

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

Deepak Roy Chittajallu, Matthew McCormick, Samuel Gerber, Tomek Czernuszewicz, Ryan Gessner, Monte Willis, Marc Niethammer, Roland Kwitt, and Stephen Aylward

**Abstract**—Ultrasound is an effective tool for rapid non-invasive assessment of cardiac structure and function. Knowledge of the cardiac and/or respiratory cycle phase of each frame in the ultrasound video is essential in many diagnostic applications. Furthermore, many applications also require the ability to capture cardiac function at a high temporal resolution 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 the instantaneous cardiac and respiratory phases directly from cardiac ultrasound videos. The method transforms the sequence of high-dimensional images in the given ultrasound video into a univariate time series with similar 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 electrocardiogram (ECG) recordings of 6 mice. These methods generate an image series of a cardiac cycle with high temporal resolution from an ultrasound system with a limited frame rate.

**Index Terms**—Ultrasound, Echocardiography, Cardiac, Phase estimation, Gating, Temporal Super-resolution

## I. INTRODUCTION

Cardiovascular disease is the leading cause of death worldwide and ultrasound is an effective tool for rapid non-invasive assessment of cardiac structure and function [1]–[4]. 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 [5], quiescence detection [6], 3D reconstruction [7], robotic beating heart surgery [8]) and the ability to capture cardiac function at a high temporal resolution is vital for accurate diagnosis (e.g. valve motion assessment). Typically, the cardiac phase or position within the cardiac cycle is tracked by a simultaneously acquired ECG or pulse-oximetry data and the respiratory phase or position within the respiratory cycle is tracked by motion of markers placed on

D.R. Chittajallu, M. McCormick, S. Gerber, and S. Aylward are with Kitware Inc.

R. Kwitt is with the Dept. of Computer Science at University of Salzburg.  
M. Niethammer is with the Dept. of Computer Science at University of North Carolina - Chapel Hill

M. Willis is with the Dept. of Pathology and Laboratory Medicine at University of North Carolina - Chapel Hill

R. Gessner and T. Czernuszewicz are with SonoVol

the subject’s body [7], [9], [10]. Setting up such hardware is cumbersome particularly in pre-clinical studies involving small animals [1]. Moreover, the frame rates of affordable commercial ultrasound transducers fall short [11] 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 1). 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 prior work.** Previous work on the estimation of cardiac and/or respiratory phases directly from ultrasound echocardiography videos is limited. In [14], Karadayi *et al.* compute a signal of the x- or y-coordinate of the center-of-mass of each frame, use a band-pass filter to remove frequencies outside the cardiac range, determine the dominant frequency in the periodogram, and apply matched filtering using single-period sine and cosine signals at the dominant frequency to estimate the instantaneous cardiac phase. The issue with this technique is that the center-of-mass signals may not be reliable in all scenarios. In [5], Sundar *et al.* compute a signal of phase correlation between consecutive frames and apply a band-pass and low-pass filter to obtain an estimate of the instantaneous cardiac and respiratory phase, respectively. Phase correlation encodes global translation in the image plane and cannot model out-of-plane motion of the beating heart present in our data. In [15], Wachinger *et al.* use a manifold learning or non-linear dimensionality reduction technique called Laplacian Eigenmap to learn the low-dimensional manifold of the image sequence embedded in high-dimensional space and project the images onto first eigen direction of the Laplacian of the image similarity graph to obtain a 1D signal encoding respiratory motion. In [16], Panayiotou *et al.* use a series of image filtering operations to obtain a binary mask of pixels predominantly affected by cardiac/respiratory motion, apply a linear dimensionality reduction technique called principal component analysis (PCA) on the intensities of image pixels within the binary mask, project the images onto the principal directions with high variation to extract 1D signals encoding cardiac/respiratory motion, and post process them by suppressing undesired frequencies in

