

An Image-based method for phase estimation, gating and temporal super-resolution of cardiac ultrasound

*

*

Abstract. Ultrasound is emerging as an increasingly effective tool for rapid non-invasive assessment of cardiac structure and function. Knowledge of the phase of each video frame within the cardiac and/or respiratory cycles is essential in many applications and the ability to capture cardiac function at a sufficient temporal resolution is essential for accurate diagnosis. These tasks are particularly challenging in pre-clinical studies involving small animals with high heart and respiration rates. In this paper, we present a novel method for estimation of instantaneous cardiac and respiratory phases directly from cardiac ultrasound videos. The method transforms the complex high-dimensional image sequence into a univariate time series with the same periodicity characteristics, decouples the periodic sources of beating heart and respiratory motion using a trend extraction technique, and estimates the cardiac and respiratory phases using the Hilbert transform. We also present a robust non-parametric regression technique for gating out respiratory frames and a kernel regression model that reconstructs images at any cardiac phase and facilitates temporal super-resolution. We validate our methods using 2D cardiac ultrasound or echocardiography videos and ECG recordings of 6 mice.

1 Introduction

Cardiovascular disease is the leading cause of death worldwide and ultrasound is emerging as an increasingly effective tool for rapid non-invasive assessment of cardiac structure and function. Knowledge of the phase or location of each video frame within the cardiac and/or respiratory cycles is essential in many applications (e.g. gating [10], quiescence detection [11], 3D reconstruction [2]) and the ability to capture cardiac function at a sufficient temporal resolution is vital for accurate diagnosis. Typically, the position within the cardiac cycle is tracked through ECG data and position within the respiratory cycle is tracked by motion of markers placed on the subject's body [9]. Setting up such hardware is cumbersome particularly in pre-clinical studies involving small animals. Moreover, the frame rates of affordable commercial ultrasound transducers fall short [5] in imaging small animals such as mice with high heart (310-840 BPM) and respiration (80-230 BPM) rates.

Contributions. In this paper, we present a novel method that estimates the instantaneous cardiac and respiratory phases directly from the cardiac ultrasound

video(Figure 2). We also present a robust non-parametric regression technique for gating out respiratory frames and a novel kernel regression model for reconstructing images at any cardiac phase to facilitate temporal super-resolution.

Related work. Previous work on estimation of cardiac and/or respiratory phases directly from ultrasound videos is very limited. In [8], the authors compute a signal of the x- or y-coordinate of the center-of-mass of each frame, use bandpass filtering to remove components outside the cardiac frequency range, and apply matched sine/cosine filters to estimate the instantaneous phase. In [10], the authors compute a signal of phase correlation between consecutive frames and use an approach similar to [8] to recover the cardiac and respiratory signals. The center-of-mass signals are not reliable and phase correlation finds global translation that does not model beating and out-of-plane heart motion well.

2 Method

In this section, we present the theory underlying the proposed methods. In Section 2.1, we describe our method for estimation of instantaneous cardiac and respiratory phases. In Section 2.2, we present a robust method to exclude video frames with significant respiratory motion. In Section 2.3, we present a kernel regression model for reconstructing images at any cardiac phase to facilitate temporal super-resolution.

2.1 Estimation of instantaneous cardiac and respiratory phases

While there have been numerous efforts for the estimation of instantaneous phase and/or frequency in periodic univariate time series data [4], there are not many methods that tackle this problem in a multivariate setting such as the case of cardiac ultrasound videos wherein thousands of variables (pixel intensities) are involved. Our strategy is to transform this complex multi-variate problem into a univariate one and take advantage of existing methods to solve the problem.

We first compute the similarity between all pairs of images/frames in the given periodic image sequence containing N frames to create a symmetric matrix $A \in R^{N \times N}$ where in the element $A(i, j)$ is equal to the similarity between the i^{th} and j^{th} frame. Here in, we use normalized correlation (also known as the Pearson correlation coefficient) to quantify inter-frame similarity; but in principle, other image similarity metrics can be used. Each row in the inter-frame similarity matrix A can now be seen as a univariate time series. If the similarity metric is chosen with care and if the corresponding frame is not significantly corrupted, this time series will preserve the periodicity characteristics of the original image sequence. Figure 2(a) shows the inter-frame similarity matrix of one of our cardiac ultrasound videos wherein the periodicity of low-frequency respiratory motion and high-frequency beating heart motion can be observed.

