CrossMark

ORIGINAL ARTICLE

Consistent reconstruction of 4D fetal heart ultrasound images to cope with fetal motion

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

Abstract

Purpose 4D ultrasound imaging of the fetal heart relies on reconstructions from B-mode images. In the presence of fetal motion, current approaches suffer from artifacts, which are unrecoverable for single sweeps.

Methods We propose to use many sweeps and exploit the resulting redundancy to automatically recover from motion by reconstructing a 4D image which is consistent in phase, space, and time. An interactive visualization framework to view animated ultrasound slices from 4D reconstructions on arbitrary planes was developed using a magnetically tracked mock probe.

Results We first quantified the performance of 10 4D reconstruction formulations on simulated data. Reconstructions of 14 in vivo sequences by a baseline, the current state-of-the-art, and the proposed approach were then visually ranked with respect to temporal quality on orthogonal views. Rankings from 5 observers showed that the proposed 4D reconstruction approach significantly improves temporal image quality in comparison with the baseline. The 4D reconstructions of the baseline and the proposed methods were then inspected interactively for accessibility to clinically important views and rated for their clinical usefulness by an ultrasound spe-

cialist in obstetrics and gynecology. The reconstructions by the proposed method were rated as 'very useful' in 71% and were statistically significantly more useful than the baseline reconstructions.

Conclusions Multi-sweep fetal heart ultrasound acquisitions in combination with consistent 4D image reconstruction improves quality as well as clinical usefulness of the resulting 4D images in the presence of fetal motion.

Keywords Ultrasound · Fetal heart · Reconstruction

✉ Christine Tanner
  tanner@vision.ee.ethz.ch

  Orcun Goksel
  ogoksel@vision.ee.ethz.ch

[1] Computer-Assisted Applications in Medicine, Computer Vision Lab, ETH Zurich, Zurich, Switzerland

[2] Departments of Obstetrics and Gynaecology, Zurich University Hospital, Zurich, Switzerland

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, mainly 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 outflow tract, allows for a more detailed examination, and contributes to the diagnostic evaluation in case of complex heart defects [1,4].

Spatio-temporal image correlation (STIC) [13] is a well-known 4D US reconstruction approach for fetal heart. Similarly to earlier works [10], STIC builds on very slow, single sweep US acquisitions; e.g., 1500 frames of roughly 25° elevational field of view in 10 s. Then, autocorrelation is used to estimate the fetal heart rate (HR) and the frames are sorted based on their resulting phases. With this, all heart phases (i.e., within ≈0.5 s) exist within a probe sweep of

Springer

merely $\approx 1°$, and interpolation on a fixed grid after sorting can yield successful reconstructions—but only in the absence of any external motion. Fetal organ screening has to take place between 18 and 22 weeks of gestation, a time when movements are already an important sign of fetal well-being. These movements and the different and changing position of the fetus' body and extremities may turn fetal heart examination into a difficult task. This exacerbated by patient breathing creates significant artifacts [16,18] with no straightforward way of compensating motion, since each sweep angle is acquired only once. To the best of our knowledge, there has been no reports on correcting fetal motion for STIC fetal heart reconstructions. Accordingly, mothers are asked to hold their breath and operators wait for a period of calmer fetal activity, which often requires several trials, and potentially yielding no successful 4D reconstructions. It is also quite operator dependent; for instance, acquisitions by non-STIC experts show more motion artifacts (42%) than those by experts (16%) [16].

With the advance of 2D-matrix arrays and ultrafast imaging [2,15], it may be possible to collect volumes at sufficiently high frame rates to reconstruct the fetal heart, e.g., within one beat. However, the image quality of individual ultrafast frames are often low, and such technology still has a long way to come to obstetrics applications in particular regarding fetal safety concerns.

We propose a method for spatio-temporal fetal heart reconstruction using image sequences from rapid sweeps of common mechanically swept probes. These yield several volumes where fetal motion can potentially be resolved. Nonetheless, sophisticated reconstruction techniques are required, since the swept probes are slow compared to the fetal heart rate; i.e., the entire heart at a phase cannot be captured in a single sweep (e.g., 5–12 sweeps/s, 2.5 beats/s results in only 2–4.8 sweeps per heartbeat). With other imaging modalities, a general approach to such a 4D reconstruction problem from continuously acquired individual 2D images is to *reorder* the slices based on their *consistency* within a reconstruction [10,13,17]. External gating is used to avoid motion and as trigger signal to extract the exact phase. For instance, adult cardiac 4D MR reconstruction is supported by ECG and respiratory signals [11]. However, these signals cannot be reliably extracted for fetus [12] and HR estimation directly from the US images avoids changing clinical practice. For fetal cardiac MRI, such self-gating has been based on optimizing the time-entropy image metric and assumes a piecewise-constant heart rate [5]. Yet this approach cannot compensate for any non-cardiac motion.

For respiratory motion, 4D US reconstruction has been studied based on extracting a gating signal per slice position by dimensionality reduction and then matching these signals across slices [17]. This relies on gathering motion statistics per slice and hence might not be robust to non-periodic

motion, e.g., drift. In order to improve reconstructions, image registration has also been used, although this is often computationally very expensive. For example, correction of fetal 3D MRIs using slice-to-volume rigid registration of local patches required 40 min on multiple GPUs in [6]. Correction of adult 3D cardiac MRIs, after gating based on ECG and breathing belt signals, took 3 h on a 16 workstation cluster in [11].

We performed a preliminary test to compensate for fetal motion by rigidly registering the frames based on the regions away from the heart (to minimize distortions from heartbeats) using normalized cross-correlation. This, however, did not yield satisfactory motion compensation. Therefore, we herein resort to an approach of selecting suitable image slices from repeated acquisitions. We focus on the consistency of a 4D reconstruction and the detection of outliers due to motion. A large range of selection criteria was first quantitatively evaluated on *simulated* US sequences including motion. For the in vivo data, in order to boost the statistical power, 3 of these methods were identified and applied: a baseline, the state-of-the-art, and our proposed method. Temporal visual quality of the reconstructions was ranked by 4 technical US experts in addition to an US specialist in obstetrics and gynecology. In contrast to our earlier study in [14], herein we additionally (i) investigate the effects of US-specific filtering on reconstructions and of a L1-norm phase constraint, which is seen to yield better results; (ii) have increased our in vivo fetal heart dataset by 40%; (iii) developed an interactive interface to view animated planes from 4D reconstructions; and (iv) have included additional user studies and evaluations on temporal consistency and clinical usefulness.