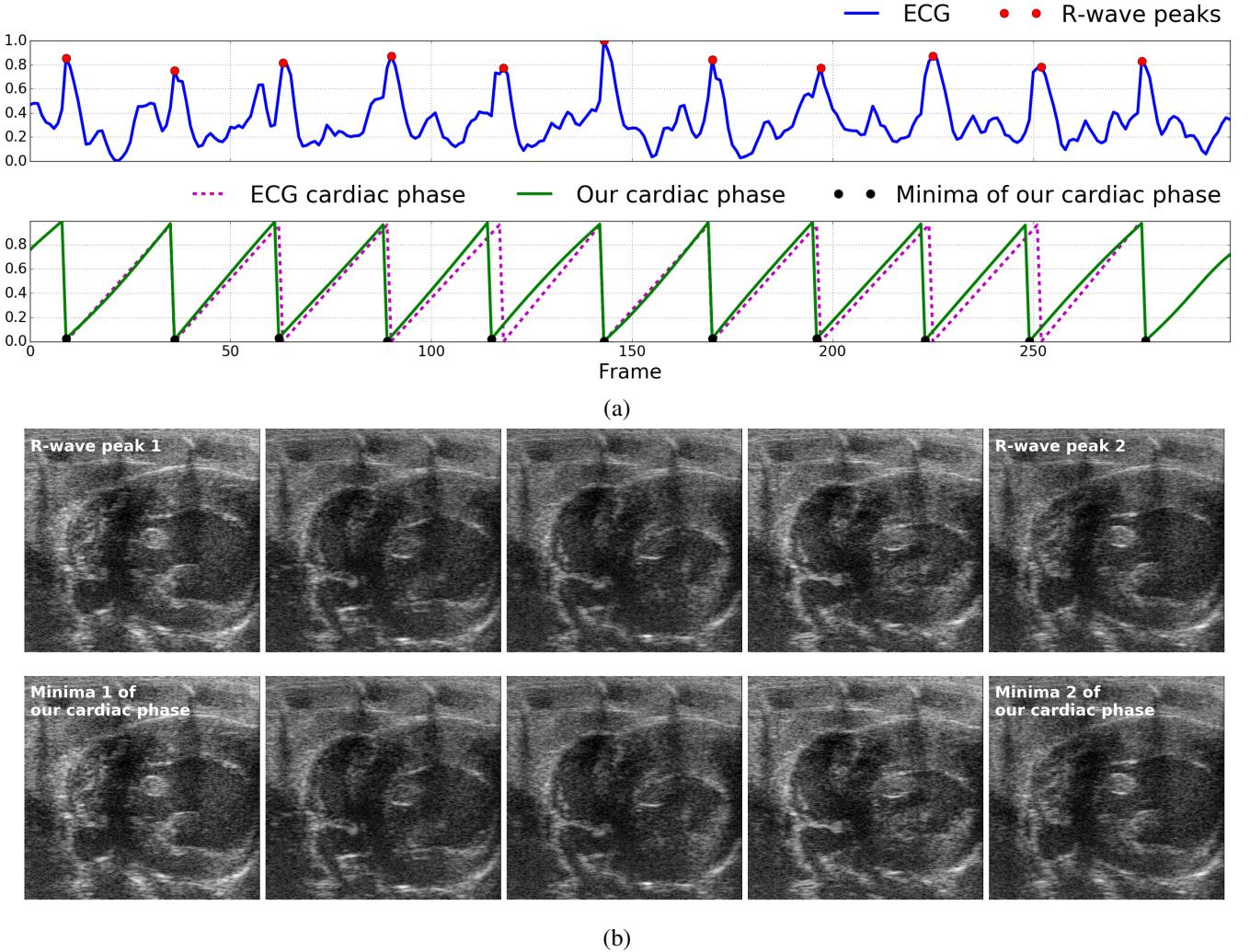


Fig. 1. Illustration of the close match between the cardiac phase derived from the ECG signal and the cardiac phase estimated using our method: (a) ECG signal simultaneously acquired with the image sequence (blue) overlaid with peaks of the R-wave in each cardiac cycle (red dot/circle), cardiac phase derived from the ECG signal (Pink dashed line) through linear interpolation between R-wave peaks [12], [13], and the cardiac phase estimated directly from the image using our method (green solid line) overlaid with the corresponding minima (black dot/circle), (b) Five video frames evenly spaced in time between the first and second R-wave peaks of the ECG signal (top-row) and the corresponding minima of the cardiac phase estimated using our method (bottom-row) constituting one cardiac cycle.

the frequency domain. While the aforementioned approaches based on manifold learning and masked PCA are promising and generally applicable, in this paper, we present a method that can estimate the cardiac phases more accurately. The proposed method extracts the univariate signal directly from the images and does not require an intermediate representation such as the PCA or manifold embedding of the image data.

## II. METHOD

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

### A. Estimation of 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 [12], [13], [17]–[19], 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 multivariate 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. A number of image similarity measures have been proposed in the context of medical image registration [20]. Here, we use normalized correlation to quantify

