applied optics

Lock-in technique for extraction of pulse rates and associated confidence levels from video

Adam Eaton,[1] Karthik Vishwanath,[1,*] Chi-Hao Cheng,[2] E. Paige Lloyd,[3] and Kurt Hugenberg[3]

[1]Department of Physics, Miami University, Oxford, Ohio 45056, USA
[2]Electrical and Computer Engineering, Miami University, Oxford, Ohio 45056, USA
[3]Department of Psychology, Miami University, Oxford, Ohio 45056, USA
*Corresponding author: vishwak@miamioh.edu

Received 15 November 2017; revised 19 April 2018; accepted 22 April 2018; posted 26 April 2018 (Doc. ID 313564); published 21 May 2018

We investigate the practical applicability of video photoplethysmography (VPPG) to extract heart rates of subjects using noncontact color video recordings of human faces collected under typical indoor laboratory conditions using commercial video cameras. Videos were processed following three previously described simple VPPG algorithms to produce a time-varying plethysmographic signal. These time signals were then analyzed using, to the best of our knowledge, a novel, lock-in algorithm that was developed to extract the pulsatile frequency component. A protocol to associate confidence estimates for the extracted heart rates for each video stream is presented. Results indicate that the difference between heart rates extracted using the lock-in technique and gold-standard measurements, for videos with high-confidence metrics, was less than 4 beats per minute. Constraints on video acquisition and processing, including natural subject motion and the total duration of video recorded required for evaluating these confidence metrics, are discussed.

OCIS codes: (000.4430) Numerical approximation and analysis; (100.2960) Image analysis; (280.4991) Passive remote sensing; (330.4300) Vision system - noninvasive assessment.

https://doi.org/10.1364/AO.57.004360

1. INTRODUCTION

Photoplethysmography (PPG) is a well-established optical method that detects the time-dependent oscillations in the intensity of transmitted light through living tissues to derive cardiovascular information such as pulse rate, respiration, and blood oxygen saturation [1]. However, light that is diffusely backscattered from skin and imaged as reflectance from the exposed skin surface (detected in the visible near infrared region) can also be used to derive such temporal oscillations [2,3]. Acquisition and processing of color video data to compute pulsatile waveforms and thus derive the heart rate (HR) has been termed video photoplethysmography (VPPG) (or remote, noncontact, imaging, or camera photoplethysmography) [4–14].

Given that VPPG can operate nonobtrusively, has relatively low cost, and works with most natural (internal and external) lighting conditions, its uses in practical applications have been growing rapidly [15–17]. With the ubiquity of high-performance video cameras in cell phones, its uses as a phone-based app have also been investigated [18–22]. Applications of VPPG for remote telemedicine, human behavioral studies, exercise and sports medicine, neonatal monitoring, and clinical medicine have been explored [13,23–28].

In its basic form, VPPG is used to process a region of interest (ROI) in each video frame into one or more scalar values to generate one or more time-varying signals of specified durations. These time-domain signals are analyzed in the frequency domain to extract dominant frequencies (where a dominant frequency in the range of 0.8–3 Hz is expected to represent the HR). A popularly used set of techniques to build the VPPG-based pulsatile signal includes Independent Component Analysis (ICA) and related statistical techniques [29]. Studies have shown that to get best performance from these methods, several times, manual interventions are needed to correctly pick out the pulsatile signal from the derived components during data processing while fully automated functioning usually requires sophisticated processing methods [5,20,30–34]. Analytical expressions have been derived to calculate the pulsatile signal from video by several methods including the modified Beer–Lambert approach, spatial filtering, weighted ROI contributions, wavelet, and sparse matrix-based decomposition methods [6,28,30,35,36].

Most reports of VPPG typically have been focused on developing improved algorithms for generating robust, noise-free, pulsatile VPPG signals. Several of these have also sought a means to identify the true HR signal in inherently noisy VPPG signals and tend to employ computationally intensive processing techniques. The focus of our work was to investigate whether VPPG signals that can be generated almost directly

from the color video data could be processed to accurately identify the HR.

We present a simple method that is based on the physical principle of using higher order harmonics to lock into fundamental frequencies in noisy power spectra, as those typically obtained in VPPG data. Additionally, we also present a process to construct a confidence metric to assess the HR extracted using VPPG without any prior knowledge or gold-standard measurements. We describe our technique as a postprocessing filter that can be used by any VPPG algorithm that relies on Fourier-based analysis for identifying a dominant pulsatile frequency. We validate the performance of our algorithm by comparing it to commercial gold-standard measures (pulse oximeter and electrocardiograms) that were also used to measure HR in these experiments.

2. METHODS

A. Overview of VPPG

Burgeoning interest in VPPG has led to many reports describing different types of algorithms for extracting the HR from video data. All these studies have a set of basic (common) steps that need to be completed. These include preprocessing the video streams, tracking subject motion, identifying regions of interest (ROIs), converting RGB components into a single value, constructing a time signal for the video stream, applying smoothing or bandpass filters, and finally processing the time-domain signal to extract its dominant component. These steps have been schematically summarized as a workflow in Fig. 1. Each of these workflow steps are described in detail below.

B. Video Streams Used and Data Acquisition

Video data were acquired from 16 different participants to yield a total of 27 distinct video recordings. Subjects were composed of a racially diverse collection of male and female college students (including Caucasians, African Americans, and Asian Americans). Videos were all acquired indoors, under fluorescent illumination while the participants were seated facing the camera for the duration of the data acquisition. A light meter (Sekonic L-308S-U Flashmate) was used to ensure the room illumination was between 640 and 900 lumens for all video recordings. Figure 1 shows video recording environments for three representative videos used here. Thirteen video streams (with consents from subjects) have been made (see Supplemental Data: Visualization 1, Visualization 2, Visualization 3, Visualization 4, Visualization 5, Visualization 6, Visualization 7, Visualization 8, Visualization 9, Visualization 10, Visualization 11, Visualization 12, and Visualization 13).

