4D Reconstruction of Fetal Heart Ultrasound Images in Presence of Fetal Motion

Christine Tanner[1], Barbara Flach[1], Céline Eggenberger[1], Oliver Mattausch[1], Michael Bajka[2], and Orcun Goksel[1]

[1]Computer Vision Laboratory, ETH Zurich, 8092 Zurich, Switzerland
[2]Department of Gynaecology, University Hospital, 8091 Zurich, Switzerland

Abstract. 4D ultrasound imaging of the fetal heart relies on reconstructions from B-mode images. In the presence of fetal or mother's motion, current approaches suffer from artifacts. We propose to use many sweeps and exploit the resulting redundancy to recover from motion by reconstructing a 4D image which is consistent in phase, space and time. We first quantified the performance of 7 formulations on simulated data. Reconstructions of the best and baseline approach were then visually compared for 10 in-vivo sequences. Ratings from 4 observers showed that the proposed consistent reconstruction significantly improved image quality.

1 Introduction

Fast acquisition rates and non-invasiveness of ultrasound (US) imaging makes it an ideal modality for screening the fetal heart to detect congenital heart malformation. Traditionally, the functioning of fetal heart is inspected in real-time during B-mode imaging. Guidelines recommend examination of the four-chamber and outflow tract views [1]. Yet prenatal detection rates vary widely, due to differences in examiner experience, maternal obesity, transducer frequency, gestational age, amniotic fluid volume and fetal position [1]. 4D US imaging simplifies the assessment of the outflow tracts, allows a more detailed examination and contributes to the diagnostic evaluation in case of complex heart defects [2, 1].

4D US of the fetal heart requires special image reconstruction methods, since the speed of 3D US acquisitions using common mechanically steered probes is too slow compared to the fetal heart rate (e.g. 7-10 vs. 2-2.5 Hz). A general approach for such a 4D reconstruction problem is to continuously acquire individual 2D images covering the region of interest [3–5], which then need reordering to extract consistent 3D images. While cardiac 4D MR reconstructions for adults can be supported by ECG and respirator signals [6], these signals cannot reliably extracted for fetus [7]. Hence sorting has relied on extracting the periodic cardiac signal from the images and that no other fetal motion is present [3, 4].

The most common method for fetal 4D US reconstruction is the STIC (Spatio-Temporal Image Correlation) method [4], where autocorrelation is used to detect the systolic peaks, the fetal heart rate (HR) is deduced and the frames are sorted according to their resulting phases. STIC builds on slow, single sweep US acquisitions (e.g. 150 frames/s, 25° in 10 s, 1500 frames) and works well if only

fetal cardiac motion is present. Yet, additional motion from spontaneous fetal activity or mother's breathing creates artifacts [8, 9]. Such artifacts cannot be remedied, as motion affects all consecutive frame positions, which in a single sweep have only been acquired once. Hence mothers are asked to hold their breath and operators may wait for a period of less fetal movement, which prolongs examination time. Volumes acquired by non STIC experts showed more motion artifacts (42%) than those by experts (16%) [8]. Reports on correcting motion for fetal heart 4D US reconstruction have not been encountered.

Image registration has been used to improve reconstructions, but is generally computationally very expensive. For example, correction of fetal 3D MRIs by slice-to-volume rigid registration of local patches required 40 min on multiple GPUs [10]. Correction of adult 3D cardiac MRIs, after gating based on ECG and breathing belt signals, took 3 h on a 16 workstation cluster [6]. For respiratory motion, 4D US reconstruction has been based on extracting a gating signal per slice position by image dimensionality reduction and then matched these signals across slices [5]. This relies on gathering reliable motion statistics per slice, and hence might not be robust to severe, non-periodic motion, e.g. drift.

To avoid time-consuming registrations, we follow the approach of selecting suitable image slices from repeated mechanically-swept US acquisitions. Herein we focus on the consistency of the 4D reconstruction and the detection of outliers due to motion. A large range of selection criteria was first quantitatively evaluated on simulated US sequences. Then, to have statistical power, only the baseline and the best method were applied to in-vivo data, and the visual appearances of the reconstructions were scored by 3 researchers and a gynecologist.