Next, we use a trend extraction technique called the Hodrick-Prescott (HP) filter [1] to decompose the frame similarity signal $u^i(t)$ corresponding to each row i of the matrix A into a sum of two components: (i) Trend component $\tau_{resp}^i(t)$ with periodicity characteristic of only respiratory motion, and (ii) Residual component $r_{heart}^i(t)$ with periodicity characteristic of only beating heart motion. The

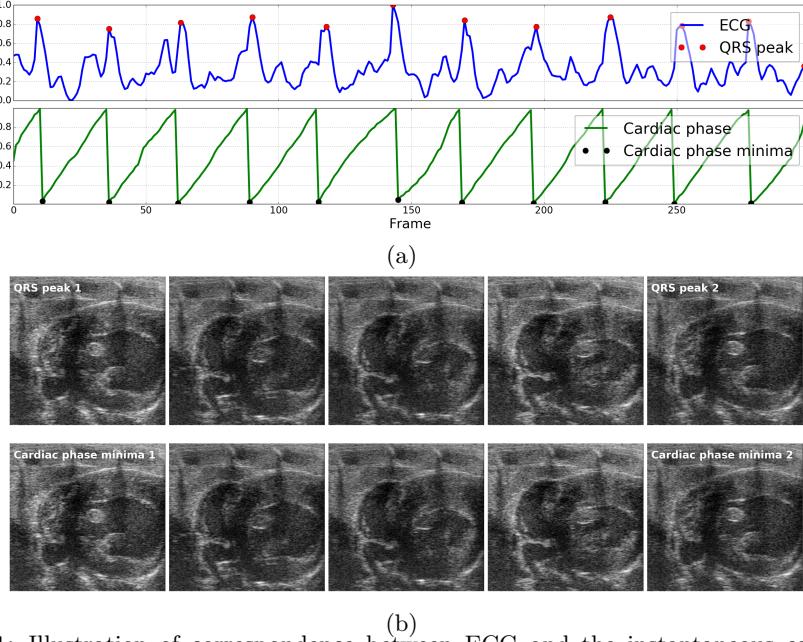


Fig. 1: Illustration of correspondence between ECG and the instantaneous cardiac phase estimated using our method: (a) ECG signal (blue) and the cardiac phase estimated using our method overlaid with peaks of the QRS complex (red circle) and the corresponding cardiac phase minima (black circle), (b) Five video frames evenly spaced in time between the first and second QRS peaks of the ECG signal (top-row) and the corresponding cardiac phase minima (bottom-row) constituting one cardiac cycle.

HP filter performs the decomposition of $u^i(t) = \tau_{resp}^i(t) + r_{heart}^i(t)$ by solving the following optimization problem:

$$\arg \min_{\tau_{resp}^i(t)} \left[\sum_{t=1}^N (u^i(t) - \tau_{resp}^i(t))^2 + \lambda \sum_{t=1}^{N-1} (\nabla^2 \tau_{resp}^i(t))^2 \right] \quad (1)$$

where $\nabla^2 \tau_{resp}^i(t) = \tau_{resp}^i(t+1) - 2\tau_{resp}^i(t) + \tau_{resp}^i(t-1)$ is the second-order difference or derivative of the trend signal and $\lambda = 6400$ is penalty parameter. Let A_{resp} and A_{heart} be the matrices whose rows contain the trend/respiratory and residual/heart-beat components (Figures 2(b,c)), respectively, of the frame similarity signal in the corresponding rows of matrix A . We now need to select the signals corresponding to one of the rows for phase estimation. To minimize the effect of any noise, we pick the row for which the periodogram or power-frequency distribution of the heart-beat component in A_{resp} has minimum entropy. Let $\hat{u}(t)$ be the selected frame similarity signal and let $\hat{\tau}_{resp}(t)$ and $\hat{r}_{heart}(t)$ be its associated trend and residual components (Figure 2(d)) that we will henceforth refer to as respiration and heart-beat signals, respectively.

