hat{i}} = \sigma_{\hat{i}}^2 / \mu_{\hat{i}}$ the ratio between the variance ($\sigma_{\hat{i}}^2$) and the mean ($\mu_{\hat{i}}$) of the pixel intensities in $B_{\hat{i},j}(t^*)$. This interpolation scheme has the advantage that it incorporates regularization (first term) and accounts for the relative image content (second term) [14]. The position of the tracked point is $P_j(t^*) = P_j(t^{ref}) + \mathbf{d}_j(t^*)$.

Reference Frame Definition. BMAs can generally only cope with small deformations and appearance changes, as they are based on the translations of local regions. Hence BM is applied to temporally consecutive frames (i.e. $t^{ref} = t^* - 1$) for tracking. However, this strategy is subject to error accumulation leading to drift. Such errors are particularly relevant in long sequences. Yet the approximate periodic nature of respiratory motion provides frequently frames which are similar to the initial frame and BM is again applicable for aligning these [19]. Errors occur also due to the quantization of $\mathbf{d}_{G_{i,j}}$. Hence we introduce the following strategy:

```

if  $NCC(ROI_j(t_0), ROI_j(t^*)) > \theta_{NCC,j}$  then  $t^{ref} = t_0$ 
else if  $\|\mathbf{d}_j(t^*)\| \leq \epsilon_d$  then  $t^{ref} = t_{prev}^{ref}$ 
else  $t^{ref} = t^* - 1$  end

```

where $NCC(A, B)$ is the NCC between image region A and B , $\theta_{NCC,j}$ is the 84th percentile of the NCC values gathered from an initial subset of the sequence, $\epsilon_d = 0.01$ pixel, and t_{prev}^{ref} denotes t^{ref} from the previous image pair.

3.2 Learning-based Tracking

During therapy, images are acquired continuously over several minutes. Hence temporal realignment of the images is crucial to ensure robust tracking and to avoid error accumulation. For repetitive motion, such as breathing, redundancy within the images can be exploited [4]. Following a similar strategy, we divide the method into a training and tracking phase. During training we learn the relationship between image appearance and the corresponding displacements, from a slower, but more robust tracking method. During the clinical application,

the displacements are computed by the proposed SA-BMA (see Sec. 3.1), with the reference frame given by the closest frame from the training set. This strategy allows temporal realignment over many more breathing states than previously.

Training Phase. In the training phase we acquire a sequence covering 10 breathing cycle, resulting in T_{10C} images $I(t_i)$, $t_i \in [t_0, \dots, T_{10C}]$.

The images $I(t_i)$ are registered to $I(t_0)$, to obtain spatial correspondence. The registration optimizes the parameters of an affine transformation with respect to NCC over a manually selected region around $P_j(t_0)$ and is initialized by the result from $I(t_{i-1})$ to $I(t_0)$.

To store the image appearance efficiently, we embed the images $I(t_i) \in \mathbb{R}^D$ into a low-dimensional representation $\mathbf{S}(t_i) = [s_1(t_i); \dots; s_L(t_i)] \in \mathbb{R}^L$, with $L \ll D$, using Principal Component Analysis (PCA) [4]. We select L such that the cumulative energy of the first L eigenvectors just exceeds 95%. In addition, we select the PCA component \mathbf{s}_B in \mathbf{S} that captures the main breathing motion, by computing the FFT of each \mathbf{s}_i and choosing the one that has a power spectral density maximum at 0.15-0.4 Hz (2.5-6 s, common breathing). \mathbf{S} and the corresponding registration results (e.g. P_j) are stored $\forall t_i$.