2   Material

Simulated data - To support method development by having a ground truth, B-mode images were simulated from an in-silicon phantom (see Fig. 1a) based on [11]. This method uses GPU ray-tracing to simulate US beam propagation and interactions with given surfaces to accurately simulate typical US attenuation, reflection, refraction, and shadowing effects. Following the mechanical sweep protocol of the real data, 3845 frames at an image frequency of f_i=279 frames/s were simulated. The in-silicon phantom consisted of an ellipsoidal object with semi-axes of $\mathbf{a}$=[9.9 11.5 12.3] mm length. The object changed in size ($\mathbf{a}\pm20\%$) according to a sinusoidal pattern with the simulated HR frequency. Regular HR was set to 143.08 beats/min (i.e. 117 frames/beat). Irregular HR was modelled by linearly changing it by 5% over 1500 frames. Fetal global motion was simulated by applying a [4 8 3] mm translation and a [4 3 8]^o rotation from frame 701 to 1100 and their inverse from frames 1701 to 2200, see Fig. 1c. Simulations included 3 scenarios, namely (Sim1) irregular HR, no global motion, (Sim2) regular HR, with global motion, and (Sim3) irregular HR, with global motion.
In-vivo data - A total of 10 US sequences of 6 fetus at 19-25 weeks of gestation (mean$\pm$SD heart semi-axes of [10.9 9.2 10.6]$\pm$[3.0 2.5 2.3] mm) were acquired. B-mode images were continuously acquired at $f_i\in$[217,372] frames/s (i.e. $\approx$[87,148]

frames/beat) during 56-128 motorized forward-backward sweeps, each covering 25° and consisting of 31 frames (i.e. $\approx[18,46]$ beats/sequence).

3 Method

Reconstruction is based on first estimating the heart rate. Then frames are selected for reconstruction according to phase, spatial and temporal consistency.

3.1 Mean heart rate (HR) estimation

We tested two approaches (A1, A2) for automatically estimating HR f_h (Hz). A1 was based on computing the autocorrelation of the intensity profile over time per pixel, taking their mean and then extracting the power spectrum via Fourier transform. For A2, the image similarity $\mathbf{J}(i,j)$ between frames i and j was measured using various (dis)similarity measures (correlation coefficient (CC), mean square difference (MSD), mutual information (MI) and US specific image similarity measures SK1, SK2, CD1, CD2 [12]). The power spectrum of each row of $\mathbf{J}$ was then extracted via Fourier transform and their mean obtained. For each method the resulting spectrum was bandpass filtered around the expected HR (100-200 beats/min) and f_h was set to the frequency at its global maxima.

3.2 4D Reconstruction

Fig. 2 illustrates the problem of reconstructing P 3D phase images from B B-mode images continuously acquired at K positions in S sweeps. From the estimated HR f_h, the phase value $q_b\in[0.5,P+0.5]$ of the B-mode image $\mathbf{I}_b$ (acquired at time $t=b/f_i$) was estimated from the fractional part of the heart beats (tf_h), i.e. $q_b = (P-1)(tf_h-\lfloor tf_h\rfloor) + 0.5$. The frame from sweep s and position k is denoted as $\mathbf{I}_s^k$ with associated phase q_s^k. For reconstructing P 3D phase images, $P{\times}K$ indices (called $\check{s}_p^k$) need to be determined.

Table 1 provides an overview of the tested reconstruction methods. Method **M0** selects frames whose phase q_s^k is closest to the desired phase p [3, 4].

Greedy methods **M1-M3** first determine for each phase p a reference B-mode image $\mathbf{I}_{\check{s}_p^m}^m$ and then sequentially minimize spatial inconsistency, i.e.

$$\check{s}_p^{k+1} = \arg\min_{s\in S_p^{k+1}} D\left(\mathbf{I}_{s_p^k}^k, \mathbf{I}_s^{k+1}\right) \text{ for } k=\{m, m+1, ..., K\text{-}1, m\text{-}1, m\text{-}2, ..., 1\} \quad (1)$$

where D is an image dissimilarity measure and $\mathcal{S}_p^k=\{s|\ |q_s^k - p| < 0.5\}$ is the set of sweep indices of frames at position k belonging to phase p. For M1, $\mathbf{I}_{\check{s}_p^m}^m$ is the first frame at position $m=1$, which belongs to phase p i.e. $\check{s}_p^1=\min\mathcal{S}_p^1$. M2 is the same as M1 apart from using the midframe ($m=\lceil K/2\rceil$). In M3 the most typical midframe is used as reference, i.e. the midframe which has the highest correlation with all other midframes within the phase range $\mathcal{S}_p^{\lceil K/2\rceil}$:

$$\check{s}_p^{\lceil K/2\rceil} = arg\max_{s\in\mathcal{S}_p^{\lceil K/2\rceil}} \sum_{r\in\mathcal{S}_p^{\lceil K/2\rceil}} CC\left(\mathbf{I}_s^{\lceil K/2\rceil}, \mathbf{I}_r^{\lceil K/2\rceil}\right). \quad (2)$$
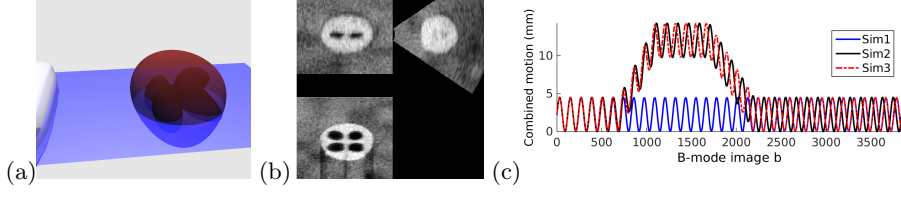
Fig. 1. Illustration of (a) the in-silicon phantom geometry with a transducer plane, (b) a simulated US image and (c) the simulated motion over time.
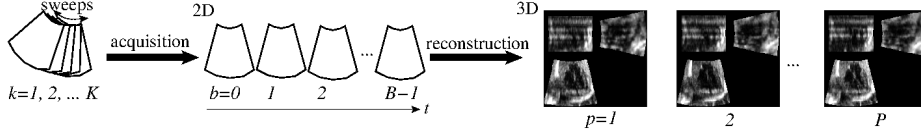
Fig. 2. Problem overview: a sequence of B B-mode images from S sweeps of K positions need to be reconstructed into P 3D images capturing individual heart-beat phases.

Table 1. Overview of methods M0 to M6. Optimization included phase difference PD, spatial inconsistency cost SC and temporal inconsistency cost TC.

Name	f_h	Cost	Optimization type	Reference image $\mathbf{I}_s^m$
M0	fixed	PD	global	n/a
M1	fixed	SC	sequential	$m{=}1, \sqrt{PD} < 0.5, \min(s)$
M2	fixed	SC	sequential	$m{=}\lceil K/2 \rceil, \sqrt{PD} < 0.5, \min(s)$
M3	fixed	SC	sequential	$m{=}\lceil K/2 \rceil, \sqrt{PD} < 0.5, \max(\text{sumCC})$
M4	fixed	PD, SC	global	n/a
M5	opt.	PD, SC	global	n/a
M6	fixed	PD, SC, TC	TC sequential	n/a

In **M4-M6** different cost functions are globally minimized using dynamic programming for determining the best $P \times K$ frame selection indices $\check{s}_p^k$. M4 balances the phase difference PD against the spatial inconsistency cost SC:

$$\check{C}_{f_h} = \min_{s_p^k \in S} \sum_{p=1}^{P} \left(\sum_{k=1}^{K} \underbrace{\left(q_{s_p^k}^k - p \right)^2}_{PD_p^k} + \alpha \sum_{k=1}^{K-1} \underbrace{D\left(\mathbf{I}_{s_p^k}^k, \mathbf{I}_{s_p^{k+1}}^{k+1} \right)}_{SC_p^k} \right) \tag{3}$$

where weight $\alpha{=}\sum_k \alpha_k / K$ was automatically determined by $\alpha_k{=}|\overline{PD}_k / \overline{SC}_k|$ with $\overline{PD}_k$ denoting the mean of PD_p^k for the $R{=}10$ closest observations to p and $\overline{SC}_k$ being the mean of SC_p^k for the R most similar spatial neighbours. M5 is the same as M4 apart from also allowing variations in the estimated HR f_h through an additional grid search over $1/f \in [1/f_h \pm 0.05]$ s to minimize $\check{C}_f = min_{f_h \in f} \check{C}_{f_h}$.

M6 extends Eq. (3) by adding a term for temporal consistency (TC):

$$\check{C}_{f_h} = \min_{s_p^k \in S} \sum_{p=1}^{P} \left(\sum_{k=1}^{K} PD_p^k + \alpha \sum_{k=1}^{K-1} SC_p^k + \beta \sum_{k=1}^{K} TC_p^k \right) \qquad (4)$$

where $TC_p^k = D(\mathbf{I}_{s_p^k}^k, \mathbf{I}_{s_{[(p-1)mod_P]}^k}^k)$, $\beta = \sum_k |\overline{PD}_p^k / (\overline{TC}_p^k K)|$ and $\overline{TC}_p^k$ denotes the mean of TC_p^k for the R most similar temporal neighbours. Eq. (4) was sequentially optimized until convergence after reconstructing a phase via Eq. (3).