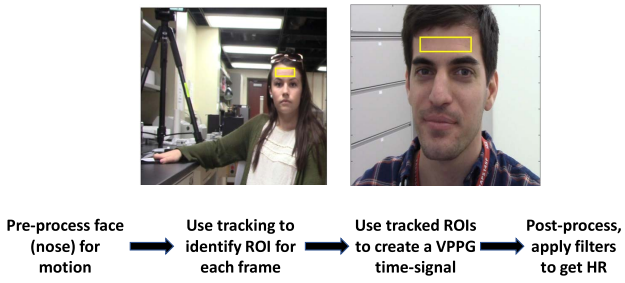

Fig. 1. Representative frames from two different videos to show recording conditions and environments. ROI locations are outlined in yellow. The workflow to extract HR from video using the VPPG technique employed is also shown.

Four of these participants were recorded under three different conditions—first the subject was asked to sit as still as possible without talking. In the second condition, the subject watched a video clip on a computer screen, and in the third condition the subject had a conversation (they responded to a questionnaire) with the experimenter. Each of these recordings lasted approximately 2 min, and they were obtained using a standard consumer video camera (Canon VIXIA HF R700). During acquisition of these videos, the participants wore a commercial pulse oximeter (Innovo 430J-PE) to get a gold-standard estimate of their HR.

Four video streams were collected from a different single participant, as a part of an ongoing behavioral study. These four recordings were acquired in two separate sessions on two different days. The first video acquisition in each session commenced after the participant had acclimatized to the room environment for about 10 min. In the first session, the first video was recorded as the participant spoke extempore about close friends. The second video in the session was recorded after the subject spent 5 min in relaxation (deep breathing) and then was asked to speak again about the same topic. In the second session, the recordings were obtained as the participant keep their head still on a support and remained silent. The first video in this session was recorded at baseline and the second after they were asked to walk up and down stairs. Gold-standard HR was acquired for these four recordings using EKG signals collected using a BIOPAC MP150 unit (BIOPAC Systems). The last 11 videos were obtained from 11 participants as they were each delivering a 5-min speech. Subjects sat facing the camera and were asked to speak naturally while the gold-standard HR was obtained using the commercial pulse oximeter.

All video recordings were made as the video camera was mounted on a tripod and were stored as MPEG-IV data. They had uniform frame rates of 29.9 fps. Videos were recorded with resolutions of 1920×1080 for the first set of 4 and last 11 subjects, and at 1280×720 for subject 5.

Stored video files were directly processed using MATLAB following the procedures described below. All studies were conducted under human subject protocols approved by the Miami University IRB.

C. Preprocessing: Motion Tracking and ROI Selection

Each of the 27 video files were first analyzed to reduce noise from natural movements and to ensure that the VPPG signals were derived from the same spatial regions, across frames. This was achieved by the Kanade–Lucas–Tomasi facial feature tracking available in MATLAB (Computer Vision Toolbox, Mathworks, Natick MA). The ROI (approximately 100×50 pixels) was manually selected for the first frame of all video streams processed such that it was situated on the forehead of the subject (see Fig. 1). Computer vision was then used to track two regions—the subject's face and the subject's nose in every frame, thus providing two rectangular regions that enclosed the face and the nose. The location of the ROI in each frame was determined such that its relative position from the

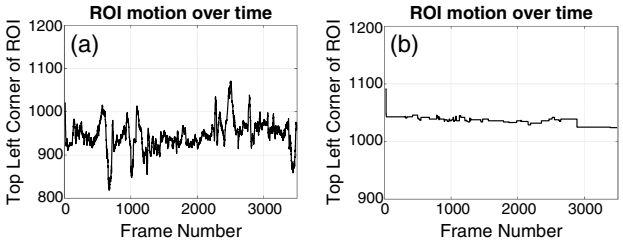

Fig. 2. Plots of the top left pixel of the ROI as a function of frame number. (a) shows a situation where the computer-vision-based tracking of the ROI was erratic, while (b) shows an example for smoothly tracked ROI.

centroid of the nose-bounding rectangle was the same across frames. The size of ROI in all frames was the same as the manually selected region in the first frame. Thus, even though the ROI was on the forehead, it was the nose feature that was tracked to position the ROI. This method is expected to be robust for translational motion of the subject's face but not to facial rotations. As described previously, failing to track subjects' motion from frame to frame could cause large errors in extracting the HR [37]. Figure 2 shows the location of the top-left pixel of the ROI for two video streams [Fig. 2(a): poorly tracked ROI; Fig. 2(b): accurately tracked ROI]. For analysis of the videos using VPPG, we use the variance in the location of the ROI to estimate the degree of movement for subjects in the videos as discussed below.

D. Constructing a VPPG Time Signal

1. Converting ROI in a Frame into a Single Value

The RGB values for each pixel in the ROI for each frame need to be converted into a single value. As noted above, there are several different algorithms that have been proposed to achieve this [17]. Here we use three such previously proposed algorithms that were computationally simple to implement.

The first method is a ratiometric technique derived as an approximation of the modified Beer–Lambert law, developed by Xu *et al.* [6]. In this method, the spatially averaged values of the red and green channels in ROI of two consecutive are used to derive the pulsatile signal, $(t) = \left[\log \frac{\bar{P}_R^2 \bar{P}_G^1}{\bar{P}_G^2 \bar{P}_R^1}, \log \frac{\bar{P}_R^3 \bar{P}_G^2}{\bar{P}_G^3 \bar{P}_R^2}, \ldots \log \frac{\bar{P}_R^n \bar{P}_G^{n-1}}{\bar{P}_G^n \bar{P}_R^{n-1}}\right]$. Following the original notation, the P_G^1 and P_R^1 refer to the spatially averaged ROI intensity values of the green and red channels (subscripts indicate color plane) of the first frame (superscripts indicate frame number), respectively [6].

The second algorithm constructs a VPPG time signal by calculating the mean value of all pixels in the green channel for the ROI, in each frame [4]. This represents the most direct means to access information from a single channel that would be spectroscopically expected to be sensitive to hemoglobin absorption in the tissue [13,38].