Material

Simulated data

To support method development based on some ground-truth data, B-mode images were simulated from a numerical phantom (see Fig. 1a) based on [9]. This method uses GPU ray tracing to simulate US beam propagation and interactions with given anatomical surface representations to accurately simulate typical US attenuation, reflection, refraction, and shadowing effects present in US images. Simulating the probe positions based on a 3D probe geometry and the mechanical sweeping action, 3658 frames at an image frequency of $f_i = 279$ frames/s (fps) were generated. The numerical phantom consisted of an ellipsoidal object representing a fetal heart with semi-axes of $\mathbf{a} = [9.9\ 11.5\ 12.3]$ mm. The size of this ellipsoid was changed sinusoidally by $\mathbf{a} \pm 20\%$ to simulate heartbeat. Regular HR was set to 143.08 beats/min (bpm), leading to 117 frames/beat. Irregular HR was modeled by increasing then decreasing the HR by 5% over 1500 frames (5.4 s) between
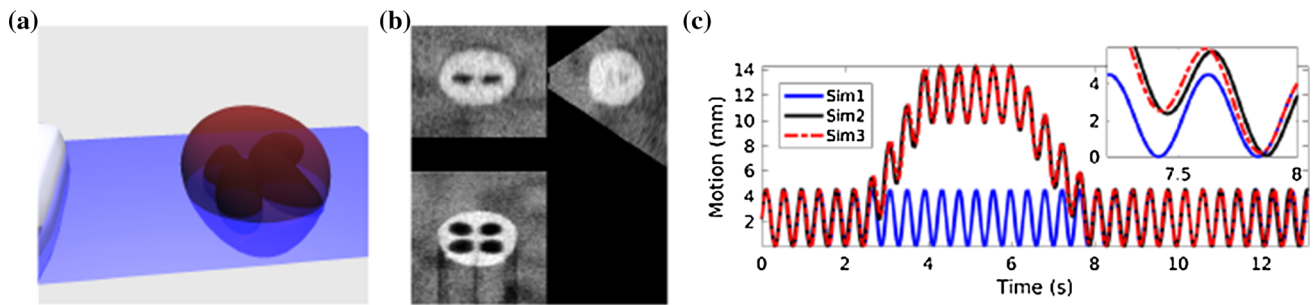
Fig. 1 Illustration of **a** the in silico phantom geometry with a transducer plane, **b** a simulated US image and **c** the simulated combined motion over time

Table 1 Acquisition details of in vivo data listing gestation age (GA) in weeks, acquisition frequency (acqF) in sweeps/s, sweep angle (swA), number of frames per sweep (K), total number of sweeps (S), total number of frames ($B = KS$), and total acquisition time (acqT). Extracted heart rate f_h using autocorrelation ('Methods' section) and deduced beats per sequence (b/sq) and sweeps per beat (sw/b). Percentage of inliers during outlier removal ('Methods' section)

No	GA	acqF	swA	K	S	B	acqT	f_h			Inlier
	w	sw/s	^{o}				s	bpm	b/sq	sw/b	%
#1	25	9	25	31	128	3968	14.2	148.8	35	3.6	100
#2	25	9	25	31	115	3565	12.8	153.9	33	3.5	100
#3	20	7	45	55	128	7040	18.3	153.8	47	2.7	78
#4	25	7	25	26	56	1456	8.0	145.1	19	2.9	93
#5	25	9	25	26	107	2782	11.9	147.7	29	3.7	95
#6	20	12	25	31	128	3968	10.7	158.7	28	4.5	100
#7	20	12	25	31	128	3968	10.7	147.4	26	4.9	100
#8	20	6	45	55	98	5390	16.3	167.0	45	2.2	100
#9	20	5	65	79	77	6083	15.4	122.4	31	2.5	99
#10	20	5	65	79	57	4503	11.4	128.6	24	2.3	100
#11	20	5	65	79	128	10,112	25.6	127.6	54	2.4	65
#12	25	8	60	43	58	2494	7.3	155.8	19	3.1	100
#13	25	8	60	43	68	2924	8.5	157.6	22	3.0	100
#14	20	8	60	43	128	5504	16.0	143.6	38	3.3	100
Min	20	5	25	26	56	1456	7.3	122.4	19	2.2	65
Mean	21	8	44	47	100	4554	13.4	147.0	32	3.2	95
Max	25	12	65	79	128	10,112	25.6	167.0	54	4.9	100

139.5 and 146.5 bpm. Fetal motion was simulated by applying a [4 8 3] mm translation and a [4 3 8]° rotation linearly during frames [701, 1100] and reverting these during frames [1701, 2200], as shown in Fig. 1c. Simulations included 3 scenarios: (**Sim1**) irregular HR, no global motion; (**Sim2**) regular HR, with global motion; and (**Sim3**) irregular HR, with global motion.

In vivo data

Fourteen US sequences from 8 fetus at 20–25 weeks of gestation with mean $\pm$ SD heart semi-axes of [13.4 9.8 11.5] $\pm$ [3.2 1.8 2.3] mm were acquired. B-mode images were continuously acquired at $f_i \in$ [182, 395] fps (i.e., 75–194 frames/beat) during 56–128 motorized forward–backward sweeps, each covering 25°–44° and consisting of 26–44 frames (i.e., 19–54 beats/sequence), see Table 1.