Next, considering the narrow-band nature of the respiration and heart-beat signals (see Figure 2(e)), we use the Hilbert transform [4] to estimate the instantaneous phase of each frame. Specifically, we compute the instantaneous

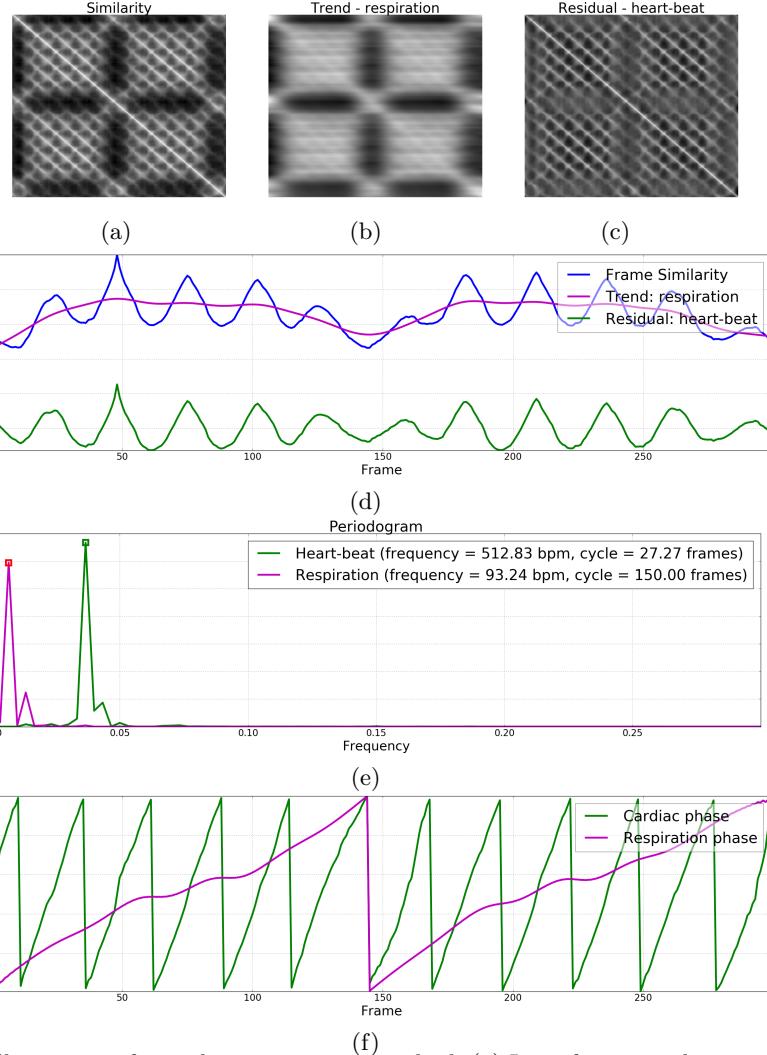


Fig. 2: Illustration of our phase estimation method: (a) Inter-frame similarity matrix, (b) Trend matrix corresponding to respiratory motion, (c) Residual matrix corresponding to beating heart motion, (d) Frame similarity $\hat{u}(t)$ selected for phase estimation with associated trend/respiration \hat{r}_{resp} and residual/heart-beat \hat{r}_{heart} signals, (e) Periodogram of heart-beat and respiration signals along with periodicity characteristics (e.g. frequency, cycle duration) calculated from the dominant frequency, and (f) Instantaneous cardiac and respiratory phases estimated using the Hilbert transform.

phase $\phi(t) \in [-\pi, \pi]$ of a periodic time series $x(t)$ using its Hilbert transform $H_x(t)$ as follows: $\phi(t) = \arctan\left(\frac{H_x(t)}{x(t)}\right)$ and map $\phi(t)$ from $[-\pi, \pi]$ to $[0, 1)$. Let $\hat{\phi}_{heart}(t)$ and $\hat{\phi}_{resp}(t)$ denote the instantaneous cardiac and respiratory phases (Figure 2(f)) computed from the respiration and heart-beat signals, respectively.