The third VPPG algorithm also used information from a single channel but converted the RGB color space to the hue-saturation-value (HSV) color space and then used the average of the hue (H) channel in the ROI of each frame to generate a time signal [39]. This method has been previously described as a better means of capturing spectroscopic information content from the red, green, and blue channels simultaneously [40].

The resulting temporal signal from each of the three methods was then used for extraction of the HR frequency either directly (no filtering) or after digital filtering using a finite impulse response bandpass filter (passband: 0.5 and 8 Hz).

2. Sliding Time Window

In any application of VPPG it is necessary to select a set of contiguous frames from which the time signal can be generated. The number of sequential frames needed, along with the total number of frames available in a data set, determine the number of ways such sets of contiguous frames can be constructed. In our data sets, we used 120 s of video that was acquired at nearly 30 fps, and thus had a total of 3600 frames. Therefore, there are 2500 different sets comprising 900 sequential frames (i.e., spanning 30 s each) that could be selected. Each of these 2500 sets can be used in turn to generate VPPG signals. Thus, we could generate 2500 different estimates of VPPG HR (from each of 2500 sets of time signals). The median value of the HR derived from each of these 2500 HR was taken to be the HR for that video stream. The idea of the sliding window is similar to the time-frequency analysis (including wavelets) described by other reports [4,35].

E. Lock-In Method to Extract the HR

The primary expectation in VPPG is that the dominant frequency component of the constructed time signal will yield the pulsatile rate. This is true when the temporal signal extracted from a video stream accurately represents a plethysmographic time signal. Most published reports recover the HR as the frequency at which the power spectrum of the VPPG time signal peaks. However, the VPPG time signal is inherently noisy, and the peak of the power spectrum for such noisy time signals does not correctly identify the HR (relative to a gold-standard measure) and is illustrated in Fig. 3.

Data in Fig. 3(a) shows the power spectra obtained from one of the participants (P3, sitting still condition), while Fig. 3(b) shows the power spectrum for another participant (P4, talking condition). Power spectra were obtained by Fourier transforming VPPG signals constructed from 30 s of each video stream, using the Xu algorithm [6]. Dashed vertical lines indicate HR values from pulse oximeter readings for the same 30 s durations.

Here, we present a method that seeks to identify the HR from such noisy data. We hypothesize that even in the presence of noise, a pulsatile frequency component would produce

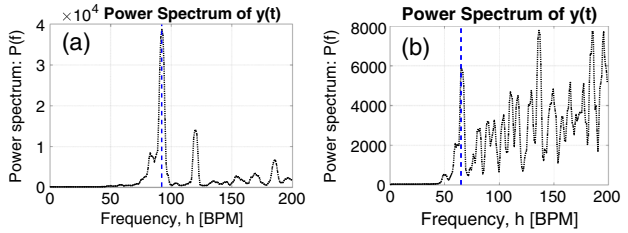

Fig. 3. Representative power spectra derived using two VPPG signals, from two different video streams, for two participants [(a) still and (b) talking]. Vertical dashed blue line shows the gold-standard HR values for these video streams.

higher order harmonics, while random noise would not have these harmonics. We call this technique of finding the HR in VPPG as the lock-in identification. The lock-in method operates by first calculating the power spectrum of any constructed VPPG time signal. Next, we select a set of (at most) 10 frequencies, f_i, where the calculated power spectrum had largest amplitudes. These were arranged in descending order of their corresponding amplitudes. For each of these peak-frequencies f_i, the power spectrum was analyzed to detect a local peak (where a local peak was defined as having a value higher than its two neighbors) closest to $2f_i$ (first harmonic) and closest to $3f_i$ (second harmonic)—denoted as ${}^2f_i$ and ${}^3f_i$. Note that ${}^2f_i$ need not be the same as $2f_i$. A scalar score s_i was then calculated for each peak frequency f_i, using the following equation:

$$s_i = \frac{1}{P(f_i)\Delta f}\begin{bmatrix} \max(\Delta f, |2f_i - {}^2f_i) \cdot P({}^2f_i) + \\ \max(\Delta f, |3f_i - {}^3f_i) \cdot P({}^3f_i) \end{bmatrix}. \quad (1)$$

In Eq. (1), $P(f_i)$ represents the amplitude of the power spectrum at each of the selected peak frequencies, while $P({}^2f_i)$ and $P({}^3f_i)$ are the amplitude of peaks detected closest to the second and third harmonics of f_i, Δf is the spacing of frequencies in the power spectrum, and the max function returns the higher of its two input arguments. Essentially, the score in Eq. (1) would have low values if the amplitude of the power spectrum $P(f_i)$ was large and small amplitude peaks were found at (or close to) the first and second harmonics (i.e., the max function returns Δf). Thus, using this lock-in method, the estimated HR was the frequency that had lowest score as calculated in Eq. (1).

3. RESULTS

A. Harmonics in Synthetic Data

Figure 4 illustrates the utility of using the lock-in scheme. A temporal signal was generated by repeating a random signal (of length T_0) N times [Fig. 4(a)], and its corresponding power spectrum was calculated [Fig. 4(b)]. It is apparent that the power spectrum of the periodic signal has peaks at harmonics of the fundamental frequency ($f_0 = T_0^{-1}$). Moreover, the largest amplitude in the power spectrum was found at $4f_0$. We tested 100 such synthetically generated signals (with fundamental frequency of 1.25 Hz to simulate a 75 BPM HR) using the lock-in technique. The temporal signals were analyzed with or without bandpass filtering (between 0.4 and 8 Hz). The lock-in technique correctly identified the fundamental frequency 90% of the time (the fundamental frequency f_0 was the tallest peak 8% of the time) with no bandpass filtering. The correct fundamental frequency was extracted 94% of the time by the lock-in technique for the bandpass filtered time signal (the fundamental frequency had largest amplitude 22% of the time).