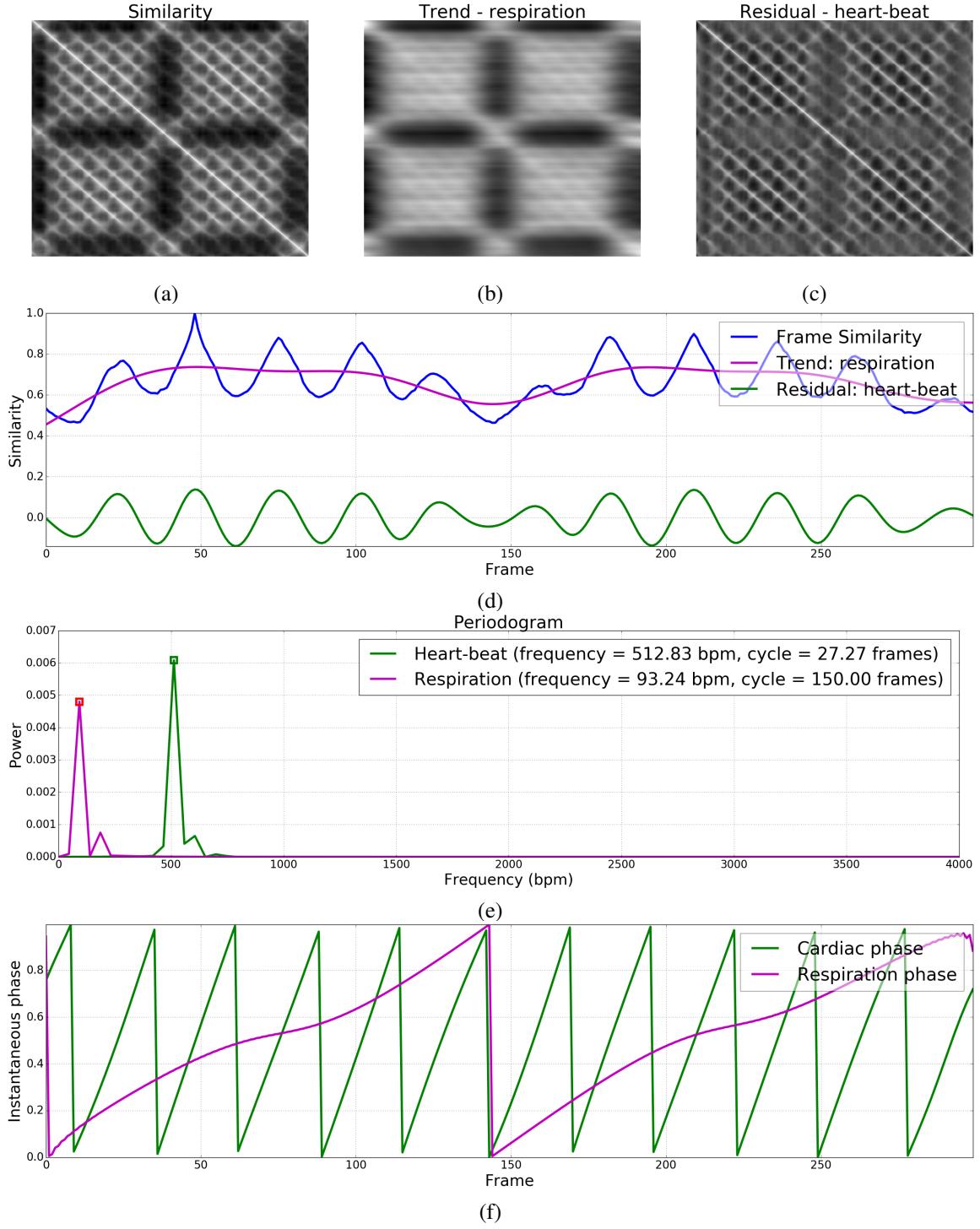


Fig. 2. Illustration of our phase estimation method: (a) Inter-frame similarity matrix, (b) Trend matrix corresponding to respiratory motion, (c) Residual matrix corresponding to beating heart motion, (d) Frame similarity  $\hat{u}(t)$  selected for phase estimation with associated trend/respiration  $\hat{\tau}_{resp}$  and residual/heart-beat  $\hat{\tau}_{heart}$  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.

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 [21] to decompose the frame similarity signal  $u^i(t)$  corresponding to each row  $i$  of the

sequence. Figure 2(b) shows the trend matrix corresponding to respiratory motion, which is a smooth curve over the frames. Figure 2(c) shows the residual matrix corresponding to heart-beat motion, which exhibits high-frequency oscillations. These three matrices together represent the decomposition of the original similarity matrix into its periodic components.

