

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 visual assessment of cardiac structure and function. Knowing the phase or location of each video frame within the cardiac and/or respiratory cycles is essential in a variety of applications. In this paper, we present a novel method for the 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 Hilbert transform. 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. We validate our methods using 2D cardiac ultrasound videos and associated 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 visual assessment of cardiac structure and function. Cardiac ultrasound videos consists of two kinds of periodic motion, one that corresponds to the beating heart motion and the other that corresponds to respiratory motion. Knowing the phase or location of each video frame within the cardiac and/or respiratory cycles is essential in a variety of applications such as gating[2], quiescence detection[11], 3D reconstruction[10], temporal super-resolution and robotic beating heart surgery[7]. Typically, the position within the cardiac cycle is tracked through ECG data and the position within respiratory cycle is tracked by motion of markers placed on the subject's body[6].

In this paper, we present a novel method that estimates the instantaneous cardiac and respiratory phases directly from the the cardiac ultrasound video. We also present a robust non-parametric regression technique for gating out respiratory frames and a novel kernel regression to reconstruct images at any cardiac phase to facilitate temporal super-resolution.

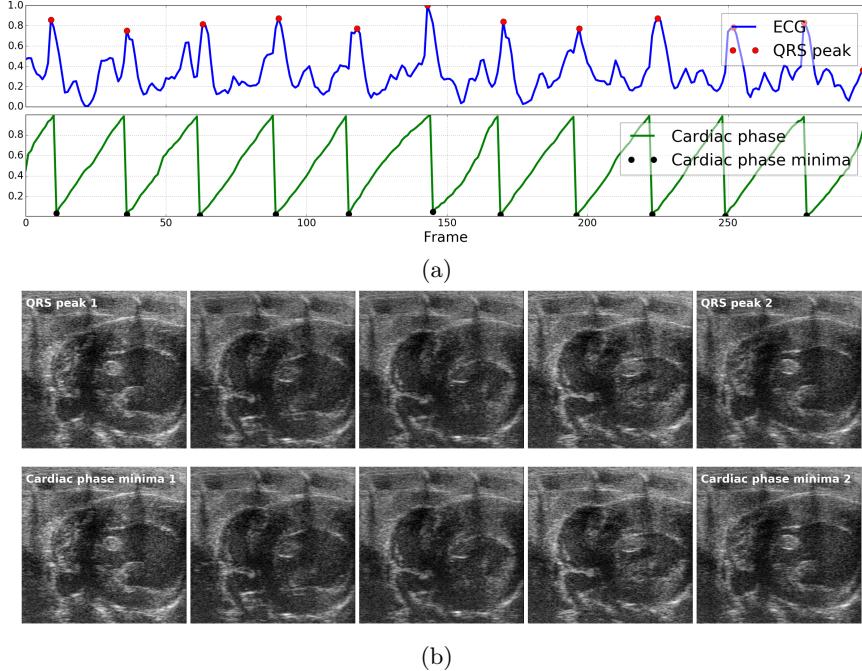


Fig. 1. Illustration of correspondence between ECG and the instantaneous cardiac phase estimated using our method: (a) ECG signal (blue) and the cardiac phase estimated using our method overlaid with peaks of the QRS complex (red circle) and the corresponding cardiac phase minima (black circle), (b) 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.

2 Method

In this section, we present the theory underlying the proposed methods for instantaneous phase estimation, gating and temporal super-resolution of cardiac ultrasound videos. Specifically, we begin by describing our method for the estimation of instantaneous cardiac and respiratory phases in Section 2.1. We then present a robust method that uses these phase estimates to gate out video frames with predominant respiratory motion in Section 2.2. Finally, in Section 2.3, we present a kernel regression model to generate the image at any cardiac phase that can be used to achieve 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, 8, 9], there are not many methods that tackle this problem in a multivariate setting such as the case of periodic cardiac ultrasound videos wherein several thousands of variables (pixel intensities) are involved. Our strategy was to find a way to transforming this complex multi-variate problem into a univariate one and take advantage of existing univariate methods to solve the problem.