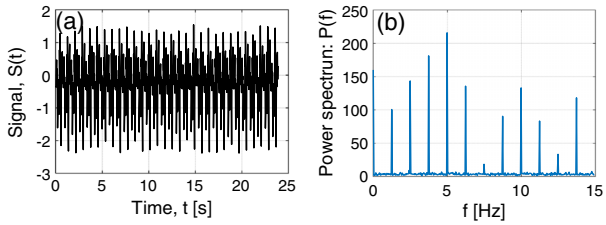

Fig. 4. Time signal formed by a short random signal (a) with period of 0.8 s that was repeated for a total duration of 30 s. (b) shows its corresponding power-frequency spectrum. This signal was sampled at 30 Hz to simulate the camera acquisition.

B. Processing with Sliding Windows

As discussed above, it is possible to use a short continuous segment (30 s) of the video stream to compute the VPPG HR while the video recording itself can be much longer (120 s). Figure 5 shows how a sliding window of 30 s can be used to analyze a 120 s long video stream (participant P2, still condition). The x axis denotes the midpoint of the time window used to construct the 30 s window, while the y axis is the HR derived for each 30 s interval (symbols—VPPG; line—pulse oximeter). The VPPG algorithm of Xu *et al.* [6] was used to derive the time signal, which was bandpass filtered, and the HR was obtained using the lock-in method.

A 2-min long video can therefore yield several estimates of HR as shown in Fig. 5. We use the median of these VPPG-derived values to provide a single value as the HR for the video and use the standard deviation of these values as a measure of uncertainty in its estimation (across the duration of 120 s). Thus, this sliding window analysis yielded both the HR and its variance. Figure 6 shows a correlation plot comparing the VPPG-derived median HR (using a 30 s sliding window to analyze 120 s of video data) and the gold-standard estimates of HR across the same 120 s interval, for all 27 videos. Here, the median VPPG HR for each video stream is plotted on the x axis while the gold-standard-derived HR on the y axis. The error bars for each point along each axis represent the variances in these values across the 120 s. The VPPG data were processed as in Fig. 5.

These data are also provided in tabular form in Table 1 for each video and list conditions under which the video data were collected, HR values derived using the peak of the Fourier power spectrum, HR values derived from the lock-in method, the gold-standard HR data, and the mean percent errors between the VPPG HR relative to gold-standard values. Videos for P1, P3, P4, and P5 listed in Table 1 are available as Visualization 1, Visualization 2, Visualization 3, Visualization 4,

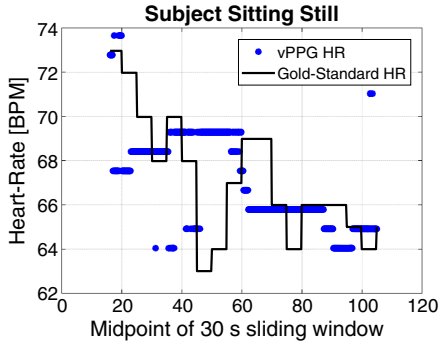

Fig. 5. Derived VPPG values of HR at each position of the sliding window (symbols) and the obtained gold-standard values during the same time (line). VPPG HR data are sampled every 1 s for clarity. The data in this figure was generated using the Xu algorithm.

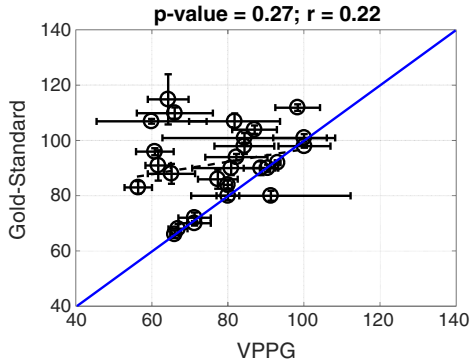

Fig. 6. Correlations between the gold-standard HR values (y axis) and the VPPG HR values (x axis) for all videos in the dataset. The dashed black line shows the line of correlation while the solid blue line shows the line $y = x$.

Visualization 5, Visualization 6, Visualization 7, Visualization 8, Visualization 9, Visualization 10, Visualization 11, Visualization 12, and Visualization 13.

The line of perfect correlation is the line with equation $y = x$ and is shown as the blue solid line in Fig. 6. However, the data in Fig. 6 clearly indicate that only a few videos yielded VPPG HR values that were near the gold-standard HR values. The correlation coefficient (r) and the p-values displayed in the figure indicate the merits of fit. In general, the closer r is to 1, the more perfectly are the data correlated. The p-value represents the likelihood of the null hypothesis (that there is no correlation between the data sets) being true. Thus, a low p-value indicates probability of the null hypothesis being false.

We observed that videos that had VPPG HR values close to those estimated by the gold standard had two features: the tracked ROI location for these videos had low variance [see for example, data shown in Fig. 2(b)]; and these video streams also had low VPPG HR variance for the sliding window analyses. Conversely, videos that had high ROI variances and high VPPG HR variances had VPPG HR values that were dissimilar from the gold-standard HR values.

C. Confidence Levels of Derived VPPG HR Values

Each of the 27 available videos were sorted into four mutually exclusive groups by thresholding variances observed in the ROI location and the VPPG HR, in increasing levels of confidence of the derived VPPG HR values. The first group contained videos where both the ROI and HR variances were larger than the respective threshold limits. The second group consisted of videos where the ROI variance was greater than the ROI threshold value and the VPPG HR variance was lower than the VPPG HR threshold. The third group videos included videos with ROI variance lower than the ROI threshold value while the VPPG HR variance was greater than the VPPG HR threshold value.