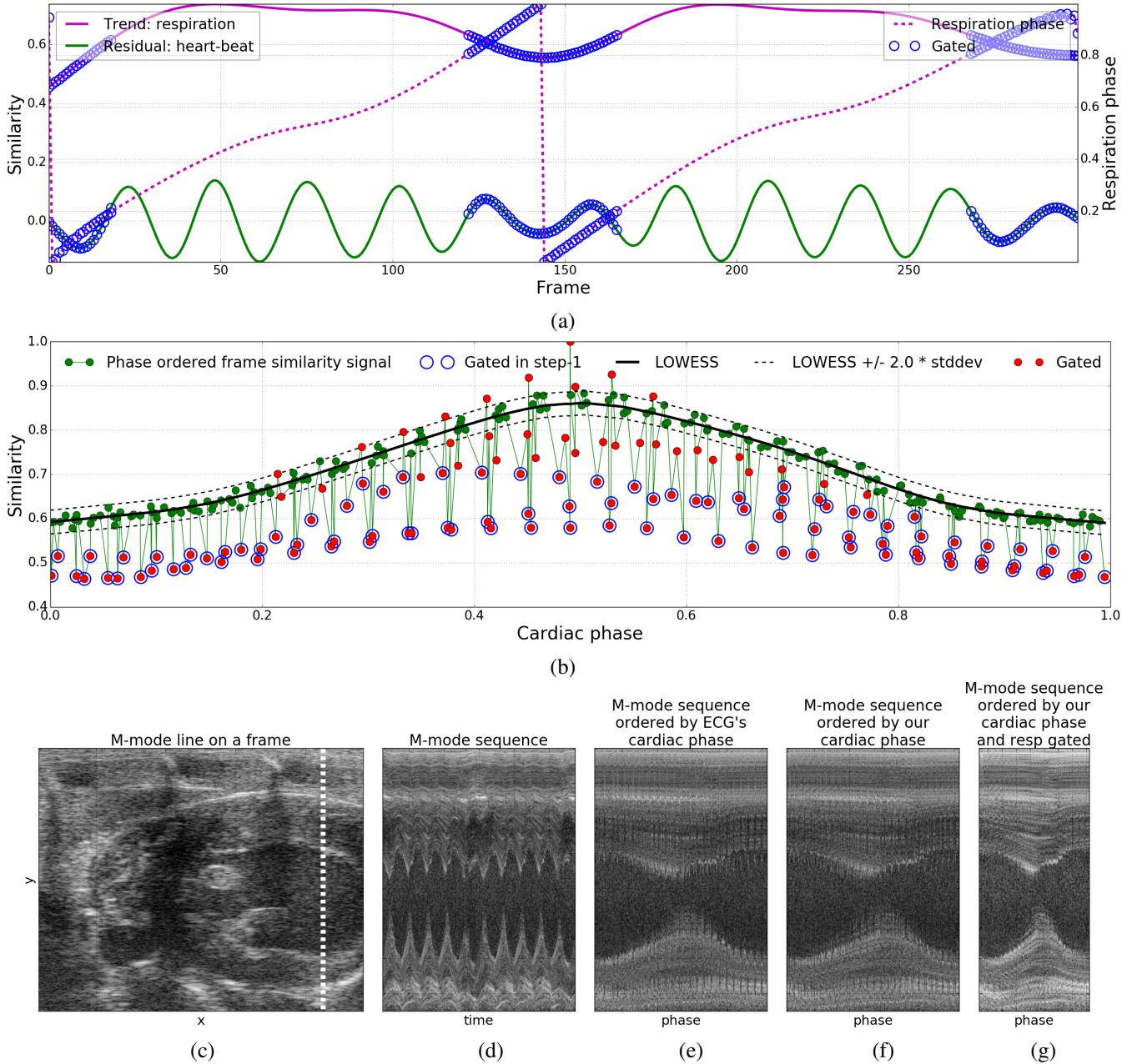


Fig. 3. Illustration of the respiratory gating method: (a) Frames  $F_{cutoff}$  discarded (blue circles) in step-1 overlaid with the respiration  $\hat{\tau}_{resp}(t)$ , heart-beat  $\hat{\tau}_{heart}(t)$ , and respiratory phase  $\hat{\phi}_{resp}(t)$  signals, (b) The frame similarity signal  $\hat{u}(t)$  vs cardiac phase  $\hat{\phi}_{heart}(t)$  overlaid with frames  $F_{cutoff}$  discarded in step-1 (blue circle), LOWESS fit  $L(\phi_{heart})$  (black solid), upper and lower bounds or 95% confidence interval ( $L(\phi_{heart}) \pm 2.0 * \hat{\sigma}_L$ ) of non-respiratory frames (black dotted), and frames  $F_{resp}$  (red circles) gated after step-2, (c) One of the frames in one of our cardiac ultrasound videos overlaid with the M-mode line shown in the next three images, (d) M-mode frames in the order they appear in the input video, (e) M-mode frames ordered by cardiac phase derived from the ECG signal, (f,g) M-mode frames ordered by cardiac phase estimated using our method before and after respiratory gating.

matrix  $A$  into a sum of two components: (i) Lower frequency trend component  $\tau_{resp}^i(t)$  with periodicity characteristic of only respiratory motion, and (ii) Higher frequency residual component  $r_{heart}^i(t)$  with periodicity characteristic of only beating heart motion. The HP filter performs the decomposition of  $u^i(t) = \tau_{resp}^i(t) + r_{heart}^i(t)$  by solving the following optimization problem:

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

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 a 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 then select the signals corresponding to one of the rows in  $A$  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. To suppress undesired frequencies, we apply a low-pass filter with a cutoff frequency of 230 BPM to the respiration signal  $\hat{\tau}_{resp}(t)$  and a band-pass filter within the frequency range of 310–840 BPM to the heart-beat signal  $\hat{r}_{heart}(t)$ .

Next, considering the narrow-band near-sinusoidal nature of the respiration and heart-beat signals (Figure 2(e)), we use the concept of the analytic signal involving the Hilbert transform as introduced by Gabor to estimate the instantaneous phase of each frame [12], [13], [22]–[24]. Specifically, we compute the instantaneous phase  $\phi(t) \in [-\pi, \pi]$  of a periodic time series  $x(t)$  to be the instantaneous phase of its analytic signal  $A_x(t) = x(t) + iH_x(t)$  involving 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.