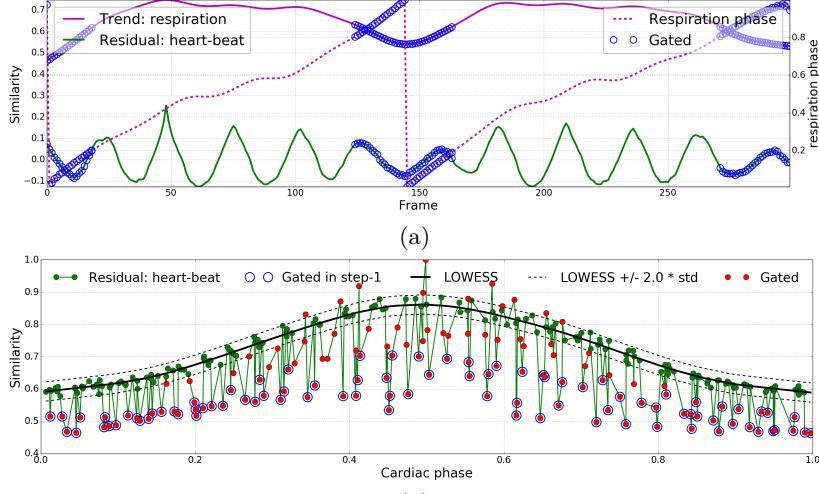


Fig. 3: Illustration of the respiratory gating method: (a) Frames F_{cutoff} discarded (blue circles) in step-1 overlaid with the respiration, heart-beat, and respiratory phase signals (b) heart-beat signal $\hat{r}_{heart}(t)$ vs cardiac phase $\hat{\phi}_{heart}(t)$ overlaid with frames F_{cutoff} discarded in step-1 (blue circle), LOWESS fit $L(\phi)$ (black solid), upper and lower bounds or 95% confidence interval ($L(\phi) \pm 2.0 * \hat{\sigma}_L$) of non-respiratory frames (black dotted), and frames F_{resp} (red circles) gated after step-2.

2.2 Gating out respiratory frames

Once the cardiac and respiratory phases of each frame have been estimated, they can be used to select or filter out frames from a desired part/point in the periodic cycle, a process commonly referred to as gating. In this section, we present a robust two-step method that uses these phase estimates to filter out video frames with significant respiratory motion.

In the first step, based on the observation that respiratory motion mainly occurs around the minima of the respiration signal $\hat{\tau}_{resp}(t)$ (pink curve in Fig. 2(d)) with respiratory phase $\hat{\phi}_{resp}(t) = 0$, we perform a rough initial gating by discarding the frames $F_{cutoff} = \{t \mid \hat{\phi}_{resp}(t) < c \vee \hat{\phi}_{resp}(t) > (1 - c)\}$ whose phase distance from $\hat{\phi}_{resp}(t) = 0$ is below a specified cutoff value $c = 0.2$. Figure 3(a) shows the discarded frames F_{cutoff} overlaid on the respiration $\hat{\tau}_{resp}(t)$, heart-beat $\hat{r}_{heart}(t)$, and respiratory phase $\hat{\phi}_{resp}(t)$ signals.

In the second step, we learn a robust mapping $L(\phi) : [0, 1] \rightarrow R$ from cardiac phase to the heart-beat component of the frame similarity signal by fitting a robust non-parametric regression model called Locally weighted regression (LOWESS) [6] to the dataset $\{(\hat{\phi}_{heart}(t), \hat{r}_{heart}(t)) \mid \forall t \notin F_{cutoff}\}$ containing the pair of the cardiac phase $\hat{\phi}_{heart}(t)$ and heart-beat signal $\hat{r}_{heart}(t)$ values of all frames that do not belong to the set of frames F_{cutoff} discarded in the first step above. Next, we compute a robust estimate of the standard deviation $\hat{\sigma}_L$ of non-respiratory frames around the LOWESS fit based on the

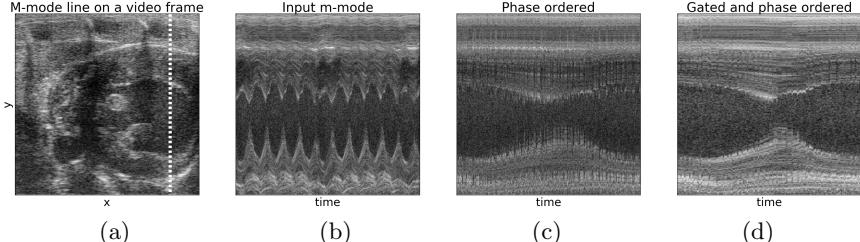


Fig. 4: Visualization of m-mode frames ordered by cardiac phase: (a) One of the frames in the video overlaid with the M-mode line shown in the next three images, (b) M-mode frames in the order they appear in the input video, (c) M-mode frames ordered by cardiac phase, and (d) Non-respiratory m-mode frames ordered by cardiac phase.

median absolute deviation between the LOWESS fit $L(\hat{\phi}_{heart}(t))$ and heart-beat signal $\hat{r}_{heart}(t)$ for all frames. Lastly, we gate out all frames $F_{resp} = \{t : |\hat{r}_{heart}(t) - L(\hat{\phi}_{heart}(t))| > k \times \hat{\sigma}_L\}$ whose heart-beat signal value deviates from the LOWESS fit by more than $k = 2.0$ (95% confidence interval) times the standard deviation $\hat{\sigma}_L$ (Figure 3(b)).