Outlier removal (OR) - Having observed on simulated and real data that motion leads to low CC values when comparing images (see Fig. 3), we also tested all method after removing low correlating sweeps. For this we created the CC matrix $\mathbf{J}$ for the midframes, determined the midframe with the lowest mean correlation to all others, and discarded the associated sweep. This was continued until the lowest mean correlation was >0.5 or only 50% of sweeps were left.

4   Experiments and Results

4.1   Mean heart rate (HR) estimation

HR ground truth for in-vivo data was estimated from M-mode traces by counting the number of heart cycles between the first and the last visible extrema.

Table 2 lists the errors in automatic HR estimation for the 3 simulations and 4 in-vivo sequences. Errors were below 0.6% for autocorrelation (A1), and below 3.5% for the image similarity measures (A2) apart from MSD for in-vivo sequence #2 (11.0%). Hence we used A1 for estimating HR for all 4D reconstructions.

4.2   4D Reconstruction of Simulated Data

The performance for the simulations was quantified by combined motion errors. For this, phase errors were converted to motion errors by assigning to each phase value the corresponding mean change in semi axis length (± 2.25 mm).

Table 3 lists the mean absolute error for the most complex simulation (Sim3) when applying methods M0-M6 using one of 3 image dissimilarity measures D, and including outlier removal (OR✓) or not (OR✗). Highest accuracy was achieved by M2 based on CD2 with or without OR (M2-CD2) and by M6-CD2-OR✓. The result overview shows that without motion (Sim1), the errors were low and OR had no impact. For simulations with motion, additional optimization of the heart rate (M5) was counter-productive, while OR generally helped. For the motion scenario with regular heart rate, lowest errors were achieved with M6-CD2 (✗: 0.12, ✓: 0.11 mm) followed by M2-CD2 (0.29 mm). When considering the 3 simulations, M2-CD2 (0.31 mm) and M6-CD2 (✗: 0.26, ✓: 0.23 mm) still provided the lowest errors. The mean runtime of M0, M2, or M6 with OR was 12 s, 191 s, or 285 s, respectively, when reconstructing Sim3 on a single CPU using non-optimized Matlab® code. Prior OR reduced the image data by 31% and the
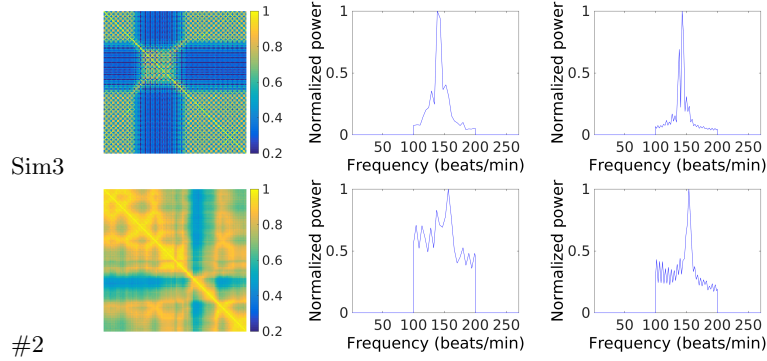
Fig. 3. (left) Example CC matrix **J** of midframes from Sim3 and for in-vivo sequence #2. (middle, right) Power spectra from (middle) **J** and (right) autocorrelation method.

Table 2. Ground truth (GT) heart rate (in beats/min) and difference (GT-estimation) for estimation methods using (A1) autocorrelation or (A2) image similarities.

Method/Sequence	Sim1	Sim2	Sim3	In-vivo1	In-vivo2	In-vivo3	In-vivo4
GT	143.08	143.08	143.08	148.06	154.29	159.34	147.86
A1	0.00	0.00	0.00	-0.75	0.34	0.60	0.46
A2 CC	0.00	0.00	4.62	-0.75	-2.03	0.60	0.46
A2 MSD	0.00	0.00	0.00	-0.75	16.92	0.60	0.46
A2 SK1	0.00	-4.62	-4.62	-0.75	-2.03	0.60	0.46
A2 MI,SK2,CD1,CD2	0.00	0.00	0.00	-0.75	-2.03	0.60	0.46

runtime of M2 (M6) by 58 (59)%. Fig. 4 shows example reconstructions for Sim3. Artifacts can be observed for M0 across the combined frames. Reconstructions by the best OR✓ methods are both very similar to the ground truth. Due to its lower runtime, we selected M2-CD2-OR✓ for the in-vivo evaluation.