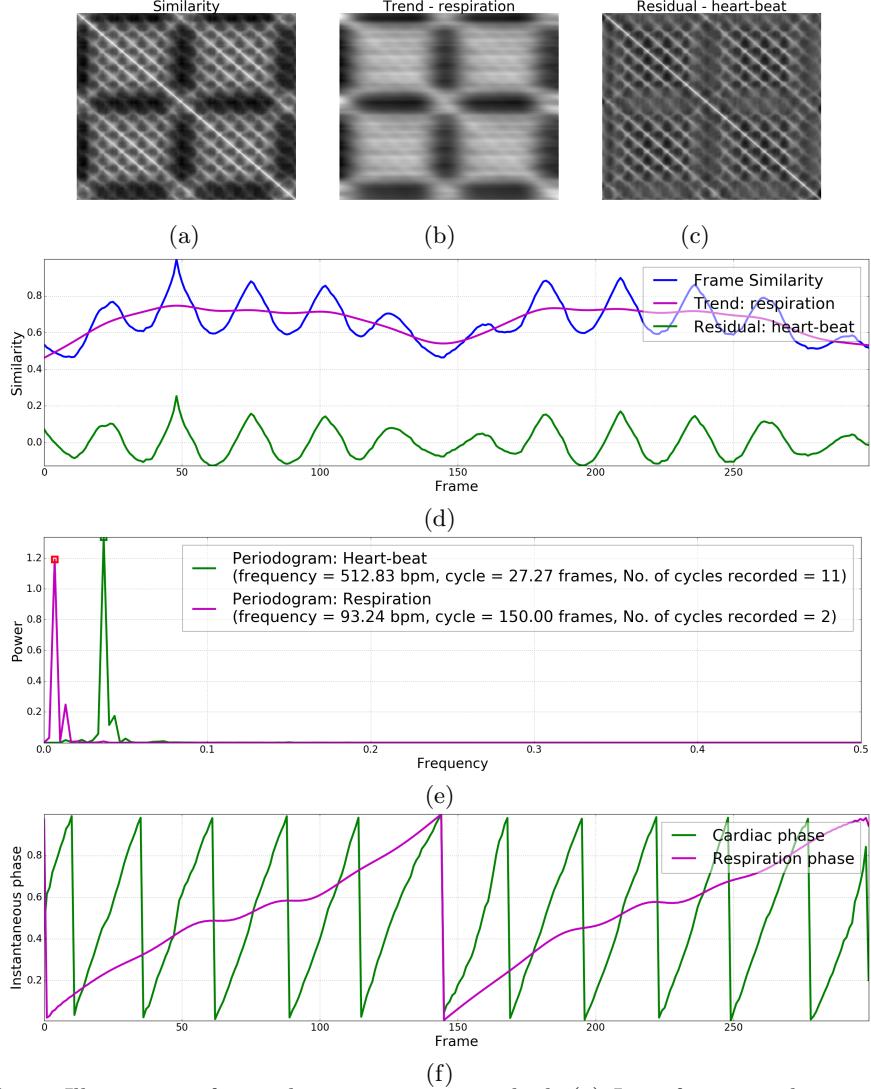


Fig. 2. Illustration of our phase estimation method: (a) Inter-frame similarity matrix, (b) Matrix of trend components corresponding to respiratory motion, (c) Matrix of residual components corresponding to beating heart motion, (d) Similarity profile chosen for phase estimation with associated trend/respiration and residual/heart-beat components, (e) Periodogram of heart-beat and respiration components 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.

We first compute the similarity between all pairs of frames in the given periodic image sequence containing N images/frames to create a 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} frames. Here in, we use normalization correlation also known as the Pearson

correlation coefficient to quantify inter-frame similarity but, in principle, any of the similarity metrics used in image registration algorithms can potentially be chosen. Each row in the inter-frame similarity matrix A can now be seen as a univariate time series. Furthermore, If the similarity metric is chosen with care and if the corresponding frame is not significantly corrupted, this time series can be expected to preserve the periodicity characteristics of the original image sequence. Two kinds of periodic motion are present in cardiac ultrasound videos, one corresponding to the beating heart motion and the other corresponding to respiratory motion. Figure 2(a) shows the inter-frame similarity matrix of one of our cardiac ultrasound videos wherein these two sources of periodicity can be clearly observed in many rows/columns.