2.3 Model to generate images by cardiac phase for super-resolution

In this section, we present a kernel regression model to reconstruct the image at any cardiac phase. This model can then be used to generate a single-cycle video representative of the subject’s heart-beat at a higher temporal resolution.

Given a cardiac ultrasound video $I(t) : \{1, \dots, N\} \rightarrow R^m$ of N images with m pixels each, we extract the heart-beat signal $\hat{r}_{heart}(t)$, estimate the instantaneous cardiac phase $\hat{\phi}_{heart}(t)$ of each frame, and compute the robust LOWESS fit $L(\phi) : [0, 1] \rightarrow R$ that maps cardiac phase to the heart-beat signal as described in Sections 2.1 and 2.2. We then use Nadarya-Watson (NW) kernel regression [3] to learn a function $M(\phi) : [0, 1] \rightarrow R^m$ that reconstructs the image for any cardiac phase ϕ using a kernel-weighted local average as follows:

$$M(\phi) = \frac{\sum_{t=1}^N K(\phi, \hat{\phi}_{heart}(t)) I(t)}{\sum_{t=1}^N K(\phi, \hat{\phi}_{heart}(t))} \quad (2)$$

where kernel $K(\phi, \hat{\phi}_{heart}(t)) = \exp\left\{-\frac{|\phi - \hat{\phi}_{heart}(t)|^2}{2\sigma_\phi^2}\right\} \times \exp\left\{-\frac{|L(\phi) - \hat{r}_{heart}(t)|^2}{2\sigma_L^2}\right\}$ is defined as the product of two radial-basis function (RBF) kernels. The first RBF kernel weighs images inversely proportional to their distance (accounting for periodicity) in the cardiac phase space. Its bandwidth σ_ϕ is set equal to a constant $k_\phi = 0.4$ times the median difference in cardiac phase between consecutive frames of the given image sequence. The second RBF kernel weighs images inversely proportional to the deviation of their residual/heart-beat signal value from the LOWESS prediction $L(\phi)$. Its bandwidth σ_L is set equal to a constant $k_L = 2.0$ times the robust estimate of standard deviation $\hat{\sigma}_L$ of non-respiratory frames ($\forall t \notin F_{resp}$) around the LOWESS fit. The function $M(\phi)$ can now be used to reconstruct a single-cycle video representative of the subject’s heart-beat at any desired resolution/sampling of phase in the range $[0, 1]$.

Table 1: Match between frame positions of QRS peaks in ECG signal and corresponding cardiac phase minima.

VID	QRS peak frame identification	
	R^2	$mean \pm std$ error
1	0.9995	1.55 ± 0.89
2	0.9995	1.36 ± 1.15
3	0.9996	1.45 ± 0.89
4	0.9998	0.73 ± 0.62
5	0.9990	2.09 ± 1.78
6	0.9992	1.40 ± 1.74

Table 2: Evaluation of our kernel regression model. Leave one out cross-validation (LOOCV) of 50 rounds.

VID	LOOCV (50)	QRS Peak Frames
	$mean \pm std$ ncorr	$mean \pm std$ ncorr
1	0.8546 ± 0.0365	0.7981 ± 0.0532
2	0.8386 ± 0.0460	0.7877 ± 0.0503
3	0.8552 ± 0.0337	0.7849 ± 0.0457
4	0.8721 ± 0.0343	0.8395 ± 0.0486
5	0.8681 ± 0.0308	0.8196 ± 0.0476
6	0.8767 ± 0.0289	0.8270 ± 0.0778

3 Experiments and Results

We use cardiac ultrasound videos and associated ECG recordings of 6 mice to validate our methods. The ultrasound videos were acquired using the VisualSonics Vevo 2100 scanner at 233 frames per second. Each video consists of approximately 300 frames, 11 cardiac cycles, and 2 respiratory cycles.

We validate the cardiac phase estimates of our method by comparing them with the ECG signal which is the gold standard for cardiac gating. Figure 1(a) shows a comparison between the ECG signal and the estimated cardiac phase signal for one of the 6 videos. Notice how well the locations of the QRS peaks in the ECG signal and the corresponding minima of the cardiac phase signal in each cycle match. Table 1 reports the R^2 correlation and the $mean \pm std$ error between the locations of the QRS peaks and the corresponding minima of the cardiac phase signals for all 6 videos. Figure 1(b) presents a visual comparison of five video frames evenly spaced in time between two consecutive QRS peaks (top-row) and corresponding minima of the cardiac phase signal (bottom-row). Figure 4 shows frames of an m-mode line through the ventricle before and after ordering by cardiac phase. The supplementary material includes a video showing the cardiac and respiratory phases of each frame for one of our datasets.