### B. 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 to enable the creation of an image sequence containing only cardiac 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 regression function  $L(\phi_{heart}) : [0, 1] \rightarrow R$  to predict the frame similarity signal value for any cardiac phase by fitting a robust non-parametric regression model called Locally weighted regression (LOESS) [25] to the dataset  $\{(\hat{\phi}_{heart}(t), \hat{u}(t)) \mid \forall t \notin F_{cutoff}\}$  containing the pair of the cardiac phase  $\hat{\phi}_{heart}(t)$  and frame similarity signal  $\hat{u}(t)$  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 LOESS fit based on the median absolute deviation between the LOESS fit  $L(\hat{\phi}_{heart}(t))$  and frame similarity signal  $\hat{u}(t)$  for all frames. Lastly, we gate out all frames  $F_{resp} = \{t : |\hat{u}(t) - L(\hat{\phi}_{heart}(t))| > k \times \hat{\sigma}_L\}$  whose

frame similarity signal value deviates from the LOESS fit by more than  $k = 2.0$  (95% confidence interval) times the standard deviation  $\hat{\sigma}_L$  (Figure 3(b)). Figure 3(d) shows an m-mode view of one of our videos along the line shown in Figures 3(c). Figures 3(e,f) show the m-mode sequence ordered by the cardiac phase derived from the ECG signal and the cardiac phase estimated using our method, respectively, that look very similar. Notice the jaggedness/discontinuities along the phase dimension in both these images that is caused by respiratory motion. Figure 3(g) shows the result obtained by ordering the m-mode sequence using the cardiac phase estimated using our method and gating out the respiratory frames using the method described above. Notice the significant reduction the jaggedness/discontinuities after gating out respiratory frames spotted using our method.

### C. Model to reconstruct images by cardiac phase

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 estimate the instantaneous cardiac phase  $\phi_{heart}(t)$  of each frame and compute the robust LOESS fit  $L(\phi_{heart}) : [0, 1] \rightarrow R$  that maps cardiac phase to the heart-beat signal as described in Sections II-A and II-B. We then use Nadarya-Watson (NW) kernel regression [26] to learn a function  $M(\phi_{heart}) : [0, 1] \rightarrow R^m$  that reconstructs the image for any cardiac phase  $\phi$  using a kernel-weighted local average as follows:

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

where in we define the kernel  $K(\phi_{heart}, \hat{\phi}_{heart}(t)) = \exp\left\{-\frac{|\phi_{heart} - \hat{\phi}_{heart}(t)|^2}{2\sigma_\phi^2}\right\} \times \exp\left\{-\frac{|L(\phi_{heart}) - \hat{u}(t)|^2}{2\sigma_L^2}\right\}$  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  $\sigma_\phi$  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 frame similarity signal value from the LOESS prediction  $L(\phi_{heart})$ . Its bandwidth  $\sigma_L$  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 LOESS fit. Figure 4(a) shows bandwidth of the kernel in phase and similarity space computed as described above. The model  $M(\phi_{heart})$  can now be used to reconstruct a single-cycle video representative of the subject's heart-beat by generating images at any desired resolution/sampling of phases in the range  $[0, 1]$ . Figure 4(b) shows m-mode views of the single-cycle videos reconstructed at 1x, 2x, 4x, and 8x temporal magnification using the NW kernel regression model with a bivariate RBF kernel defined in both cardiac