Methods

Figure 2 illustrates the problem of reconstructing P 3D images of heartbeat phases from a sequence of B B-mode images (also called frames) continuously acquired at K discrete angles in S sweeps. The frame from sweep s and angle k is denoted as $\mathbf{I}_s^k$. Our reconstruction is based on first estimating the dominant HR from the sequence of midframes of the sweeps $\mathbf{I}_s^{\lceil K/2 \rceil}$, and then selecting frames for 4D reconstruction according to phase, spatial, and temporal consistency criteria. In contrast to the baseline method [13], the devised
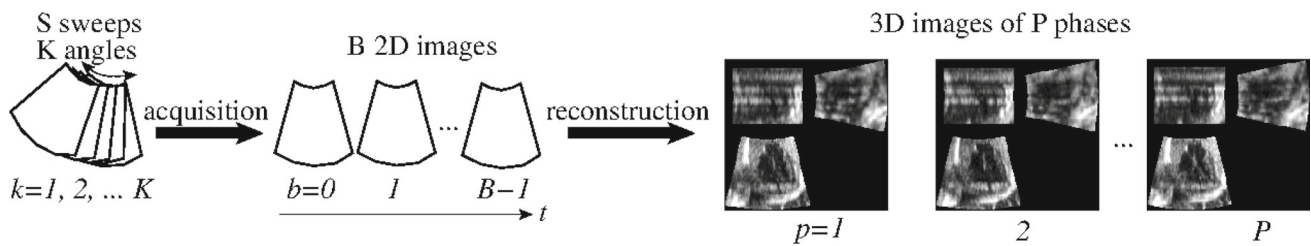
Fig. 2 Problem overview: Reconstruct P 3D volumes of different heartbeat phases from a sequence of B images from S sweeps at K discrete angles

reconstruction methods allow selected frames to deviate from the estimated dominant HR if this improves spatial (or temporal) consistency.

Mean heart rate (HR) estimation

We tested two approaches (A1, A2) for automatically estimating HR f_h (Hz). Approach A1 is based on the autocorrelation of the intensity profile of a pixel **x** over time ($\mathbf{I}_s^{\lceil K/2 \rceil}(\mathbf{x})$). From the mean autocorrelation of all pixels, the power spectrum is then extracted via Fourier transform, where the peak estimates the dominant HR. For approach A2, the image similarity $\mathbf{J}(i, j)$ between every midframe $\mathbf{I}_i^{\lceil K/2 \rceil}$ and $\mathbf{I}_j^{\lceil K/2 \rceil}$ is computed using various image similarity metrics (herein, the correlation coefficient (CC), negative mean square difference (MSD), mutual information (MI), and US-specific measures SK1, SK2, CD1, CD2 from [3]). The power spectra of each row of matrix **J**, computed via Fourier transform, are then averaged to incorporate the information from the comparisons of all frames, to increase signal-to-noise ratio, and to provide the dominant heart rate even with motion. After bandpass filtering the resulting mean spectra between an expected fetal HR of [100, 200] bpm, the maximum yields the dominant HR f_h.

4D reconstruction

Based on the estimated HR f_h, we estimate the phase value $q_b \in [0.5, P+0.5]$ associated with frame $\mathbf{I}_b$ (acquired at time $t = b/f_i$) from the fractional part of the heartbeats (tf_h), i.e., $q_b = (P-1)(tf_h - \lfloor tf_h \rfloor) + 0.5$. The frame from sweep s and angle k is denoted as $\mathbf{I}_s^k$ with associated estimated phase q_s^k. For reconstructing P 3D phase images, $P \times K$ sweep indices (called $\check{s}_{p,k}$) need to be determined.

Next we describe the baseline (M0) and the devised reconstruction methods (M1–M6), which employ increasing levels of sophistication. Baseline method **M0** selects frames whose estimated phases q_s^k are closest to the desired phases p [10,13]. Greedy methods **M1-M3** first determine for each desired phase p a reference B-mode image $\mathbf{I}_{\check{s}_{p,m}}^m$ and then sequentially minimize the inconsistency to spatially neighboring frames, i.e.,

$$\check{s}_{p,k+1} = \arg\min_{s \in \mathcal{S}_{p,k+1}} d\left(\mathbf{I}_{\check{s}_{p,k}}^k, \mathbf{I}_s^{k+1}\right)$$
$$\text{for } k = \{m, m+1, ..., K-1, m-1, m-2, ..., 1\} \quad (1)$$

where d is an image dissimilarity measure ($d_{-\text{CC}}$, d_{MSD}, $d_{-\text{MI}}$, $d_{-\text{SK1}}$, $d_{-\text{SK2}}$, $d_{-\text{CD1}}$, $d_{-\text{CD2}}$) and $\mathcal{S}_{p,k} = \{s \in \mathcal{S} : |q_s^k - p| < 0.5\}$ is the set of sweep indices of frames at angle k belonging to phase p. In **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 similar to M1, apart from using the midframe as reference ($m = \lceil K/2 \rceil$). In **M3**, the most typical midframe is used as the reference, i.e., the midframe which has the highest correlation with all other midframes within the phase range $\mathcal{S}_{p,\lceil K/2 \rceil}$:

$$\check{s}_{p,k} = \arg\min_{s \in \mathcal{S}_{p,k}} \sum_{r \in \mathcal{S}_{p,k}} d_{-\text{CC}}\left(\mathbf{I}_s^k, \mathbf{I}_r^k\right) \text{ for } k=\lceil K/2 \rceil. \quad (2)$$

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 spatial inconsistency cost $c_k^S(s, r) = d(\mathbf{I}_s^k, \mathbf{I}_r^{k+1})$ with the absolute or squared phase difference cost $[c_{p,k}^P(s)]^n = |q_s^k - p|^n$, for $n \in \{1, 2\}$:

$$\check{c}_{f_h} = \min_{s,r \in \mathcal{S}} \sum_{p=1}^{P} \left(\sum_{k=1}^{K} \left[c_{p,k}^P(s) \right]^n + \alpha \sum_{k=1}^{K-1} c_k^S(s, r) \right) \quad (3)$$

where desired phase p depends on the estimated HR f_h and weight α is automatically determined from the relationship between the typical phase difference values and spatial inconsistency costs. In detail, $\alpha = \sum_k |\bar{c}_k^P / \bar{c}_k^S| / K$ with $\bar{c}_k^P$ denoting the mean of $c_{p,k}^P$ for the $R = 10$ closest observations to the desired phase p and $\bar{c}_k^S$ being the mean of c_k^S for the R most similar spatial neighbors. **M5** is similar to M4, while 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 the combined cost $\check{c}_{f_h}$. **M6** extends Eq. (3) with an additional temporal consistency term $c_{p,k}^T(t, s) = d(\mathbf{I}_t^k, \mathbf{I}_s^k)$ where $\mathbf{I}_t^k$ and $\mathbf{I}_s^k$ are temporal neighbors in the sense that they