Tracking Phase. New images are continuously acquired during treatment. Given the current image $I(t^*)$, we first project it into the PCA space ($\mathbf{S}(t^*) = [s_1(t^*); \dots; s_L(t^*)]$). Then, depending on its similarity to the training data and the previous frame, a reference frame is chosen. The logic is as follows:

```

outlierFlag = false
if  $\|\mathbf{S}(t^*) - \mathbf{S}(t_0)\|_2 < \theta_1$  then  $t^{ref} = t_0$ 
else if  $\operatorname{argmin}_{t_x \in [t_0, \dots, T_{10C}]} \|\mathbf{S}(t^*) - \mathbf{S}(t_x)\|_2 < \theta_2$  then  $t^{ref} = t_x$ 
else if  $\|\mathbf{S}(t^*) - \mathbf{S}(t^* - 1)\|_2 < \theta_2$  then  $t^{ref} = t^* - 1$ 
else outlierFlag = true end
if (outlierFlag == false) then do SA-BMA
else do affine registration and update  $\mathbf{S}$  end

```

The threshold θ_1 is the 5th percentile of the Euclidean distance between $\mathbf{S}(t_0)$ and $\mathbf{S}(t_i) \forall t_0 < t_i \leq T_{10C}$. θ_2 is the 95th percentile of the distribution of the minimum Euclidean distances between the $\mathbf{S}(t_i)$ in the training set [4].

3.3 Evaluation

We compared the performance of SA-BMA (Sec. 3.1) and LB-BMA (Sec. 3.2). As baseline BMA, we modified the SA-BMA to have fixed block size $\Delta b_j = 16$. The methods were tested for a total of 25 vessels in 9 sequences, see Fig. 1. We visually inspected the tracking quality for all vessels. We quantitatively evaluated the tracking error for the 15 vessels, which appeared to allow reliable annotations. We randomly selected 10% of the tracking phase images and manually annotated the position (denoted as \bar{P}_j) corresponding to $P_j(t_0)$. For the annotated frame (\hat{t}), we calculated the tracking error $\text{TE}_j(\hat{t}) = \|P_j(\hat{t}) - \bar{P}_j(\hat{t})\|$. We summarize the results by the mean (MTE), standard deviation (SD) and 95th percentile of all $\text{TE}(\hat{t})$, considering all landmarks as a single distribution. We computed the

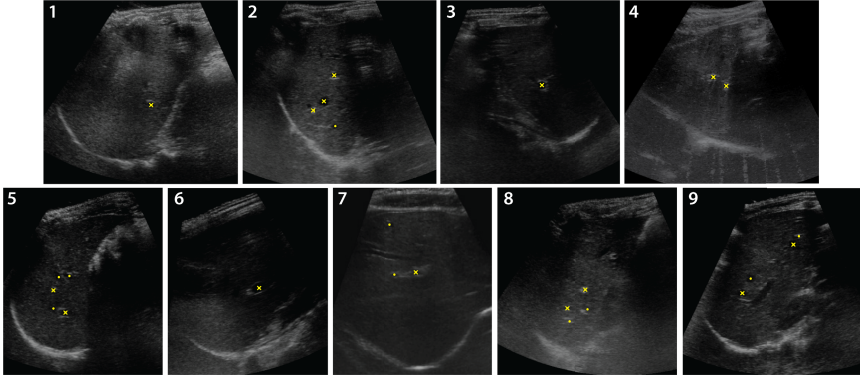


Fig. 1. $I(t_0)$ of the 9 sequences and manual annotation of the tracked vessel centers $P_j(t_0)$, $j \in [1, \dots, 25]$. Quantitative evaluation was based on the 15 P_j marked by 'x'. Visible artifacts include MR-RF interferences (4), and small acoustic windows (2,3,6,8).

MTE for each landmark j (MTE_j) and report the range for the 15 vessels. We included the motion magnitude of the vessels, i.e. $\|P_j(t_0) - \bar{P}(\hat{t})\|$.

We estimated the inter-observer variability of the annotations. Two additional experts annotated 3% of randomly selected images from the tracking phase. We then defined as ground truth the mean position over the 3 annotations and calculated the tracking error as before.