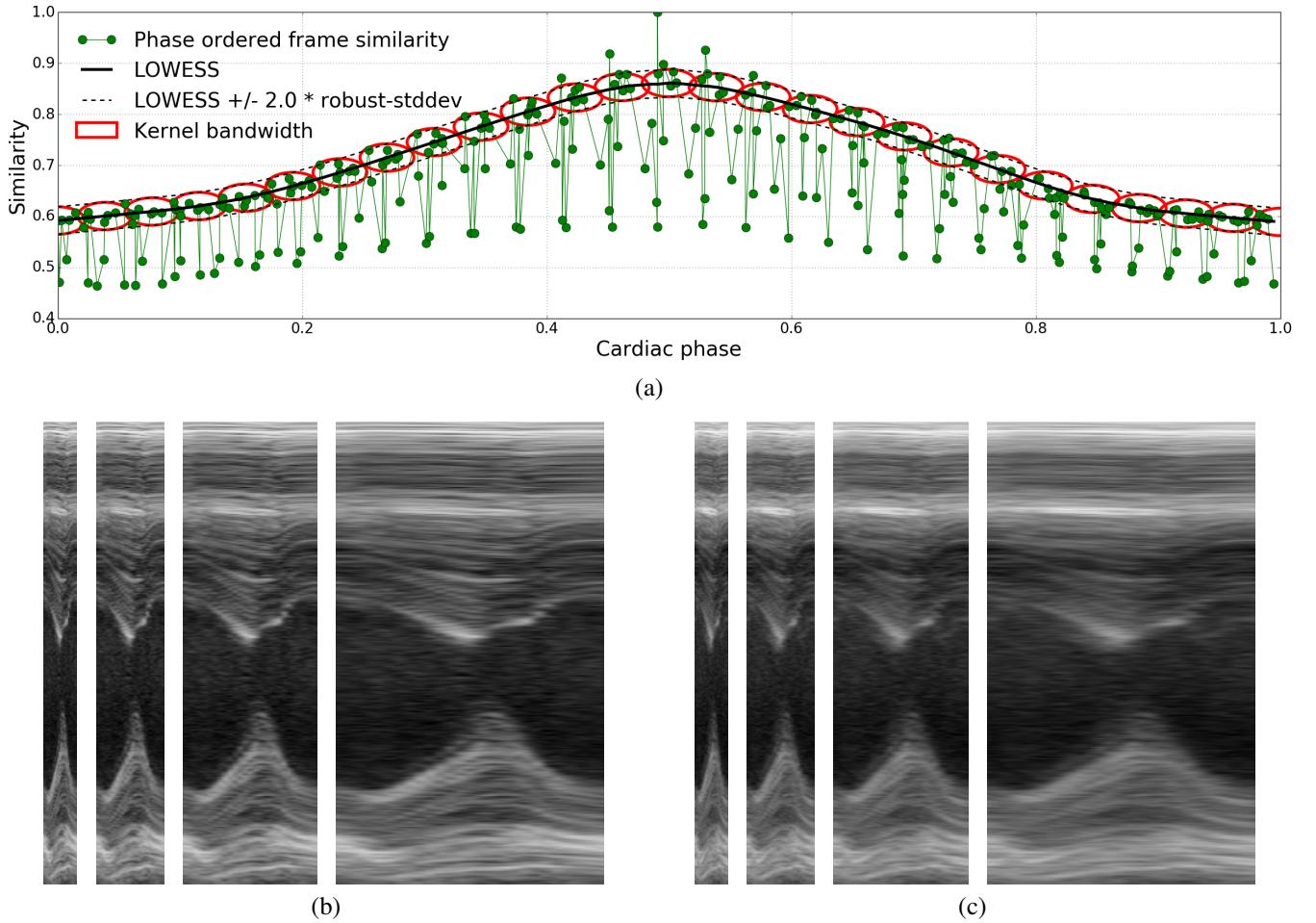


Fig. 4. Illustration of temporal super-resolution using our kernel regression model: (a) Shows the frame similarity signal  $\hat{u}(t)$  vs cardiac phase  $\hat{\phi}_{heart}(t)$  overlaid with LOWESS fit  $L(\phi_{heart})$  (black solid), upper and lower bounds or 95% confidence interval ( $L(\phi_{heart}) \pm 2.0 * \hat{\sigma}_L$ ) of non-respiratory frames (black dotted), and the kernel bandwidth shown at a series of cardiac phases as ellipses (red) whose length along the phase and similarity axis is set to  $\sigma_\phi$  and  $\hat{\sigma}_L$  as described in Section II-C, (b) M-mode view of single cardiac cycle videos reconstructed at 1x, 2x, 4x, and 8x temporal magnification using our NW kernel regression model with a bivariate kernel defined in both phase and similarity space, (c) M-mode view of single cardiac cycle videos reconstructed at 1x, 2x, 4x, and 8x temporal magnification using the NW kernel regression model with a kernel defined in phase space only.

phase and frame similarity phase. To justify the use of the proposed bivariate kernel, Figure 4(c) shows m-mode views of the single-cycle videos reconstructed at 1x, 2x, 4x, and 8x temporal magnification using the NW kernel regression model with only a univariate kernel defined in cardiac phase space. Observe that the reconstructions obtained using the bivariate kernel are much more crisp (less blurry) than the reconstructions obtained using the univariate kernel. This is due to the capability of effectively suppressing the effects of respiratory motion by the additional RBF kernel defined in frame similarity phase included in the bivariate kernel.

### III. RESULTS

We used 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 validated the cardiac phase estimates of our method by comparing them with those obtained from the ECG signal

which is the gold standard for cardiac gating. Figure 1(a) presents a visual comparison between the instantaneous cardiac phase derived from the ECG signal through linear interpolation between R-wave peaks [12], [13] and the instantaneous cardiac phase estimated directly from the image data using our method on one of the 6 videos. Notice that the cardiac phases estimated using our method match quite well with those obtained from the ECG signal. Furthermore, notice how well the locations of the peaks of the R-wave in the ECG signal and the corresponding minima of the cardiac phase signal estimated by our method match. Figure 1(b) presents a visual comparison of five video frames evenly spaced in time between two consecutive R peaks (top-row) and the corresponding minima of the cardiac phase signal computed using our method (bottom-row) which are visually very similar. Table I reports statistics of the error between the cardiac phases estimated by our method and those obtained from the ECG signal for all the 6 videos. Note that the duration of the time interval between two consecutive frames is 4.29 ms and the average length of cardiac cycle in our videos is 27.27

frames or 116.98 ms. The supplementary material includes a video showing the cardiac and respiratory phases of each frame for one of our datasets.

TABLE I

ERROR BETWEEN THE CARDIAC PHASES ESTIMATED USING OUR METHOD AND THOSE OBTAINED FROM THE ECG SIGNAL.

VID	$\text{mean} \pm \text{stddev}$	median	IQR	range
1	$0.06 \pm 0.03$	0.06	0.05	[0.00, 0.14]
2	$0.06 \pm 0.03$	0.05	0.05	[0.00, 0.12]
3	$0.06 \pm 0.03$	0.06	0.06	[0.00, 0.12]
4	$0.04 \pm 0.02$	0.04	0.02	[0.00, 0.10]
5	$0.06 \pm 0.03$	0.06	0.03	[0.00, 0.12]
6	$0.03 \pm 0.02$	0.04	0.03	[0.00, 0.05]