Table 2 Overview of methods M0 to M6

Name	f_h	Cost	Optimization type	Reference image $\mathbf{I}_s^m$
M0	Fixed	c^P	Global	n/a
M1	Fixed	c^S	Sequential	$m = 1, c^P < 0.5, \ \min(s)$
M2	Fixed	c^S	Sequential	$m = \lceil K/2 \rceil, c^P < 0.5, \ \min(s)$
M3	Fixed	c^S	Sequential	$m = \lceil K/2 \rceil, c^P < 0.5, \ \min(\sum d_{-CC})$
M4	Fixed	c^P, c^S	Global	n/a
M5	Opt.	c^P, c^S	Global	n/a
M6	Fixed	c^P, c^S, c^T	c^T sequential	n/a

Optimization included phase difference c^P, spatial inconsistency cost c^S, and temporal inconsistency cost c^T

will belong to neigboring phases in the reconstruction, i.e., $t \in \check{S}_{(p-1) \bmod_P, k}$ and $s \in \check{S}_{p,k}$:

$$\check{c}_{f_h} = \min_{s,r \in \mathcal{S}} \sum_{p=1}^{P} \left(\sum_{k=1}^{K} \left[c_{p,k}^P(s) \right]^n + \alpha \sum_{k=1}^{K-1} c_k^S(s,r) \right.$$
$$\left. + \beta \sum_{k=1}^{K} c_{p,k}^T(t,s) \right) \tag{4}$$

where weight β is also automatically determined by using $\beta = \sum_k |\bar{c}_k^P / \bar{c}_k^T| / K$ where $\bar{c}_k^T$ denotes the mean of $c_{p,k}^T$ for the R most similar temporal neighbors. Equation (4) is optimized iteratively, after initializing it by a phase reconstructed via Eq. (3). An overview of methods M0 to M6 is provided in Table 2.

Outlier removal (OR)

Having observed that motion leads to low CC values when comparing images (see Fig. 4), we also tested all methods after removing low correlating sweeps—indicating those acquired while the fetus was at a different location. We use the CC matrix $\mathbf{J}$ of the midframes, pick the midframe with the lowest mean correlation to all others, and discard the associated sweep. This is repeated until the lowest mean correlation is >0.5 or only 50% of sweeps are left. These thresholds were set empirically based on the observed pattern of overall mean correlation values.

Image filtering (IF)

We also tested an US-specific filtering method to reduce the impact of US speckles before the calculation of image similarity methods. Assuming speckle as a multiplicative noise, different filtering algorithms were compared in [7], where a moving window using local statistics was reported to work well regarding several metrics for vessel imaging. We use this filter [8] with an empirically set filter size of 3.

Visualizing 4D reconstructions

Clinical examinations are performed on standardized views and planes, which are not always easy to image during acquisitions. These also proved difficult to find in 4D reconstructions using standard graphical interfaces for image viewing and rotation. Therefore, we developed a visualization interface in which 4D reconstructions are loaded and animated views from these are shown interactively on a plane controlled by a magnetically tracked mock transducer. This allows the physician to easily and intuitively manipulate the viewing plane to find clinically relevant orientations.

Experiments and results

Estimating the heart rate

Gold-standard dominant HR for the in vivo data was estimated by counting the number of heartbeats observed from the heart wall between the first and the last visible beat on M-mode images from the midframes, see Fig. 3b. 10–27 heartbeats, covering 30–87% of the sequence, could be identified for 4 in vivo cases. Hence, quantification differences are likely to introduce small errors when compared to the whole sequence.

Figure 4 illustrates the stages of our HR estimation process. The correlation matrices of midframes are seen in Fig. 4a, where variations from heartbeat and other motion can observed as colored bands. The spectra from the autocorrelation method A1 (Fig. 4c) provided better defined peaks compared to deriving those with A2 from the CC matrices $\mathbf{J}$ (Fig. 4b).