4 Results

We tracked a total of ~ 50000 frames, acquired over a total of ~ 50 min. Δb_j ranges in $[4, 22]$ pixels and the size of the tracked vessels varies from 2 to 9 mm. The PCA space is characterized by L in the range of $[86, 287]$, vs. $D > 3660$.

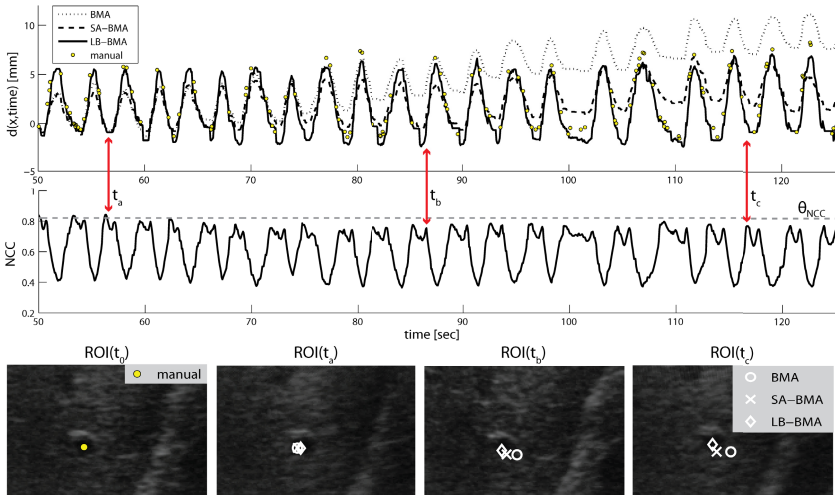
We firstly evaluated the registration error for the training images. The affine registration achieves an accuracy of 0.63 ± 0.36 mm (1.30 mm) on average (MTE \pm SD (95th percentile of TE)), with a MRE $_j$ range of $[0.42, 0.84]$ mm.

Table 1 lists the results for the proposed approaches, SA-BMA and LB-BMA. We compared LB-BMA considering \mathbf{S} (LB-BMA $_{95}$) and \mathbf{s}_B (LB-BMA $_B$). The best performance is achieved by LB-BMA $_{95}$ with a MTE of 0.96 mm. Fig. 2 illustrates the benefit of the proposed methods for the worst case of BMA. For LB-BMA $_{95}$ (LB-BMA $_B$) and all 25 tracked vessels, t^{ref} is picked from the training set for 68.3% (99.0%) of the frames, while 1.4% (1.0%) require affine registration.

The inter-observer MTE varies from 0.30 to 0.34 mm (95th TE from 0.63 to 0.68 mm, MTE $_j$ range $[0.16, 0.70]$). For the inter-observer data set, the median TE $_j$ of BMA and SA-BMA, SA-BMA and LB-BMA $_B$, and SA-BMA and LB-BMA $_{95}$ were statistically significantly different at the 0.001 level (Wilcoxon sign-rank test), while LB-BMA $_B$ and LB-BMA $_{95}$ were not ($p=0.53$). No other statistical tests were performed.

Table 1. Tracking results (in mm) for the different methods w.r.t. manual annotation from one and three observers. Best results are in bold face.

	1 Obs - 10%, ~7500 images		3 Obs - 3%, ~2500 images	
	MTE \pm SD (95 th TE)	rangeMTE _j	MTE \pm SD (95 th TE)	rangeMTE _j
VesselMotion	5.17 \pm 3.21 (10.59)	[2.81, 11.48]	5.22 \pm 3.23 (10.57)	[3.00, 11.30]
BMA	3.22 \pm 2.26 (7.24)	[1.25, 12.35]	3.20 \pm 2.26 (7.17)	[1.26, 12.16]
SA-BMA	2.19 \pm 1.46 (4.90)	[1.20, 5.79]	2.18 \pm 1.45 (4.83)	[1.22, 5.78]
LB-BMA _B	1.24 \pm 1.41 (3.81)	[1.04, 1.49]	1.21 \pm 1.39 (3.67)	[0.98, 1.49]
LB-BMA ₉₅	0.96 \pm 0.64 (2.26)	[0.38, 2.34]	0.97 \pm 0.65 (2.20)	[0.36, 2.24]