We also compared the performance of our phase estimation method with three previously published methods, namely: Phase correlation approach of Sundar *et al.* [5], Manifold learning approach of Wachinger *et al.* [15], and the Masked PCA approach of Panayiotou *et al.* [16]. For the phase correlation approach, we apply a bandpass filter in the cardiac frequency range of mice (310-840 bpm) to the phase correlation signal to extract the cardiac signal. For the manifold learning approach, we project the high-dimensional image sequence onto the second principal direction and apply a band-pass filter in the cardiac frequency range of mice (310-840 bpm) to extract the cardiac signal. For the Masked PCA approach, we compute the binary mask by thresholding the response of Frangi's vesselness filter tuned to enhance the moving heart chamber walls that look like ridges, perform PCA on the intensities of pixels within the binary mask, project the high-dimensional image sequence onto the second principal direction (projection onto the first principal direction only encodes respiratory motion in our data), and apply a band-pass filter in the cardiac frequency range of mice (310-840 bpm) to extract the cardiac signal. Note that each of these three prior works only propose a method for deriving a 1D signal reflecting the periodicity characteristics of the cardiac motion in the given image sequence. They do not provide an approach to compute the instantaneous cardiac phase from the derived signal. Hence, we compare the performance of these methods with ours in localizing the R-wave peaks of the ECG signal that correspond to the peaks/valleys of the 1D signals derived by these methods. Table II reports the  $\text{mean} \pm \text{stddev}$  error of different methods in locating the frames corresponding to the R-wave peaks of the ECG signal for all 6 videos.

TABLE II

ERROR IN LOCALIZATION OF THE FRAMES CORRESPONDING TO THE PEAKS OF THE R-WAVE IN ECG SIGNAL.

VID	mean $\pm$ stddev error in frames			
	Proposed	Phase corr [14]	Manifold learn [15]	Mask PCA [16]
1	<b>1.09 +/- 1.08</b>	4.09 +/- 2.97	1.45 +/- 0.78	1.55 +/- 0.78
2	<b>1.36 +/- 0.08</b>	2.91 +/- 2.23	1.82 +/- 0.94	1.73 +/- 1.05
3	<b>1.18 +/- 0.03</b>	3.45 +/- 1.97	1.45 +/- 1.08	1.55 +/- 0.78
4	<b>0.73 +/- 0.86</b>	4.64 +/- 3.75	1.36 +/- 0.98	1.18 +/- 1.03
5	<b>1.27 +/- 0.86</b>	4.70 +/- 3.77	<b>1.27 +/- 0.86</b>	1.36 +/- 0.88
6	<b>0.80 +/- 0.40</b>	6.00 +/- 3.52	1.00 +/- 0.63	1.10 +/- 0.54

We validated 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 III shows the  $\text{mean} \pm \text{stddev}$  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 III reports the  $\text{mean} \pm \text{stddev}$  of the normalized correlation between frames at R-wave peaks of the ECG signal. The supplementary material includes single-cycle videos at 1x, 2x, 4x, and 8x temporal magnification generated using the proposed kernel regression model wherein the motion of heart chambers and valves is clearer than the original video.

TABLE III

EVALUATION OF OUR KERNEL REGRESSION MODEL. LEAVE ONE OUT CROSS-VALIDATION (LOOCV) OF 50 ROUNDS.

VID	LOOCV (50)	QRS Peak Frames
	$\text{mean} \pm \text{stddev ncorr}$	$\text{mean} \pm \text{stddev ncorr}$
1	$0.8228 \pm 0.0311$	$0.6997 \pm 0.0605$
2	$0.8111 \pm 0.0250$	$0.7064 \pm 0.0527$
3	$0.8123 \pm 0.0256$	$0.7008 \pm 0.0421$
4	$0.8356 \pm 0.0302$	$0.7439 \pm 0.0612$
5	$0.8453 \pm 0.0254$	$0.7241 \pm 0.0591$
6	$0.8477 \pm 0.0370$	$0.7735 \pm 0.0695$

#### IV. 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 track 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 more resilient to noise [18], [19]. 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 explore 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 a manifold kernel regression approach where in the weighted average is computed in a diffeomorphic registration sense [27]. Source code<sup>1</sup> and data<sup>2</sup> used in the analysis are available online.

<sup>1</sup><https://github.com/KitwareMedical/CardiacUltrasoundPhaseEstimation>

<sup>2</sup><https://data.kitware.com/todo-add-collection-url>

## ACKNOWLEDGMENT

The proposed work is being supported through the NIH grants NIBIB/NIGMS:1R01EB021396-01A1 and NCI:5R44CA165621-04.