Table 3 lists the errors in automatic HR estimation for the 3 simulations and 4 in vivo sequences. Errors were below 0.8% for autocorrelation (A1), and below 4.7% for the image similarity metrics (A2) except MSD for in vivo sequence #2 (16.9%). Among similarity metrics for A2, CC performed consistently well. Hence, we used A1 for estimating HR for all 4D reconstructions.
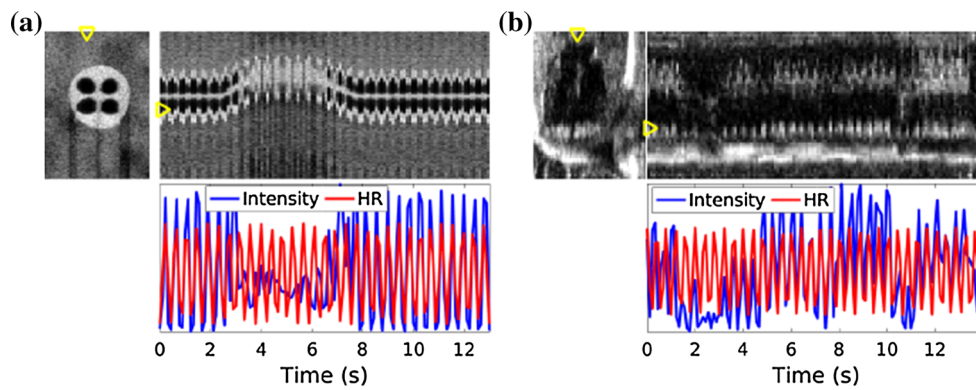
Fig. 3 (*Top*) First midframe and M-mode image of midframes from column marked by *yellow* ▽ and (*bottom*) Intensity: intensity values at pixel location marked by *yellow* ▷, and HR: sinusoidal illustration of estimated dominant HR for **a** Sim3 and **b** #1
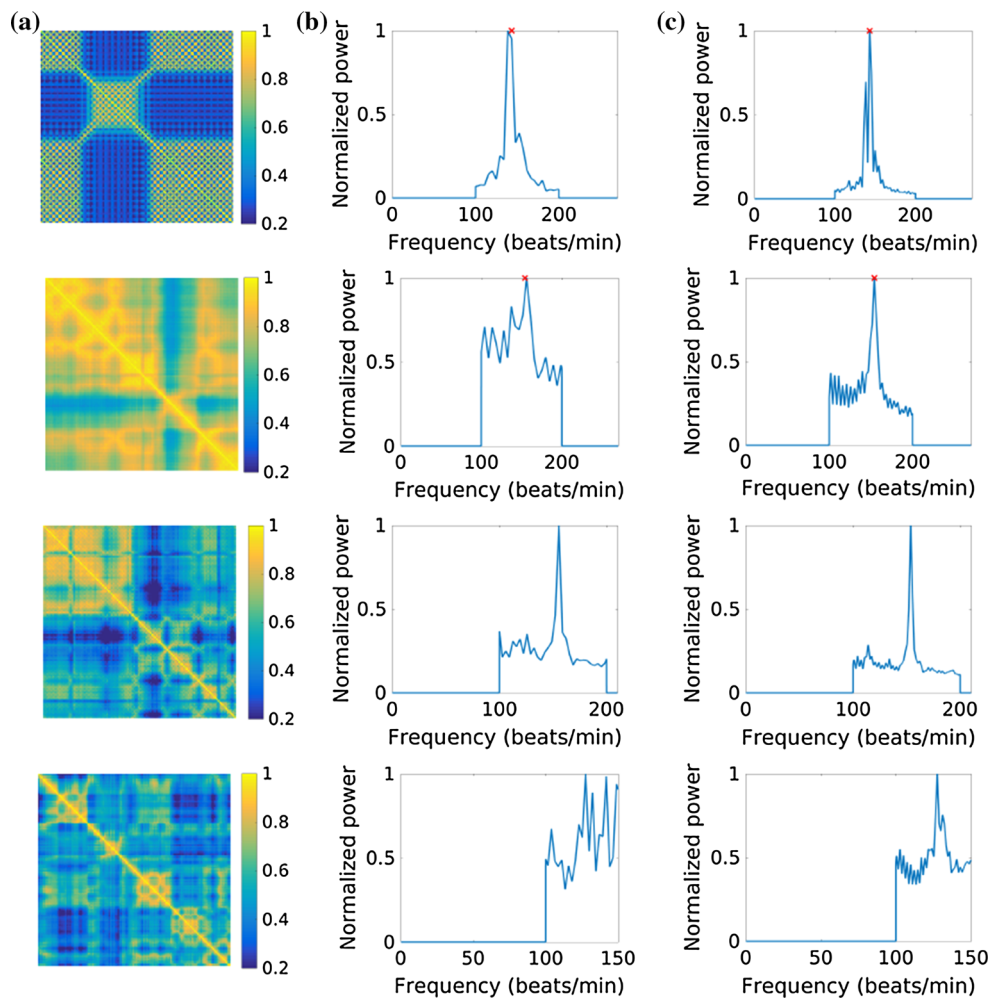
Fig. 4 Illustration of heart rate (HR) estimation for (*top to bottom*) Sim3 and in vivo #2, #3, #11. **a** Correlation coefficient matrix **J** between midframes. Heartbeats introduce repetitive patterns with relatively high correlation, while large motion causes decorrelation. **b, c** Power spectra from **b J** and **c** autocorrelation method, with ground truth marked by *red* × for Sim3 and #2

4D reconstruction of simulated data

We reconstructed $P = 8$ phases. The performance for the simulations was quantified by combined motion errors. For this, phase errors were converted to motion errors by assigning each unit of phase difference to a position error equivalent to mean motion of heart between two consecutive phases (4.5 mm/P). To find a method which can cope with all 3

Table 3 Gold-standard (GS) heart rate (in bpm) and difference (GS-estimation) for estimation methods using (A1) autocorrelation or (A2) image similarities

Method	Sequence						
	*Sim*1	*Sim*2	*Sim*3	#1	#2	#6	#7
GS	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

Table 4 (Top) Table with mean absolute errors (in mm) for all 3 simulations (Sim123)

| IF	OR	-d	M0	M1	M2	M3	Sq. L2-norm M4	M5	M6	L1-norm M4	M5	M6
×	×	CC		0.95	0.44	0.53	0.52	3.11	0.35	0.62	3.19	0.45
×	×	CD2	1.74	0.69	0.38	0.44	0.96	3.25	0.81	1.37	3.46	**0.24**
×	×	MI		1.19	0.83	1.09	3.33	4.47	2.71	2.13	4.26	0.48
×	✓	CC		0.43	0.42	0.54	0.29	1.33	0.26	0.30	1.34	0.37
×	✓	CD2	0.32	0.41	0.37	0.45	0.31	1.23	0.29	0.32	1.25	**0.25**
×	✓	MI		0.38	0.48	0.56	0.34	1.42	0.35	0.32	1.42	0.54
✓	×	CC		0.96	0.62	0.65	0.43	1.95	0.33	0.61	2.74	0.46
✓	×	CD2	1.74	0.68	0.47	0.52	0.98	3.27	0.79	1.37	3.46	**0.23**
✓	×	MI		0.69	0.56	0.62	3.29	4.59	1.32	1.46	4.13	0.46
✓	✓	CC		0.45	0.58	0.61	0.29	1.34	**0.23**	0.30	1.24	0.32
✓	✓	CD2	0.33	0.42	0.46	0.51	0.32	1.24	0.29	0.33	1.26	**0.23**
✓	✓	MI		0.41	0.53	0.59	0.30	1.43	0.32	0.31	1.37	0.47

Errors within 10% of the lowest error (0.23 mm) are marked in bold. The accuracy of the baseline, state-of-the-art, and proposed method is 1.74, 0.37, and 0.23 mm, respectively (highlighted by boxes). (bottom) Visualization of results for all simulations and their mean

simulation scenarios, methods were compared on the basis of the mean error over all 3 simulations.

Table 4 lists the mean absolute error for all simulation (Sim123) when applying methods M0–M6 using one of 3 image dissimilarity measures d on filtered (IF✓) or not filtered (IF×) images, including outlier removal (OR✓) or not (OR×), and measuring phase differences via the squared L2 or L1 norm ($[c^P]^n$) in methods M4-M6. The highest accuracy of 0.23 was achieved by three methods, namely M6-L1 based on CD2-IF✓ with or without OR, and by M6-L2 based on CC-IF✓ and OR✓. Any M6-L1-CD2 method achieved results within 10% of the minimum. The results with and without filtering (IF) were highly correlated with $r \in [0.92, 0.99]$. Without motion (Sim1), the errors were low

and OR had no impact as no outliers were detected. For simulations with motion (Sim2, Sim3), additional optimization of the heart rate (M5) was counter-productive, while OR generally helped. Image similarity MI was the worst at detecting inconsistent frames due to motion.