4.3   4D Reconstruction of In-Vivo Data

The reconstructed 4D US images were visually inspected using the *vv* image viewer [13]. Four raters were asked to compare the image quality provided by 2 methods for 10 cases and to rate them in a Likert scale as A either (score 1: 'much better', 2: 'better', 3: 'equal', 4: 'worse', or 5: 'much worse') than B. The mean score when comparing M2-CD2-OR✓ against M0 was 2.1 (close to 'better'). The distribution of the 5 categories was 1: 20.0%, 2: 57.5%, 3: 15.0%, 4: 7.5% and 5: 0%. The mean score of the 4 raters ranged from 1.8 to 2.4, with the clinician's result being closest to the mean (2.2). The median score (2: 'better') was statistically significantly different from score 3 ('equal') at the <0.0001 level using the Wilcoxon signed rank test. Raters observed reduced artifacts across frames and fewer motion artifacts. Fig. 5 shows sample reconstructions. Misalignment artifacts are clearly reduced by M2-CD2-OR✓.

D	OR	M0	M1	M2	M3	M4	M5	M6
CC	×	2.59	1.68	0.50	0.58	0.88	6.04	0.93
CD2	×	2.59	0.92	**0.36**	0.47	2.14	4.49	0.39
MI	×	2.59	1.81	0.64	0.77	3.95	5.59	1.57
CC	✓	0.71	0.57	0.50	0.59	0.54	1.25	0.55
CD2	✓	0.71	0.62	**0.36**	0.47	0.68	1.51	**0.36**
MI	✓	0.71	0.88	0.64	0.77	0.61	1.59	1.03
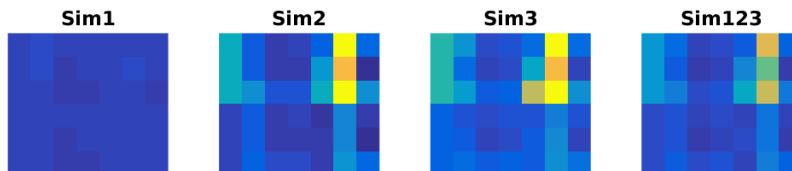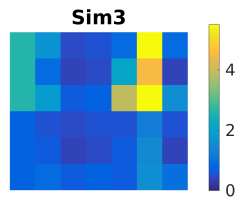
Table 3. (top-left) Table with mean absolute errors (in mm) for simulation Sim3. The lowest errors are marked in bold. (top-right) Visualization of table results. (bottom) Visualization of results for all simulations and their mean (Sim123).
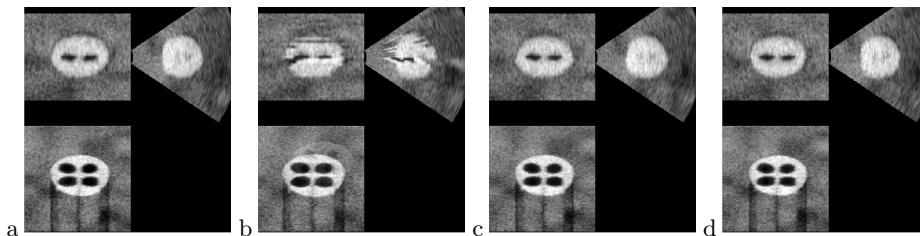

Fig. 4. Orthogonal example slices from reconstruction of simulation Sim3 for (a) ground truth and methods (b) M0, (c) M2-CD2-OR✓ and (d) M6-CD2-OR✓.

5 Discussion and Conclusion

We developed a fast reconstruction method, which improved reconstructions of 4D fetal heart US images noticeable in comparison to neglecting the presence of fetal motion. Based on evaluations on simulated data, the proposed outlier removal was found beneficial. The most successful methods were M6 by optimizing phase, spatial and temporal consistency, and M2 by using the first midframe within a phase and iteratively selecting the most similar neighbouring slice.