Table 1. Human Participants, Video Recording Conditions, and HR Values Obtained Using Different Methods under a Variety of Conditions

Subject ID	Motion Condition	True HR	Lock-In HR	FFT Peak HR	Error (%) (Lock-In)	Error (%) (FFT Peak)	σ_{ROI}	Group ID
P1[a]	Still	90 ± 5.0	87.7 ± 6.3	95.6 ± 26.7	2.6	6.2	16	4
	Watch	84 ± 3.1	80.7 ± 3.0	83.3 ± 10.9	3.9	0.8	12	4
	Talk	90 ± 5.1	82.5 ± 19.3	93.9 ± 28.4	8.3	4.3	5	3
P2[a]	Still	70 ± 1.7	71.1 ± 4.0	71.1 ± 27.4	1.6	1.6	5	4
	Watch	72 ± 3.8	71.9 ± 3.9	71.9 ± 31.6	0.1	0.1	122	2
	Talk	80 ± 3.6	76.3 ± 18.5	128.1 ± 14.1	4.6	60.1	176	1
P3[a]	Still	92 ± 1.5	93.0 ± 1.5	93.0 ± 1.5	1.0	1.0	7	4
	Watch	90 ± 2.5	90.4 ± 3.7	90.4 ± 2.0	0.4	0.4	19	4
	Talk	101 ± 2.7	100.0 ± 15.7	101.8 ± 6.3	1.0	0.8	9	3
P4[a]	Still	66 ± 2.7	65.8 ± 2.1	119.3 ± 24.2	0.3	80.8	4	4
	Watch	68 ± 3.4	66.7 ± 5.7	119.3 ± 25.2	1.9	75.4	11	4
	Talk	86 ± 6.9	77.2 ± 9.3	119.3 ± 21.9	10.2	38.7	2	4
P5[b]	Talk (pre)	94 ± 2.4	86.3 ± 15.1	102.5 ± 20.3	8.2	9.0	66	1
	Talk (post)	98 ± 5.7	91.2 ± 15.1	119.3 ± 15.2	6.9	21.7	44	1
	Still (pre)	98 ± 1.2	99.2 ± 16.0	119.3 ± 10.2	1.2	21.7	205	1
	Still (post)	112 ± 2.3	98.3 ± 14.1	119.3 ± 8.6	12.2	6.5	205	1
P6[a]	Talk	110 ± 4.1	65.8 ± 20.0	135.2 ± 32.8	40.2	22.9	92	1
P7[a]	Talk	91 ± 11.0	61.5 ± 6.7	119.3 ± 21.2	32.4	31.1	85	2
P8[a]	Talk	101 ± 3.4	84.3 ± 43.4	157.1 ± 26.4	16.5	55.5	47	1
P9[a]	Talk	83 ± 4.5	56.2 ± 7.3	119.3 ± 30.0	32.3	43.7	34	2
P10[a]	Talk	107 ± 5.7	81.6 ± 24.0	160.7 ± 26.0	23.7	50.2	28	1
P11[a]	Talk	104 ± 3.0	86.9 ± 11.8	137.8 ± 27.6	16.4	32.5	33	1
P12[a]	Talk	80 ± 3.3	79.9 ± 6.0	90.4 ± 22.2	0.1	13.0	25	2
P13[a]	Talk	96 ± 2.6	60.6 ± 9.9	119.3 ± 19.6	36.9	24.3	23	2
P14[a]	Talk	88 ± 7.2	65.0 ± 12.5	77.3 ± 25.8	26.1	12.2	117	1
P15[a]	Talk	107 ± 1.7	59.7 ± 29.0	123.8 ± 34.3	44.2	15.7	162	1
P16[a]	Talk	115 ± 18.2	64.1 ± 10.7	119.3 ± 5.6	44.3	3.7	274	1
					Mean error = 14.0%	Mean error = 23.5%		

[a]Gold-standard HR obtained using finger pulse oximeter.
[b]Gold-standard HR obtained using EKG.

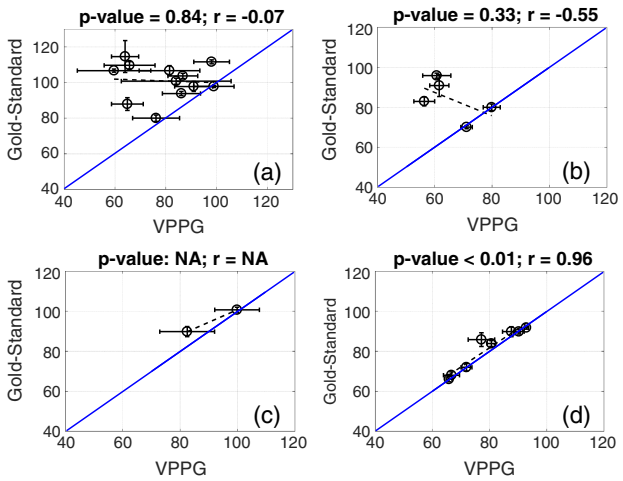

Fig. 7. Correlations between the gold-standard HR values (y axis) and the VPPG HR values (x axis) grouped using tolerance of variance in ROI position and VPPG HR as confidence limits. The confidence limits were arranged in increasing order from (a)–(d), where the confidence level of the predicted VPPG HR was lowest in (a) and highest in (d) (see text).

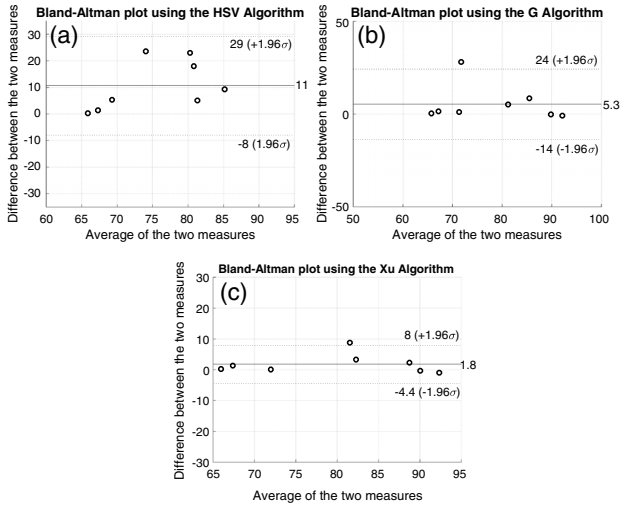

Fig. 8. Bland–Altman plots comparing different algorithms to produce a VPPG time signal for the high-confidence group videos. VPPG HR in (a) was constructed using the Hue channel, while (b) used the Green channel and (c) used the Xu algorithm. All units are shown in BPM.

The fourth group of videos had the ROI variance and VPPG HR variance lower than their respective thresholds. For our analysis here, we selected an ROI variance threshold of 20 pixels and a VPPG HR variance threshold of 10 BPM. The ROI variance was computed by calculating the fraction of area occupied by the smallest square required to enclose the participant's face to the total area of the image and then multiplying this fraction with the length of the face-bounding square. The variance in the HR was selected by setting a tolerance level of ±5 BPM (for a total variance of 10 BPM).