The mean runtime of M0, M2, or M6 with OR was 12, 191, 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 runtime of M2 (M6) by 58 (59)%. Image filtering IF increased the runtime by 28 s. Figure 5 illustrates the frame selection. Without OR (left plots), M6 avoids by itself the frames with additional motion, while M0 (M2) includes many (a few) of these. The lines connecting the selected frames per phase are more straight
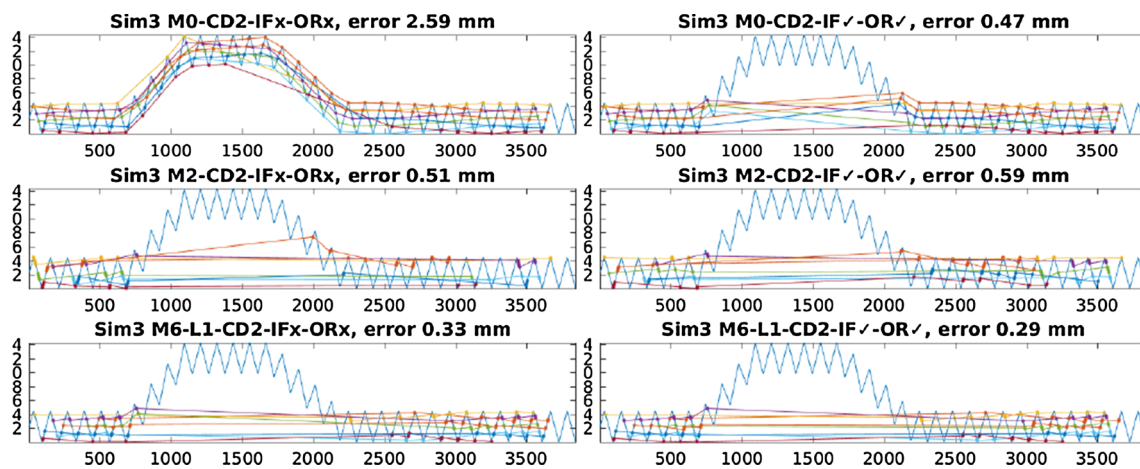
Fig. 5 Illustration of selected frames (*dots* connected by a line per phase) overlaid on motion trace for simulation Sim3, CD2 and (*left*) without and (*right*) with outlier removal (OR) and image filtering (IF) showing (*top to bottom*) M0, M2, and M6 results
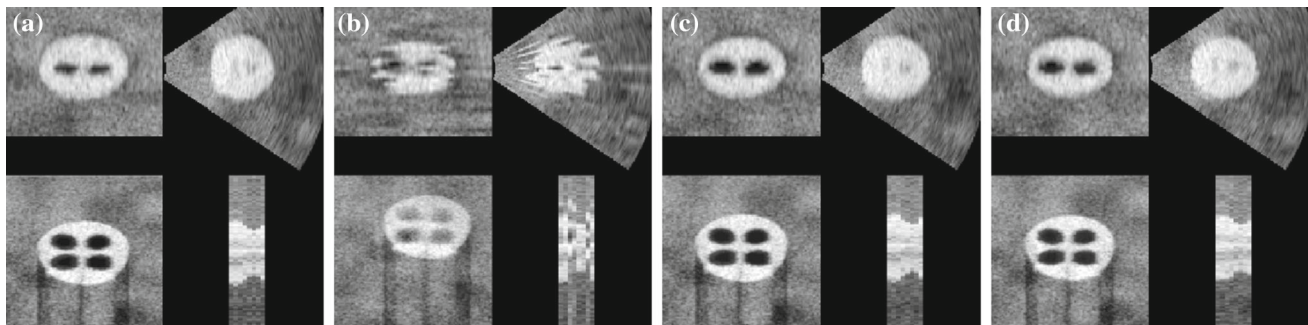
Fig. 6 Sample orthogonal slices and (*bottom-right*) M-mode image across 8 phases from reconstructions of Sim3 phase 3 for **a** ground truth, **b** baseline, **c** state-of-the-art, and **d** proposed method

and less crossing for M6, supporting its higher reconstruction accuracy.

Due to the consistent performance of M6–L1, the lower runtime for OR✓ and the slightly better performance of IF✓, we selected **M6–L1–CD2–IF✓–OR✓** as the best method of this study, which we call from now onwards **proposed method**. In all further tests, the proposed method is compared to the **baseline** (**M0–OR×**) and the **state-of-the-art** method [14] (**M2–CD2–IF×–OR✓**).