## REFERENCES

- [1] R. W. Cootney, "Ultrasound Imaging: Principles and Applications in Rodent Research," *ILAR Journal*, vol. 42, no. 3, pp. 233–247, Jan. 2001.
- [2] C. Scholten, R. Rosenhek, T. Binder, M. Zehetgruber, G. Maurer, and H. Baumgartner, "Hand-held miniaturized cardiac ultrasound instruments for rapid and effective bedside diagnosis and patient screening," *Journal of Evaluation in Clinical Practice*, vol. 11, pp. 67–72, 2005.
- [3] Y. Beaulieu, "Bedside echocardiography in the assessment of the critically ill," *Critical care medicine*, vol. 35, pp. S235–S249, 2007.
- [4] A. Oren-Grinberg, D. Talmor, and S. M. Brown, "Focused critical care echocardiography," *Critical care medicine*, vol. 41, pp. 2618–26, 2013.
- [5] H. Sundar, A. Khamene, L. Yatziv, and C. Xu, "Automatic image-based cardiac and respiratory cycle synchronization and gating of image sequences," in *MICCAI*. Springer, 2009, pp. 381–388.
- [6] C. A. Wick, J. H. McClellan, L. Ravichandran, and S. Tridandapani, "Detection of cardiac quiescence from b-mode echocardiography using a correlation-based frame-to-frame deviation measure," *IEEE J. Transl. Eng. Health Med.*, vol. 1, pp. 1 900 211–1 900 211, 2013.
- [7] C. von Birgelen, E. A. de Vrey, G. S. Mintz, A. Nicosia, N. Bruining, W. Li, C. J. Slager, J. R. T. C. Roelandt, P. W. Serruys, and P. J. de Feyter, "ECG-gated three-dimensional intravascular ultrasound," *Circulation*, vol. 96, no. 9, pp. 2944–2952, 1997.
- [8] G. Kurz and U. D. Hanebeck, "Heart phase estimation using directional statistics for robotic beating heart surgery," in *18th International Conference on Information Fusion (Fusion)*, July 2015, pp. 703–710.
- [9] A. Khamene, J. K. Warzelhan, S. Vogt, D. Elgort, C. ChefHotel, J. L. Duerk, J. Lewin, F. K. Wacker, and F. Sauer, "Characterization of internal organ motion using skin marker positions," in *MICCAI*. Springer, 2004, pp. 526–533.
- [10] C. Hennersperger, A. Karamalis, and N. Navab, "Vascular 3d+t freehand ultrasound using correlation of doppler and pulse-oximetry data," in *International Conference on Information Processing in Computer-Assisted Interventions*. Springer, 2014, pp. 68–77.
- [11] E. Chérin, R. Williams, A. Needles, G. Liu, C. White, A. S. Brown, Y.-Q. Zhou, and F. S. Foster, "Ultrahigh frame rate retrospective ultrasound microimaging and blood flow visualization in mice in vivo," *Ultrasound Med. Biol.*, vol. 32, no. 5, pp. 683–691, 2006.
- [12] M. Rosenblum, A. Pikovsky, J. Kurths, C. Schäfer, and P. A. Tass, "Phase synchronization: from theory to data analysis," *Handbook of biological physics*, vol. 4, pp. 279–321, 2001.
- [13] J. A. Freund, L. Schimansky-Geier, and P. Hänggi, "Frequency and phase synchronization in stochastic systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 13, no. 1, pp. 225–238, 2003.
- [14] K. Karadayi, T. Hayashi, and Y. Kim, "Automatic image-based gating for 4D ultrasound," in *EMBC*, Aug 2006, pp. 2388–2391.
- [15] 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, no. 4, pp. 806–818, 2012.
- [16] M. Panayiotou, A. P. King, R. J. Housden, Y. Ma, M. Cooklin, M. O'Neill, J. Gill, C. A. Rinaldi, and K. S. Rhode, "A statistical method for retrospective cardiac and respiratory motion gating of interventional cardiac x-ray images," *Medical physics*, vol. 41, no. 7, 2014.
- [17] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. ii. algorithms and applications," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 540–568, Apr 1992.
- [18] Y. Luo, S. Al-Dossary, M. Marhoon, and M. Alfaraj, "Generalized hilbert transform and its applications in geophysics," *The Leading Edge*, vol. 22, no. 3, pp. 198–202, 2003.
- [19] W.-k. Lu and C.-K. Zhang, "Robust estimation of instantaneous phase using a time-frequency adaptive filter," *GEOPHYSICS*, vol. 78, no. 1, pp. O1–O7, 2013.
- [20] T. S. Yoo, *Insight into images: principles and practice for segmentation, registration, and image analysis*. AK Peters Ltd, 2004.
- [21] T. Alexandrov, S. Bianconcini, E. B. Dagum, P. Maass, and T. S. McElroy, "A review of some modern approaches to the problem of trend extraction," *Econometric Reviews*, vol. 31, no. 6, pp. 593–624, 2012.
- [22] D. Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, 1946.
- [23] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*. McGraw-Hill New York, 1986, vol. 31999.
- [24] P. Kuklik, S. Zeemering, B. Maesen, J. Maessen, H. J. Crijns, S. Verheule, A. N. Ganesan, and U. Schotten, "Reconstruction of instantaneous phase of unipolar atrial contact electrogram using a concept of sinusoidal decomposition and Hilbert transform," *IEEE transactions on bio-medical engineering*, vol. 62, no. 1, pp. 296–302, Jan. 2015.
- [25] W. S. Cleveland and S. J. Devlin, "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *JASA*, vol. 83, no. 403, pp. 596–610, Sep. 1988.
- [26] C. M. Bishop, *Pattern Recognition and Machine Learning*, 2006, vol. 4.
- [27] B. C. Davis, P. T. Fletcher, E. Bullitt, and S. Joshi, "Population shape regression from random design data," *IJCV*, vol. 90, no. 2, pp. 255–266, 2010.