Figure 7 shows the correlations between the gold-standard HR and the VPPG derived HR for these four groups [Fig. 7(a): Group1; Fig. 7(b): Group2; Fig. 7(c): Group3; and Fig. 7(d): Group4]. As in Figs. 5 and 6, the VPPG time signals were derived using the Xu algorithm [6], bandpass filtered, and HR values extracted using the lock-in method. Videos that were selected into Group4 showed very high positive correlations with the gold-standard values. The VPPG HR in groups 1 and 2 were clearly not well correlated to the gold-standard values. Since there were only two videos that fell into Group 3, there were no correlation coefficients computed for this group. As in Fig. 6, dashed black lines show the line of regression, while the solid blue line shows the $y = x$ line. Table 1 also lists the grouping for each video stream.

D. VPPG HR with Different Algorithms

We next compared the impact of using different algorithms to construct the VPPG time signal for extracting the HR data. We only focused on testing three different algorithms described above, on the video streams from Group 4 [high-confidence group shown in Fig. 7(d)]. All algorithms used a sliding window length of 30 s for analysis across 120 s of each video stream, and the derived temporal VPPG data was bandpass filtered and then analyzed using the lock-in method. Figure 8 shows the results of these analyses as Bland–Altman plots. Figure 8(a) shows data for the Hue channel, Fig. 8(b) for the Green channel, and Fig. 8(c) for the Xu algorithm. It is evident that our lock-in technique worked best for the Xu ratiometric algorithm [6]. The Green channel VPPG HR was closer to the gold-standard measured values than the VPPG HR from the Hue channel. Mean error in the derived HR from the Green channel was 5.3 BPM across all subjects, while it was 11 BPM when derived from the Hue channel. Additionally, the correlation between the VPPG HR and the gold standards was significant for the Green channel (with correlation coefficient $r = 0.6$, data not shown) and the Xu algorithm. Correlations were absent for the VPPG HR derived from the Hue channel and the gold standard.

E. VPPG HR with Varying Sliding Window Sizes

Lastly, we compare the impact of using different sliding window lengths on the derived VPPG HR. We processed these data using the Xu algorithm [6] to create the VPPG time signal, which was bandpass filtered and analyzed using the lock-in technique to extract the VPPG HR. We investigated sliding window lengths of 10 s, 20 s (in addition to the 30 s length used earlier) across the full video duration of 120 s. Figure 9

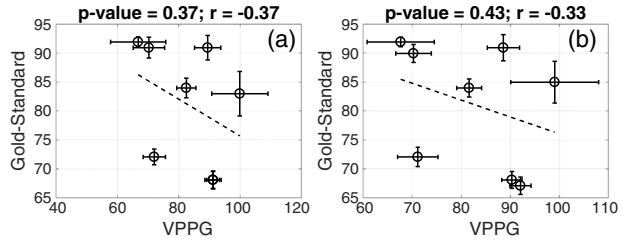

Fig. 9. Correlations between the gold-standard HR values (y axis) and the VPPG HR values (x axis) for sliding window lengths of (a) 10 s and (b) 20 s for all the same videos shown in Fig. 7(d). Black dashed line shows line of regression.

shows the results of these analyses as correlation plots comparing the median gold-standard HR values against the VPPG HR values. Figures 9(a) and 9(b) show these data for the 10 s and 20 s window lengths, respectively.

It is apparent from these data [relative to Fig. 7(d)] that the 30 s sliding window yielded substantially better results than both the 20 s and 10 s window. Additionally, the variance in derived VPPG HR were also dependent on the length of the sliding windows used. We infer that our technique for extracting VPPG HR was accurate only when at least 30 s of video data (sampled at 30 fps) was used to construct the VPPG time signal.

4. DISCUSSION

VPPG presents an attractive means for nonobtrusively and continually monitoring heart rates in humans via standard color video data recorded using commercial cameras. However, as has been previously discussed, this simplicity in data acquisition is concomitant with several sources of noise artifacts that need to be appropriately processed to accurately obtain the HR [17]. Many earlier studies have described a variety of processing techniques to derive clear pulsatile (plethysmographic) signals from the video data [9,12,30,35,36,41]. Additionally, it is also well known that VPPG algorithms necessarily need to deal with finding the correct HR frequency in the presence of high noise levels [15,16]. Our primary objective was to explore the use of a physically motivated (lock-in) simple technique that could be adapted to previously described VPPG algorithms as a "plug-in" for improving their performance. Furthermore, we also present a strategy to create a metric reflecting the confidence levels in the reliability of the extracted VPPG HR, directly from the recorded video data itself.

We applied our lock-in technique to obtain VPPG-based HR from 16 human participants across 27 independently recorded video streams and tested its performance on three previously published VPPG algorithms. As noted in Table 1, the lock-in technique mostly yielded HR values that were closer to the gold-standard values in comparison to those obtained using the peak power spectrum frequency of the VPPG signal. Of the three VPPG algorithms tested, the one based on the modified Beer–Lambert law reported by Xu *et al.* [6] proved the most robust. We hypothesize that the ratiometric form of this algorithm effectively eliminated noisy fluctuations in the VPPG signal and therefore yielded a time trace containing the pulsatile frequency along with higher harmonics, which made the lock-in technique function more effectively.

We also described a procedure for assigning a confidence metric to a video stream, reflecting the likelihood that a VPPG HR extracted from the video recording accurately represents the true HR. These confidence intervals were generated using the variance in the motion-tracked ROI and the VPPG HR (using a sliding window method) for a given segment of a video stream. If these two values were lower than threshold values, the video streams tended to produce accurate results (relative to the gold-standard measurements). Alternately, if these threshold criteria were not satisfied, the VPPG derived HR values were not guaranteed to be accurate.