Figure 6 shows example reconstructions for Sim3. Artifacts can be observed for the baseline method across the combined frames. Reconstructions by the state-of-the-art and proposed method are very similar to the ground truth.
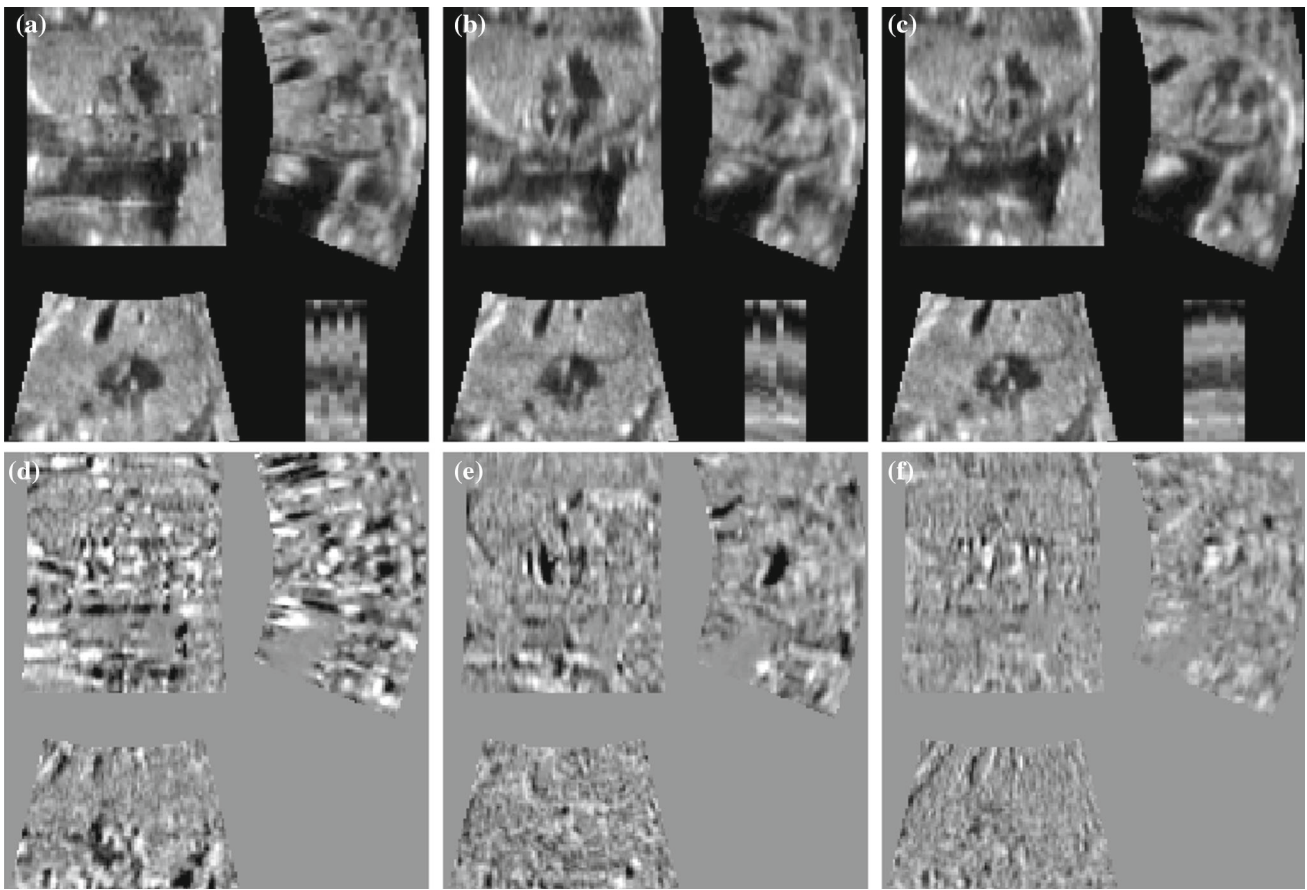
4D reconstruction of in vivo data

Temporal image quality The temporal quality of the 4D reconstructions by the baseline, the state-of-the-art, and the proposed method was blindly ranked by 5 observers (1 US specialist, 4 technical experts). Observers were shown movies of orthogonal heart slices from the 4D reconstruc-

tions, as shown in Fig. 7a–c, and asked to rank these (1: 'best', 2: 'second best', 3: 'worst') with respect to temporal image quality. The mean (standard deviation (SD)) of the ranks for these 3 methods pooled for the 5 observers was 2.83, 1.69, 1.42 (0.42, 0.69, 0.56), respectively. Figure 8a shows the distribution of the 5 mean ranks from the observers, with the result from the clinician following the overall pattern. Observers agreed completely on the ranking for case #8 and otherwise for 7 cases where the baseline method ranked third. The median rank of baseline method was statistically significantly different than the other two methods at the <0.0001 level by Wilcoxon signed rank test. Figure 7 shows sample reconstructions for #8, where misalignment artifacts are most reduced by the proposed method.

Clinical usefulness The 4D reconstructions of the baseline and the proposed method were then inspected for their clinical usefulness by the US specialist, who inspected the 4D volume interactively using the developed visualization interface, see Fig. 9. Clinically relevant planes, such as the four-chamber and outflow tract views, were found and the clinical usefulness of reconstructions on these planes was rated on

Fig. 7 Example of in vivo reconstruction where all observers agreed on rank (#8) for **a**, **d** baseline, **b**, **e** state-of-the-art and **c**, **f** proposed method for **a**–**c** phase 2 showing also (*bottom-right*) M-mode image across 8 phases and **d**–**f** difference phase 3–phase 2
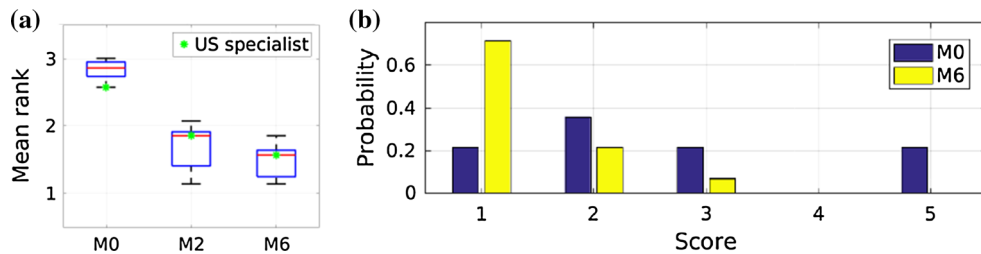
Fig. 8 a Boxplots showing distribution of temporal image quality mean rank per observer for baseline (M0), state-of-the-art (M2) and proposed method (M6) with green star for US specialist only. **b** Probability distribution of clinical usefulness score from 1: 'very useful' to 5: 'not useful at all'
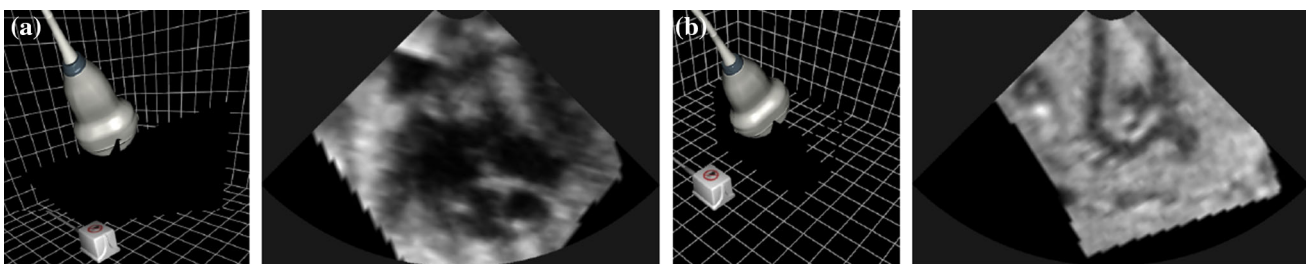
Fig. 9 Illustration of interactive tool for real-time extraction of planes from 4D volumes. (*left*) Position of electromagnetic tracker device, mock probe and plane. (*right*) Extracted plane **a** near four-chamber view from #1 and **b** for aortic arch view from #11

a Likert scale as 1: 'very useful,' 2: 'somewhat useful,' 3: 'neutral,' 4: 'not very useful,' or 5: 'not useful at all.' The mean score was 2.6 and 1.4 for the baseline and the proposed method, respectively. The reconstructions with the proposed method were very useful in 71%, somewhat useful in 21% and neutral in 7%, while the reconstructions by the baseline method were not useful at all in 21%, see Fig. 8b. The median scores of the two methods were statistically significantly different at the <0.012 level (Wilcoxon signed rank test).