Next, we use a trend extraction technique [1] called the Hodrick-Prescott (HP) filter [5], widely used in econometrics, to decompose the frame similarity signal $u(t)$ corresponding to each row of A into a sum of two components: (i) Trend component $\tau_{resp}(t)$ encompassing the periodicity characteristic of only the lower-frequency respiratory motion, and (ii) Residual component $r_{heart}(t)$ encompassing the periodicity characteristic of only the higher-frequency beating heart motion. The HP filter performs the decomposition of $u(t) = \tau_{resp}(t) + r_{heart}(t)$ by solving the following optimization problem:

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

where $\nabla^2 \tau_{resp}(t) = \tau_{resp}(t+1) - 2\tau_{resp}(t) + \tau_{resp}(t-1)$ is the second-order difference or derivative that penalizes curvature of the trend signal. We use a ridge-regression implementation of the HP filter [12] with $\lambda = 6400$. Let A_{resp} and A_{heart} be the matrices whose rows contain the trend/respiratory and residual/heart-beat components, respectively, of the frame similarity signals in the corresponding rows of the similarity matrix A . Figures 2(b,c) show the trend A_{res} and residual A_{heart} matrices for one of our cardiac ultrasound videos. To minimize the effect of any noise on phase estimation, we pick frame similarity signal corresponding to the row of A_{heart} whose periodogram or power-frequency distribution has minimum entropy. Let $\hat{u}(t)$ be the chosen frame similarity signal and let $\hat{\tau}_{resp}(t)$ and $\hat{r}_{heart}(t)$ be its associated trend/respiration and residual/heart-beat component signals, respectively. Figure 2(d) shows the selected frame similarity signal (blue) along with the associated trend/respiration (pink) and residual/heart-beat (green) components. Figure 2(e) shows the periodogram of these two components that are nicely segregated in frequency domain. The periodicity characteristics (e.g. frequency in beats-per-minute (bpm), cycle duration in frames, number of cycles) of heart-beat and respiration can be calculated based on the maximum-power frequency of their periodograms.

Next, considering the narrow-band nature of the trend/respiration and residual/heart-beat signals (see Figure 2(e)), we use the Hilbert transform [8] to estimate the instantaneous phase of each frame. Specifically, we compute the instantaneous intra-period phase $\phi(t) \in [-\pi, \pi]$ of the 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

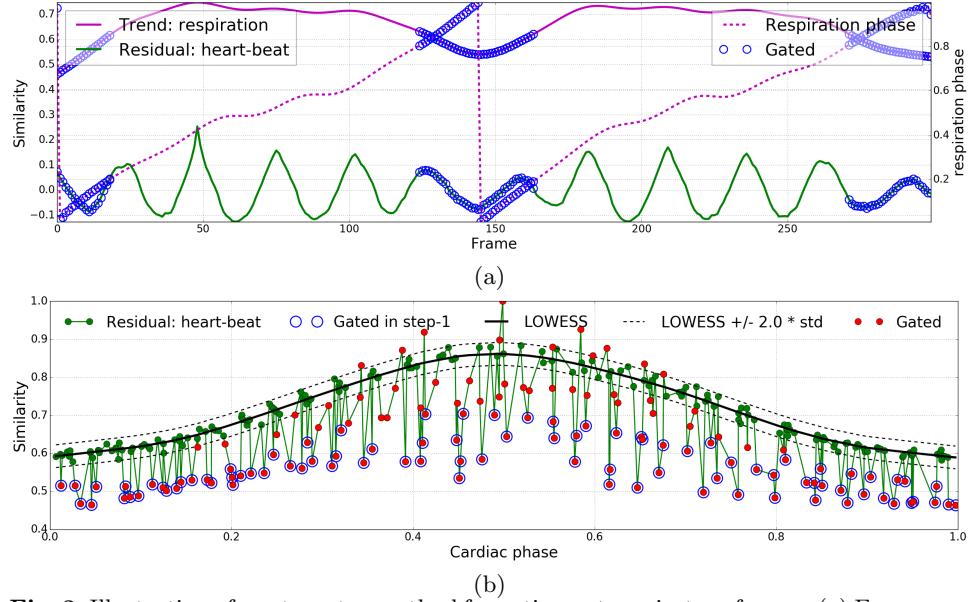


Fig. 3. Illustration of our two-step method for gating out respiratory frames: (a) Frames F_{cutoff} discarded (blue circles) in the first step overlaid with the trend/respiration, residual/heart-beat, and instantaneous respiratory phase signals (b) residual/heart-beat signal $\hat{r}_{heart}(t)$ vs 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 valid non-respiratory frames (black dotted), and frames F_{resp} (red circles) gated at the end of step-2.

the range $[-\pi, \pi]$ to $[0, 1]$. Let $\hat{\phi}_{heart}(t)$ and $\hat{\phi}_{resp}(t)$ denote the instantaneous cardiac and respiratory phases computed from the trend/respiration $\hat{\tau}_{resp}(t)$ and residual/heart-beat $\hat{r}_{heart}(t)$ signals, respectively. Figure 2(f) shows the instantaneous cardiac and respiratory phases computed from the heart-beat and respiration signals in Figure 2(d).