**Fig. 2.** Comparison of the tracking performance for a sequence where BMA failed (MTE_j=12.4 mm). (Top) Main motion component of manual annotation and d_j from 3 methods for a temporal subset. (Middle) Corresponding NCC to first image. (Bottom, left to right) First image with annotation ($P(t_0)$), image with tracking results at last realignment (t_a) of (SA-)BMA, at $t_a + 30$ s and $t_a + 60$ s. Drift occurs in a significant (moderate) way for BMA (SA-BMA) for $t > t_a$, while LB-BMA remains robust.

The average time to compute the motion of the tracked vessel per frame was 100 ms (range [30, 350] ms). The PCA projection in LB-BMA required \sim 13 ms per frame. These measures were obtained using non-optimized Matlab software and no GPU parallel computing (single PC with Intel®Core™i7-920 at 2.66 GHz processor and 8 GB RAM), and exclude outliers, i.e. images that required affine registration. The latter was computed in approximately 0.8-2.5 s per image region, using the Insight Segmentation and Registration Toolkit (ITK).

5 Conclusion

We proposed a novel and robust framework for vessel tracking in long US sequences. The method is based on learning the relationship between image appearance and feature displacements to allow frequent reinitialization of a scale-adaptive block-matching algorithm. The method was evaluated on long US sequences of the liver of 9 volunteers under free breathing and achieved a mean accuracy of 0.96 mm for tracking vessels for 5-10 min. To our knowledge, this is the first evaluation for tracking such long US sequences. Our performance also improves the state-of-the-art in 2D US tracking of the human liver (1.6 mm [23]). The proposed method is robust to interference, noise (see Fig. 1), and frame dropouts. Moreover, it is potentially real-time [9,18,4].

Standard BMA might fail in long sequences, due to an inappropriate block size, changes in the image similarity values and error accumulation. The introduction of scale-adaptive blocks and the learning strategy were both significant for the improvement of the results. While adaption to the feature size reduces the error caused by ambiguous matches, the use of NCC for measuring the feasibility of temporal realignment can be misleading. Even with adaptation to the individual US sequence, temporal realignment of the tracking was often too sparse. In contrast, the proposed learning based approach enables more frequent realignments to relevant images by exploiting the repetition in the images and learning the main variation in image appearance. This allows us to detect outliers and then adapt to these previously unseen variations by affine registration, which is slow but able to handle larger displacements.

Reducing computational costs by using only the breathing signal for measuring image similarity increased mean errors slightly (0.96 vs. 1.24 mm). While affine registration performed well on the training set, it was only applied to outliers (1%) during real-time tracking due to its computational complexity [4].

The achieved accuracy and robustness of the proposed tracking method for long and very difficult US sequences makes us confident of its success for real-time US guidance during radiation therapy under free-breathing.

References

1. Boukerroui, D., Noble, J.A., Brady, J.M.: Velocity Estimation in Ultrasound Images: A Block Matching Approach. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 586–598. Springer, Heidelberg (2003)
2. Byram, B., Holley, G., Giannantonio, D., Trahey, G.: 3-D phantom and in vivo cardiac speckle tracking using a matrix array and raw echo data. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 57(4), 839 (2010)
3. Cifor, A., Risser, L., Chung, D., Anderson, E.M., Schnabel, J.A.: Hybrid feature-based Log-Demons registration for tumour tracking in 2-D liver ultrasound images. In: *Proc. IEEE Int. Symp. Biomed. Imaging*, p. 724 (2012)
4. De Luca, V., Tanner, C., Szekely, G.: Speeding-up Image Registration for Repetitive Motion Scenarios. In: *Proc. IEEE Int. Symp. Biomed. Imaging*, p. 1355 (2012)
5. Demi, M., Bianchini, E., Faita, F., Gemignani, V.: Contour tracking on ultrasound sequences of vascular images. *Pattern Recognition and Image Anal.* 18, 606 (2008)