The idea of exploiting the presence of harmonics of a fundamental HR frequency for VPPG analysis is not unique to our work and has been observed and used previously [4,42,43]. Studies have reported that filtering of the VPPG time-domain signals by construction of adaptive bandpass or lock-in filters, designed using harmonics of the fundamental HR frequency, improved the accuracy of extracted HR signals [42,44–46]. However, the design of these filters first required a robust estimate of the frequency band of the fundamental HR (which was usually estimated from the peak of the power spectrum) or by using a gold-standard measurement. As shown in Fig. 3, the peak of the power spectrum might not occur at the fundamental HR frequency, for noisy VPPG signals. The lock-in technique presented here, on the other hand, attempts to identify the fundamental pulsatile frequency without any previous knowledge of the true HR frequency.

Finally, the use of the sliding-window method to extract HR information across time has also been presented as time frequency or wavelet-based methods by previous studies. However, to the best of our knowledge, a technique to establish a confidence interval metric on a per-video basis has not been reported earlier. The ability to assess the performance of a particular VPPG algorithm on a given video recording adds novelty to this work.

Overall, there still are several factors that must be explored thoroughly before VPPG can routinely be used to accurately measure HR in humans. Particularly, it would be ideal to have VPPG HR estimates to be available in durations far shorter than 30s, as reported by others [7,47]. However, given the low frequency of resting HR (the resting heart beats only slightly more frequently than once per second), a reliable HR using simplistic VPPG algorithms as utilized in this here may be difficult to achieve. Further, our analysis of facial regions focused on using the forehead was predicated on earlier studies that established this region as ideal for VPPG and remains to be explored more thoroughly [40,48]. The potential of VPPG techniques to routinely provide the HR and if possible the HR variability in human behavioral studies holds great promise and potential, but VPPG artifacts such as natural subject movement including rotations of the face and the impact of ballistocardiographic signal contamination need to be comprehensively addressed.

Acknowledgment. We would like to acknowledge support from Miami University's Office of Research for Undergraduates. We would also like to thank Dr Joseph Johnson, Mitchell Dandignac, and undergraduates in the first-year research experience program (Ann Haralson, Taylor Hurt, Linh Le, and David Noviski) for helping with study design and acquisition of the data.

REFERENCES

1. J. Allen, "Photoplethysmography and its application in clinical physiological measurement," Physiol. Meas. **28**, R1–R39 (2007).
2. W. J. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," IEEE Trans. Biomed. Eng. **64**, 1479–1491 (2017).
3. Y. Sun and N. Thakor, "Photoplethysmography revisited: from contact to noncontact, from point to imaging," IEEE Trans. Biomed. Eng. **63**, 463–477 (2016).
4. W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," Opt. Express **16**, 21434–21445 (2008).