We validate the accuracy of our kernel regression model for reconstructing images at any cardiac phase using leave-one-out-cross-validation (LOOCV). In each round of cross-validation, we randomly pick one of the non-respiratory frames in the video, exclude the selected frame along with corresponding frames (closest in phase) in each cycle, fit our kernel-regression model on the remaining frames, use the fitted model to reconstruct the image at the cardiac phase of the selected frame, and compute the similarity between the reconstructed and original image using normalized correlation. The second column of Table 2 shows the $mean \pm std$ of normalized correlation between the reconstructed and the original images over 50 rounds of LOCCV for all 6 videos. As a *baseline*, the third column of Table 2 reports the $mean \pm std$ of the normalized correlation between frames at QRS peaks of the ECG signal showing the effect of noise on similarity. The supplementary material includes single-cycle videos at 2x and 4x temporal resolution generated using the kernel regression model wherein the motion of heart chambers and valves is much clear than the original video.

4 Conclusion

In this paper, we have presented a novel method to estimate the instantaneous cardiac and respiratory phases directly from the cardiac ultrasound video, thereby eliminating the need of additional hardware to monitor them in, for example, small animal studies. We have also presented a robust non-parametric regression technique for gating out respiratory frames and a novel kernel regression model for reconstructing images at any cardiac phase to facilitate temporal super-resolution. We next plan to evaluate our methods on more datasets and address remaining pitfalls. Our phase estimation method makes a strong assumption of periodicity which may not hold in the case of subjects with cardiac arrhythmia. To address this, we will look into univariate methods for phase estimation in quasi-periodic signals that are less susceptible to noise. The use of normalized correlation to measure inter-frame similarity relies on an inherent assumption that a large part of the image is pulsating. To relax this, we will devise local patch-based similarity measures. Lastly, the local weighted average used by our NW kernel regression model may cause blurring at high temporal magnification. We plan to alleviate this using manifold kernel regression [7] that computes the weighted average in a diffeomorphic registration sense.

References

1. Alexandrov, T., Bianconcini, S., Dagum, E.B., Maass, P., McElroy, T.S.: A review of some modern approaches to the problem of trend extraction. *Econometric Reviews* 31(6), 593–624 (2012)
2. von Birgelen, C., de Vrey, E.A., Mintz, G.S., Nicosia, A., Bruining, N., Li, W., Slager, C.J., Roelandt, J.R.T.C., Serruys, P.W., de Feyter, P.J.: ECG-gated three-dimensional intravascular ultrasound. *Circulation* 96(9), 2944–2952 (1997)
3. Bishop, C.M.: *Pattern Recognition and Machine Learning*, vol. 4 (2006)
4. Boashash, B.: Estimating and interpreting the instantaneous frequency of a signal. ii. algorithms and applications. *Proceedings of the IEEE* 80(4), 540–568 (Apr 1992)
5. Chérin, E., Williams, R., Needles, A., Liu, G., White, C., Brown, A.S., Zhou, Y.Q., Foster, F.S.: Ultra-high frame rate retrospective ultrasound microimaging and blood flow visualization in mice *in vivo*. *Ultrasound Med. Biol.* 32(5), 683–691 (2006)
6. Cleveland, W.S., Devlin, S.J.: Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *JASA* 83(403), 596–610 (Sep 1988)
7. Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S.: Population shape regression from random design data. *IJCV* 90(2), 255–266 (2010)
8. Karadayi, K., Hayashi, T., Kim, Y.: Automatic image-based gating for 4D ultrasound. In: EMBC. pp. 2388–2391 (Aug 2006)
9. Khamene, A., Warzelhan, J.K., Vogt, S., Elgort, D., ChefHotel, C., Duerk, J.L., Lewin, J., Wacker, F.K., Sauer, F.: Characterization of internal organ motion using skin marker positions. In: MICCAI. pp. 526–533. Springer (2004)
10. Sundar, H., Khamene, A., Yatziv, L., Xu, C.: Automatic image-based cardiac and respiratory cycle synchronization and gating of image sequences. In: MICCAI. pp. 381–388. Springer (2009)
11. Wick, C.A., McClellan, J.H., Ravichandran, L., Tridandapani, S.: Detection of cardiac quiescence from b-mode echocardiography using a correlation-based frame-to-frame deviation measure. *IEEE J. Transl. Eng. Health Med.* 1, 1900211–1900211 (2013)