2.2 Gating out respiratory frames

Once the cardiac and respiratory phases of each frame have been estimated, it can be used to filter out or guide the extraction of quantitative measurements from a desired part/point of 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 gate out video frames predominantly influenced by respiratory motion.

In the first step, based on the observation that respiratory motion mainly occurs around the minima of the trend/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 $F_{cutoff} = \{t \mid \hat{\phi}_{resp}(t) < c \parallel \hat{\phi}_{resp}(t) > (1 - c)\}$ whose respiratory 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 (blue circles)

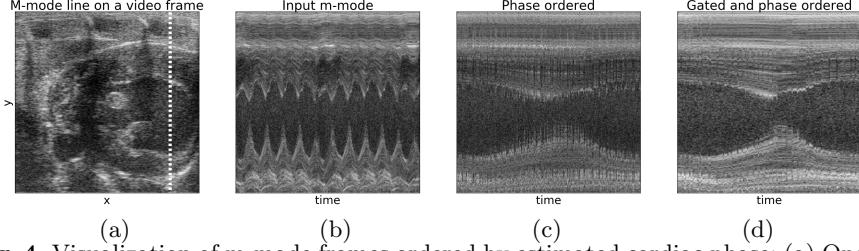


Fig. 4. Visualization of m-mode frames ordered by estimated cardiac phase: (a) One of the frames in the video overlaid with the M-mode line shown in the next three images, (b) M-mode frames in the order they appear in the input video, (c) M-mode frames ordered by estimated cardiac phase, and (d) Non-respiratory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

on the trend/respiration $\hat{\tau}_{resp}(t)$, residual/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 residual/heart-beat component of the frame similarity signal by fitting a robust non-parametric regression model called Locally weighted regression (LOWESS) to the dataset $\{(\hat{\phi}_{heart}(t), \hat{r}_{heart}(t)) \mid t \notin F_{cutoff}\}$ containing the pair of the cardiac phase $\hat{\phi}_{heart}(t)$ and residual/heart-beat $\hat{r}_{heart}(t)$ signal 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/dispersion $\hat{\sigma}_L$ of non-respiratory frames around the LOWESS fit based on the median absolute value of the difference between the LOWESS fit $L(\hat{\phi}_{heart}(t))$ and residual/heart-beat signal $\hat{r}_{heart}(t)$ for all frames. Lastly, we gate out all frames $F_{resp} = \{t \mid |L(\hat{\phi}_{heart}(t)) - \hat{r}_{heart}(t)| > k \times \hat{\sigma}_L\}$ for which the absolute difference between the LOWESS fit and heart-beat signal value is greater than $k = 0.2$ times the standard deviation $\hat{\sigma}_L$.

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 residual/heart-beat component $\hat{r}_{heart}(t)$ of the frame similarity signal, 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 kernel regression[3] to learn a phase-parametrized image manifold in the form of a function $M(\phi) : [0, 1] \rightarrow R^m$ that maps any cardiac phase ϕ to an image using kernel-weighted local average as follows:

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

where kernel $K\left(\phi, \hat{\phi}_{heart}(t)\right) = \exp\left\{-\frac{|\phi - \hat{\phi}_{heart}(t)|^2}{2\sigma_\phi^2}\right\} \times \exp\left\{-\frac{|L(\phi) - \hat{r}_{heart}(t)|^2}{2\sigma_L^2}\right\}$ 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 their deviation from the LOWESS prediction $L(\phi)$ in the residual/heart-beat signal space. Its bandwidth σ_L is set equal to a constant $k_L = 2.0$ times the robust estimate of standard deviation/dispersion $\hat{\sigma}_L$ of non-respiratory frames around the LOWESS fit as described in Section 2.2.