5. M. Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," Opt. Express **18**, 10762–10774 (2010).
6. S. C. Xu, L. Y. Sun, and G. K. Rohde, "Robust efficient estimation of heart rate pulse from video," Biomed. Opt. Express **5**, 1124–1135 (2014).
7. Y. P. Yu, P. Raveendran, and C. L. Lim, "Dynamic heart rate measurements from video sequences," Biomed. Opt. Express **6**, 2466–2480 (2015).
8. P. S. Addison, D. Jacquel, D. M. H. Foo, A. Antunes, and U. R. Borg, "Video-based physiologic monitoring during an acute hypoxic challenge: heart rate, respiratory rate, and oxygen saturation," Anesth. Analg. **125**, 860–873 (2017).
9. M. van Gastel, S. Stuijk, and G. de Haan, "Motion robust remote-PPG in infrared," IEEE Trans. Biomed. Eng. **62**, 1425–1433 (2015).
10. B. D. Holton, K. Mannapperuma, P. J. Lesniewski, and J. C. Thomas, "Signal recovery in imaging photoplethysmography," Physiol. Meas. **34**, 1499–1511 (2013).
11. Y. Sun, S. J. Hu, V. Azorin-Peris, R. Kalawsky, and S. Greenwaldc, "Noncontact imaging photoplethysmography to effectively access pulse rate variability," J. Biomed. Opt. **18**, 061205 (2013).
12. A. A. Kamshilin, I. S. Sidorov, L. Babayan, M. A. Volynsky, R. Giniatullin, and O. V. Mamontov, "Accurate measurement of the pulse wave delay with imaging photoplethysmography," Biomed. Opt. Express **7**, 5138–5147 (2016).
13. L. A. M. Aarts, V. Jeanne, J. P. Cleary, C. Lieber, J. S. Nelson, S. B. Oetomo, and W. Verkruysse, "Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—a pilot study," Early Hum. Dev. **89**, 943–948 (2013).
14. S. Rasche, A. Trumpp, T. Waldow, F. Gaetjen, K. Plotze, D. Wedekind, M. Schmidt, H. Malberg, K. Matschke, and S. Zaunseder, "Camera-based photoplethysmography in critical care patients," Clin. Hemorheol. Microcirc. **64**, 77–90 (2016).
15. J. Kranjec, S. Begus, G. Gersak, and J. Drnovsek, "Non-contact heart rate and heart rate variability measurements: a review," Biomed. Signal Process. Control **13**, 102–112 (2014).
16. A. Al-Naji, K. Gibson, S. H. Lee, and J. Chahl, "Monitoring of cardiorespiratory signal: principles of remote measurements and review of methods," IEEE Access **5**, 15776–15790 (2017).
17. M. A. Hassan, A. S. Malik, D. Fofi, N. Saad, B. Karasfi, Y. S. Ali, and F. Meriaudeau, "Heart rate estimation using facial video: a review," Biomed. Signal Process. Control **38**, 346–360 (2017).
18. E. Jonathan and M. J. Leahy, "Cellular phone-based photoplethysmographic imaging," J. Biophoton. **4**, 293–296 (2011).
19. R. Y. Huang and L. R. Dung, "Measurement of heart rate variability using off-the-shelf smart phones," Biomed. Eng. Online **15**, 11 (2016).
20. R. C. Peng, W. R. Yan, N. L. Zhang, W. H. Lin, X. L. Zhou, and Y. T. Zhang, "Investigation of five algorithms for selection of the optimal region of interest in smartphone photoplethysmography," J. Sens. **2016**, 6830152 (2016).
21. T. Coppetti, A. Brauchlin, S. Muggler, A. Attinger-Toller, C. Templin, F. Schonrath, J. Hellermann, T. F. Luscher, P. Biaggi, and C. A. Wyss, "Accuracy of smartphone apps for heart rate measurement," Eur. J. Prev. Cardiol. **24**, 1287–1293 (2017).
22. M. Z. Poh and Y. C. Poh, "Validation of a standalone smartphone application for measuring heart rate using imaging photoplethysmography," Telemed. e-Health **23**, 678–683 (2017).
23. N. Blanik, K. Heimann, C. Pereira, M. Paul, V. Blazek, B. Venema, T. Orlikowsky, and S. Leonhardt, "Remote vital parameter monitoring in neonatology—robust, unobtrusive heart rate detection in a realistic clinical scenario," Biomedizinische Technik./Biomed. Eng. **61**, 631–643 (2016).
24. J. Kuo, S. Koppel, J. L. Charlton, and C. M. Rudin-Brown, "Evaluation of a video-based measure of driver heart rate," J. Saf. Res. **54**, 55. e29–59 (2015).
25. A. Dingli and A. Giordimaina, "Webcam-based detection of emotional states," Vis. Comput. **33**, 459–469 (2017).
26. G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. G. Simos, K. Marias, and M. Tsiknakis, "Stress and anxiety detection using facial cues from videos," Biomed. Signal Process. Control **31**, 89–101 (2017).
27. W. J. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Robust heart rate from fitness videos," Physiol. Meas. **38**, 1023–1044 (2017).
28. J. Moreno, J. Ramos-Castro, J. Movellan, E. Parrado, G. Rodas, and L. Capdevila, "Facial video-based photoplethysmography to detect HRV at rest," Int. J. Sports Med. **36**, 474–480 (2015).
29. D. J. McDuff, J. R. Estepp, A. M. Piasecki, and E. B. Blackford, "A survey of remote optical photoplethysmographic imaging methods," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2015), pp. 6398–6404.
30. M. Kumar, A. Veeraraghavan, and A. Sabharwal, "Distance PPG: robust non-contact vital signs monitoring using a camera," Biomed. Opt. Express **6**, 1565–1588 (2015).
31. F. Bousefsaf, C. Maaoui, and A. Pruski, "Automatic selection of webcam photoplethysmographic pixels based on lightness criteria," J. Med. Biol. Eng. **37**, 374–385 (2017).
32. A. Garcia-Agundez, T. Dutz, and S. Goebel, "Adapting smartphone-based photoplethysmograpy to suboptimal scenarios," Physiol. Meas. **38**, 219–232 (2017).
33. H. Qi, Z. Y. Guo, X. Chen, Z. Q. Shen, and Z. J. Wang, "Video-based human heart rate measurement using joint blind source separation," Biomed. Signal Process. Control **31**, 309–320 (2017).
34. D. Wedekind, A. Trumpp, F. Gaetjen, S. Rasche, K. Matschke, H. Malberg, and S. Zaunseder, "Assessment of blind source separation techniques for video-based cardiac pulse extraction," J. Biomed. Opt. **22**, 035002 (2017).
35. F. Bousefsaf, C. Maaoui, and A. Pruski, "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate," Biomed. Signal Process. Control **8**, 568–574 (2013).
36. W. J. Wang, S. Stuijk, and G. de Haan, "Unsupervised subject detection via remote PPG," IEEE Trans. Biomed. Eng. **62**, 2629–2637 (2015).
37. A. Eaton, V. Vincely, P. Kurt, K. Hugenberg, K. Vishwanath, D. Levitz, A. Ozcan, and D. Erickson, "The reliability and accuracy of estimating heart-rates from RGB video recorded on a consumer grade camera," Proc. SPIE **10055**, 100550Y (2017).
38. U. Rubins, V. Upmalis, O. Rubenis, D. Jakovels, and J. Spigulis, "Real-time photoplethysmography imaging system," in *15th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics*, K. Dremstrup, S. Rees, and O. Jensen, eds. (2011), pp. 183–186.
39. G. R. Tsouri and Z. Li, "On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras," J. Biomed. Opt. **20**, 048002 (2015).
40. M. Abul Hassan, A. S. Malik, N. Saad, B. Karasfi, Y. S. Ali, and D. Fofi, "Optimal source selection for image photoplethysmography," in *IEEE International Instrumentation and Measurement Technology Conference Proceedings* (2016), pp. 455–459.
41. W. J. Wang, S. Stuijk, and G. de Haan, "A novel algorithm for remote photoplethysmography: spatial subspace rotation," IEEE Trans. Biomed. Eng. **63**, 1974–1984 (2016).
42. L. T. Feng, L. M. Po, X. Y. Xu, Y. M. Li, and R. Y. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," IEEE Trans. Circuits Syst. Video Technol. **25**, 879–891 (2015).
43. A. V. Moco, S. Stuijk, and G. de Haan, "Motion robust PPG-imaging through color channel mapping," Biomed. Opt. Express **7**, 1737–1754 (2016).
44. J. P. Lomaliza and H. Park, "A highly efficient and reliable heart rate monitoring system using smartphone cameras," Multimedia Tools Appl. **76**, 21051–21071 (2017).
45. L. Yang, M. Liu, L. Q. Dong, Y. J. Zhao, and X. H. Liu, "Motion-compensated non-contact detection of heart rate," Opt. Commun. **357**, 161–168 (2015).
46. A. A. Kamshilin, S. Miridonov, V. Teplov, R. Saarenheimo, and E. Nippolainen, "Photoplethysmographic imaging of high spatial resolution," Biomed. Opt. Express **2**, 996–1006 (2011).
47. Y. P. Yu, P. Raveendran, C. L. Lim, and B. H. Kwan, "Dynamic heart rate estimation using principal component analysis," Biomed. Opt. Express **6**, 4610–4618 (2015).
48. S. Kwon, J. Kim, D. Lee, and K. Park, "ROI analysis for remote photoplethysmography on facial video," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2015), pp. 4938–4941.