The developed framework is suitable for continuous, long acquisitions. Dissimilarity calculation of neighbouring slices (97% of M3 runtime) is easily parallelizable. A real-time implementation can also use the outlier removal criterion to process midframes immediately for providing real-time feedback on acquisition quality. The out-of-plane image resolution can be improved by denser sampling (slower speed) of the sweep. Given the relatively low number of rejected outliers in this study, reconstruction of more phases should also be possible, if needed.
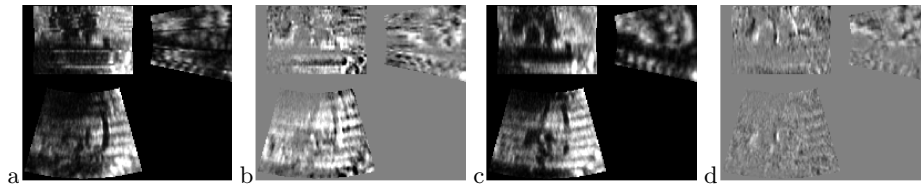
Fig. 5. Example of a representative in-vivo reconstruction (mean score 1.75) for (a,b) M0 and (c,d) M2-CD2-OR✓ for (a,c) phase 2 and (b,d) difference phase 3 - phase 2.

References

1. JS Carvalho, LD Allan, R Chaoui, JA Copel, GR DeVore, K Hecher, et al., "ISUOG Practice Guidelines (updated): sonographic screening examination of the fetal heart," *Ultrasound Obstet Gynecol*, vol. 41(3), pp. 348, 2013.
2. GR DeVore, P Falkensammer, MS Sklansky, and LD Platt, "Spatio-temporal image correlation (STIC): new technology for evaluation of the fetal heart," *Ultrasound Obstet Gynecol*, vol. 22(4), pp. 380, 2003.
3. TR Nelson, DH Pretorius, M Sklansky, and S Hagen-Ansert, "Three-dimensional echocardiographic evaluation of fetal heart anatomy and function: acquisition, analysis, and display.," *J Ultras Med*, vol. 15(1), pp. 1, 1996.
4. A Schoisswohl and P Falkensammer, "Method and apparatus for obtaining a volumetric scan of a periodically moving object," Nov. 22 2005, US Patent 6,966,878.
5. C Wachinger, M Yigitsoy, E-J Rijkhorst, and N Navab, "Manifold learning for image-based breathing gating in ultrasound and MRI," *Med Image Anal*, vol. 16(4), pp. 806, 2012.
6. F Odille, A Bustin, B Chen, P Vuissoz, and J Felblinger, "Motion-corrected, super-resolution reconstruction for high-resolution 3D cardiac cine MRI," in *Medical Image Computing and Computer-Assisted Intervention*, p. 435. Springer, 2015.
7. I Peterfi, L Kellenyi, and A Szilagyi, "Noninvasive recording of true-to-form fetal ECG during the third trimester of pregnancy," *Obstet Gynecol Int*, vol. 2014, 2014.
8. LB Uittenbogaard, MC Haak, MD Spreeuwenberg, and JMG Van Vugt, "A systematic analysis of the feasibility of four-dimensional ultrasound imaging using spatiotemporal image correlation in routine fetal echocardiography," *Ultrasound Obstet Gynecol*, vol. 31(6), pp. 625, 2008.
9. S Yagel, A Benachi, D Bonnet, Y Dumez, D Hochner-Celnikier, SM Cohen, et al., "Rendering in fetal cardiac scanning: the intracardiac septa and the coronal atrioventricular valve planes," *Ultrasound Obstet Gynecol*, vol. 28(3), pp. 266, 2006.
10. B Kainz, A Alansary, C Malamateniou, K Keraudren, M Rutherford, JV Hajnal, and D Rueckert, "Flexible reconstruction and correction of unpredictable motion from stacks of 2D images," in *Medical Image Computing and Computer-Assisted Intervention*, p. 555. Springer, 2015.
11. B Bürger, S Bettinghausen, M Radle, and J Hesser, "Real-time GPU-based ultrasound simulation using deformable mesh models," *IEEE T Med Imaging*, vol. 32(3), pp. 609, 2013.
12. B Cohen and I Dinstein, "New maximum likelihood motion estimation schemes for noisy ultrasound images," *Pattern Recogn*, vol. 35(2), pp. 455, 2002.
13. P Seroul and D Sarrut, "VV: a viewer for the evaluation of 4D image registration," in *MIDAS Journal*, 2008, p. 1.