6. Foroughi, P., Abolmaesumi, P., Hashtrudi-Zaad, K.: Intra-subject elastic registration of 3D ultrasound images. *Med. Image Anal.* 10(5), 713 (2006)
7. Harris, E.J., Miller, N.R., Bamber, J.C., Evans, P.M., Symonds-Taylor, J.R.N.: Performance of ultrasound based measurement of 3D displacement using a curvilinear probe for organ motion tracking. *Phys. Med. Biol.* 52(18), 5683 (2007)
8. Harris, E.J., Miller, N.R., Bamber, J.C., Symonds-Taylor, J.R.N., Evans, P.M.: Speckle tracking in a phantom and feature-based tracking in liver in the presence of respiratory motion using 4D ultrasound. *Phys. Med. Biol.* 55(12), 3363 (2010)
9. Hsu, A., Miller, N.R., Evans, P.M., Bamber, J.C., Webb, S.: Feasibility of using ultrasound for real-time tracking during radiotherapy. *Med. Phys.* 32(6), 1500 (2005)
10. Kaluzynski, K., Chen, X., Emelianov, S.Y., Skovoroda, A.R., O'Donnell, M.: Strain rate imaging using two-dimensional speckle tracking. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 48(4), 1111 (2001)
11. 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)
12. King, A.P., Rhode, K.S., Ma, Y., Yao, C., Jansen, C., Razavi, R., Penney, G.P.: Registering preprocedure volumetric images with intraprocedure 3-D ultrasound using an ultrasound imaging model. *IEEE Trans. Med. Imaging* 29(3), 924 (2010)
13. Leung, C., Hashtrudi-Zaad, K., Foroughi, P., Abolmaesumi, P.: A Real-Time Intra-subject Elastic Registration Algorithm for Dynamic 2-D Ultrasound Images. *Ultrasound. Med. Biol.* 35(7), 1159 (2009)
14. Lin, C.H., Lin, M.C.J., Sun, Y.N.: Ultrasound motion estimation using a hierarchical feature weighting algorithm. *Comput. Med. Imaging Graph.* 31(3), 178 (2007)
15. Lindeberg, T.: Feature detection with automatic scale selection. *Int. J. Comput. Vision* 30, 79 (1998)
16. Morsy, A.A., Von Ramm, O.T.: FLASH correlation: a new method for 3-D ultrasound tissue motion tracking and blood velocity estimation. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 46(3), 728 (1999)
17. 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)
18. Pinton, G.F., Dahl, J.J., Trahey, G.E.: Rapid tracking of small displacements with ultrasound. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 53(6), 1103 (2006)
19. Revell, J., Mirmehdi, M., McNally, D.: Computer Vision Elastography: Speckle Adaptive Motion Estimation for Elastography Using Ultrasound Sequences. *IEEE Trans. Med. Imaging* 24(6), 755 (2005)
20. Schneider, R.J., Perrin, D.P., Vasilyev, N.V., Marx, G.R., del Nido, P.J., Howe, R.D.: Real-time image-based rigid registration of three-dimensional ultrasound. *Med. Image Anal.* 16, 402 (2012)
21. 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)
22. Wein, W., Cheng, J.Z., Khamene, A.: Ultrasound based Respiratory Motion Compensation in the Abdomen. In: *MICCAI Workshop on Image Guidance and Computer Assistance for Soft-Tissue Interventions* (2008)
23. Zhang, X., Günther, M., Bongers, A.: Real-Time Organ Tracking in Ultrasound Imaging Using Active Contours and Conditional Density Propagation. In: Liao, H., "Eddie" Edwards, P.J., Pan, X., Fan, Y., Yang, G.-Z. (eds.) *MIAR 2010. LNCS*, vol. 6326, pp. 286–294. Springer, Heidelberg (2010)