A representative single-cycle video of the subject’s heart-beat can now be reconstructed by sampling the manifold $M(\phi)$ at any desired resolution from the range $[0, 1]$ of the cardiac phase. Note that this approach will facilitate temporal super-resolution only if the set of images observed in each distinct period of the given image sequence differ in cardiac phase by some non-zero amount. Also, the amount of temporal super-resolution possible will depend both on the number of cycles captured and the amount corresponding images are perturbed in phase across cycles.

3 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.

VID	Correlation R^2	QRS peak prediction error $mean \pm std$
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 1. Comparision between frame positions of QRS peaks in the ECG signal and the corresponding minima of the estimated cardiac phase for 6 videos

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 head-to-head comparison between the ECG signal and the instantaneous cardiac phase signal obtained using the method described in Section 2.1 for one of the 6 videos. Notice how well the locations of the peaks of the QRS complex (red circles) in the ECG signal and the corresponding minima (black circles) 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 in our dataset. 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). Notice the high-level of similarity in the images of each column. Figure 4(b,c) shows the frames of an m-mode line before and after ordering with the estimated cardiac phase. Figure 4(d) shows the phase-ordered m-mode frames after performing respiratory gating as described in Section 2.2.

We validate the accuracy of our kernel regression model for reconstructing the image at any cardiac phase using leave-one-out-cross-validation (LOOCV). In each round of cross-validation, we randomly exclude one of the non-respiratory frames in the video, fit our kernel-regression model on the remaining frames, use the fitted model to reconstruct the image at the cardiac phase of the excluded frame, and compute the similarity between the reconstructed and original image using normalized correlation. The second column in Table 2 shows the mean and standard deviation of normalized correlation between the reconstructed and the original images over 50 rounds of LOOCV for each of the 6 videos in our dataset. As a baseline, the third column of Table 2 shows the mean and standard deviation of the normalized correlation between frames at QRS peaks of the ECG signal for each video.

Lastly, since this paper deals with video analysis we encourage the readers to review the supplementary material for a better understanding of the utility and effectiveness of the proposed methods.

VID	LOOCV (50) NCORR $mean \pm std$	QRS Peak Frame NCORR $mean \pm std$
1	0.8562 ± 0.0366	0.7981 ± 0.0532
2	0.8558 ± 0.0525	0.7877 ± 0.0503
3	0.8592 ± 0.0348	0.7849 ± 0.0457
4	0.8713 ± 0.0348	0.8395 ± 0.0486
5	0.8655 ± 0.0312	0.8196 ± 0.0476
6	0.8736 ± 0.0278	0.8270 ± 0.0778

Table 2. Evaluation of our kernel regression model for reconstructing the image at any cardiac phase using 50 rounds of Leave one out cross-validation (LOOCV)

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. 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. Having

demonstrated a proof of concept, we next plan to evaluate our methods on more datasets, especially of subjects with irregular heart beats, and address any remaining pitfalls. One drawback of our phase estimation method is a strong assumption of periodicity which may breakdown in the case of subjects with cardiac arrhythmia. To address this, we will look into generalizations of the hilbert transform[8] that are more susceptible noise. Another drawback is that the use 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 try to devise more sophisticated local patch-based similarity measures.

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. Hodrick, R., Prescott, E.: Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking* 29(1), 1–16 (1997)
6. Khameh, 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: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 526–533. Springer (2004)
7. Kurz, G., Hanebeck, U.D.: Heart phase estimation using directional statistics for robotic beating heart surgery. In: 18th International Conference on Information Fusion (Fusion). pp. 703–710 (July 2015)
8. Lu, W.k., Zhang, C.K.: Robust estimation of instantaneous phase using a time-frequency adaptive filter. *GEOPHYSICS* 78(1), O1–O7 (2013)
9. Luo, Y., Al-Dossary, S., Marhoon, M., Alfaraj, M.: Generalized hilbert transform and its applications in geophysics. *The Leading Edge* 22(3), 198–202 (2003)
10. Wachinger, C., Yigitsoy, M., Rijkhorst, E.J., Navab, N.: Manifold learning for image-based breathing gating in ultrasound and mri. *Medical image analysis* 16(4), 806–818 (2012)
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 journal of translational engineering in health and medicine* 1, 1900211–1900211 (2013)
12. Yamada, H.: Ridge regression representations of the generalized hodrick-prescott filter. *JOURNAL OF THE JAPAN STATISTICAL SOCIETY* 45(2), 121–128 (2015)