Discussion and conclusion

We developed a fast reconstruction method, which improved quality as well as clinical usefulness of 4D fetal heart US images noticeable in comparison with neglecting the presence of fetal motion. Based on evaluations on simulated data, the most successful method optimized phase, spatial and temporal consistency in combination with a US-specific similarity measure (CD2) and a less restrictive cost for phase consistency (L1-norm). Note that this combined optimization allows for deviations from a regular heart rate. Its performance was confirmed by observer studies on in vivo data when comparing it to the baseline and the state-of-the-art method from the initial study [14].

The developed framework is suitable for continuous, long acquisitions. Dissimilarity calculation of neighboring slices (97% of runtime) is easily parallelizable. A real-time implementation can also use the outlier removal criterion 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.

Our interactive visualization interface was received very positively by the physician. 4D US reconstruction is hoped to aid the diagnosis of fetal heart malfunctions, also facilitating the navigation to clinically relevant planes through post-reconstruction interaction. Reconstructed volumes can also be used in image-based US simulations for medical training.

Compliance with ethical standards

Conflict of interest All authors declare no conflict of interest.

Human and animal rights All procedures performed in studies involving human participants were in accordance with the ethical standards of the provincial ethics committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Carvalho J, Allan L, Chaoui R, Copel J, DeVore G, Hecher K, Lee W, Munoz H, Paladini D, Tutschek B, Yagel S (2013) ISUOG Practice Guidelines (updated): sonographic screening examination of the fetal heart. Ultrasound Obstet Gynecol 41(3):348
2. Cikes M, Tong L, Sutherland G, Dhooge J (2014) Ultrafast cardiac ultrasound imaging: technical principles, applications, and clinical benefits. JACC Cardiovasc Imaging 7(8):812–823
3. Cohen B, Dinstein I (2002) New maximum likelihood motion estimation schemes for noisy ultrasound images. Pattern Recognit 35(2):455
4. DeVore G, Falkensammer P, Sklansky M, Platt L (2003) Spatio-temporal image correlation (STIC): new technology for evaluation of the fetal heart. Ultrasound Obstet Gynecol 22(4):380
5. Jansz M, Seed M, van Amerom J, Wong D, Grosse-Wortmann L, Yoo SJ, Macgowan C (2010) Metric optimized gating for fetal cardiac MRI. Magn Reson Med 64(5):1304–1314
6. Kainz B, Alansary A, Malamateniou C, Keraudren K, Rutherford M, Hajnal JV, Rueckert D (2015) Flexible reconstruction and correction of unpredictable motion from stacks of 2D images. In: Navab N, Hornegger J, Wells W, Frangi A (eds) Medical image computing and computer-assisted intervention—MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9350. Springer, Cham, pp 555–562. doi:10.1007/978-3-319-24571-3_66
7. Loizou C, Pattichis C, Christodoulou C, Istepanian R, Pantziaris M, Nicolaides A (2005) Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery. IEEE Trans Ultrason Ferroelect 52(10):1653
8. Loizou C, Theofanous C, Pantziaris M, Kasparis T (2014) Despeckle filtering software toolbox for ultrasound imaging of the common carotid artery. Comput Methods Programs Biomed 114(1):109
9. Mattausch O, Goksel O, Orcun (2016) Monte-Carlo ray-tracing for realistic interactive ultrasound simulation. In: Bruckner, Preim B, Vilanova A, Hauser H, Hennemuth A, Lundervold A (eds), Eurographics workshop on visual computing for biology and medicine (vcbm.20161285), The Eurographics Association. doi:10.2312/vcbm.20161285
10. Nelson T, Pretorius D, Sklansky M, Hagen-Ansert S (1996) Three-dimensional echocardiographic evaluation of fetal heart anatomy and function: acquisition, analysis, and display. J Ultrasound Med 15(1):1
11. Odille F, Bustin A, Chen B, Vuissoz PA, Felblinger J (2015) Motion-corrected, super-resolution reconstruction for high-resolution 3D cardiac cine MRI. In: Navab N, Hornegger J, Wells W, Frangi A (eds) Medical image computing and computer-assisted intervention—MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham, pp 435–442. doi:10.1007/978-3-319-24574-4_52
12. Peterfi I, Kellenyi L, Szilagyi A (2014) Noninvasive recording of true-to-form fetal ECG during the third trimester of pregnancy. Obstet Gynecol Int 2014. Article ID 285636
13. Schoisswohl A, Falkensammer P (2005) Method and apparatus for obtaining a volumetric scan of a periodically moving object. US Patent 6,966,878
14. Tanner C, Flach B, Eggenberger C, Mattausch O, Bajka M, Goksel O (2016) 4D Reconstruction of fetal heart ultrasound images in presence of fetal motion. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W (eds) Medical image computing and computer-assisted intervention—MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, vol 9900. Springer, Cham, pp 539–601. doi:10.1007/978-3-319-46720-7_69

15. Tanter M, Fink M (2014) Ultrafast imaging in biomedical ultrasound. IEEE Trans Ultrason Ferroelect 61(1):102–119

16. Uittenbogaard L, Haak M, Spreeuwenberg M, Van Vugt J (2008) A systematic analysis of the feasibility of four-dimensional ultrasound imaging using spatiotemporal image correlation in routine fetal echocardiography. Ultrasound Obstet Gynecol 31(6):625

17. Wachinger C, Yigitsoy M, Rijkhorst EJ, Navab N (2012) Manifold learning for image-based breathing gating in ultrasound and MRI. Med Image Anal 16(4):806

18. Yagel S, Benachi A, Bonnet D, Dumez Y, Hochner-Celnikier D, Cohen S, Valsky D, Fermont L (2006) Rendering in fetal cardiac scanning: the intracardiac septa and the coronal atrioventricular valve planes. Ultrasound Obstet Gynecol 28(3):266