DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-159

Improving Video-Based Heart Rate Estimation

Dahjung Chung [1], *Jeehyun Choe* [1], *Marguerite E. O'Haire* [2], *A. J. Schwichtenberg* [3], *and Edward J. Delp* [1]
[1] *Video and Image Processing Laboratory (VIPER), School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA*
[2] *Center for the Human-Animal Bond, Department of Comparative Pathobiology, College of Veterinary Medicine, Purdue University, West Lafayette, Indiana, USA*
[3] *Department of Human Development and Family Studies, Purdue University, West Lafayette, Indiana, USA*

Abstract

Over the past 5 years several video-based heart rate (HR) estimation methods have been developed. These non-contact methods of HR estimation use video processing techniques to estimate the HR of humans in the scene. This is known as videoplethysmography (VHR) and has applications to the medical and surveillance fields. In this paper, we review two previous VHR techniques and describe techniques to improve VHR accuracy. These include: (1) targeted skin detection within the facial region, (2) recursive temporal difference filtering and small variation amplification, (3) periodic signal detection within the expected human HR range while considering background periodic signals, and (4) reduction of signal range using a cutoff frequency search. These improvements increased our HR estimate accuracy in two conditions (no-motion, non-random motion) when compared to earlier VHR methods but were not significantly better than those that employ an adaptive frequency analysis.

Introduction

Remote health monitoring is a growing field and non-contact video-based heart rate (HR) estimates are now possible. Video-based HR estimation, known as videoplethysmography (VHR), is a technique that can detect blood volume changes in the microvascular tissue [1] and allows HR to be estimated from a video of person in the scene. Similar to the process used in medical grade pulse oximeters, VHR assesses small blood volume changes in cheek/face capillaries from the video sequence. Although pulse oximeters are accurate and inexpensive, they are not well tolerated by patients over long periods of monitoring or by patients with tactile sensitivities. Therefore, a non-contact means of HR monitoring could be a valuable addition to the medical field. Additionally, non-contact monitoring of HR could also inform surveillance systems and provide an alert when someone's heart rate is too high or low.

Several VHR methods have been developed over the past 5 years with each based on two basic assumptions. First, small color variations in the cheek/face region reflect blood volume changes (i.e., heart-beats). This is sometimes known as "micro-blushing." Second, given the rhythmic nature of heart-beats, the color variations will also follow an oscillatory pattern. Beyond these assumptions, each technique uses a slightly different approach in detecting the HR signal from video. For example, in one of the earliest VHR methods, Poh (and Picard) *et al.* [2], [3] obtained the mean color pixel intensity for the facial region in each RGB frame to form 1D signals and then used Independent Component Analysis (ICA) [4] to separate the HR signal from the other noise signals. Monkaresi *et al.* [5] improved this method by choosing the ICA components using K-Nearest Neighbor classification. McDuff *et al.* [6] also extended Poh *et al.* [3] by using a five band digital camera. Additionally, we [7] improved the Poh *et al.* [3] method by computing the 1D signal from each RGB frame and using an adaptive cutoff frequency for the bandpass filter (AFR) to achieve a more stable HR estimate. For the remainder of this paper, the Poh *et al.* [3] method will be referred to as the Picard method (or just "Picard") and the our previous approach [7] will be referred to as the AFR method (or just "AFR"). Both of these methods use ICA to estimate a HR signal (see Figure 1).

Apart from ICA, Kumar *et al.* [8] used only the green channel signal to estimate HR. They also combined signals from different tracked facial regions using a weighted average. Yan *et al.* [9] used Red, Green, Blue signal weight averages instead of ICA (and weights for each signal were determined by maximizing the signal noise ratio). Haan *et al.* [10], [11] proposed a chrominance-based VHR method using the ratio of two normalized color signals. Other approaches have proposed using more subtle changes to estimate HR. For example, Balakrishnan *et al.* [12] estimated HR by detecting subtle head motions. Additionally, others have used spatial decomposition and temporal filtering to magnify small video-captured motions to estimate HR [13], [14]. Current challenges facing VHR include diverse skin tones, a wide range of typical human HR, and several noise-related factors (i.e., room lighting, camera dependent noise, subject motion).

To mitigate noise-related factors the present study employed small variation amplification, described in detail later, instead of ICA. Building on the strengths of previous approaches we will only assess the green channel signal. Additionally, since HR is reflected in small color changes in skin region, using small variation amplification allows us to amplify the small color variations and attenuate large variations. To obtain a more stable estimation of the HR, we use a new approach to estimate the cutoff frequencies of the bandpass filters used before the Inter-Beat-Interval (IBI) computation (see Figure 1). In this paper, we will (1) describe in detail two related methods Picard [3] and our AFR [7]) and compare them with our proposed approach, (2) present experimental data on how each method preformed in no-motion and non-random motion conditions, (3) empirically test the improvements generated from the proposed method (i.e., is the proposed method significantly better than previous methods).

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-159

The Picard and AFR Methods

Picard [3] used ICA to decompose 1D signals from the video and then selected the signal component with the highest power to estimate HR. We have previously extended the Picard method using an Adaptive Frequent Range (AFR) filter that uses ICA [7]. We will describe the Picard method [3] and our AFR technique [7] in detail for future comparison. In Figure 1, the white blocks depict Picard method [3] and the gray blocks illustrate extensions/adaptations used in the our AFR method [7].

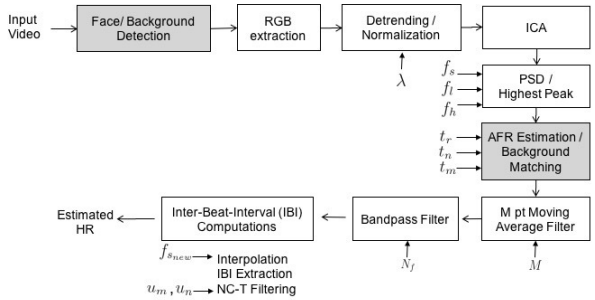


Figure 1: The block diagram of the Picard method [3] and our previously published AFR method [7].

Picard begins with face detection. The average average of the pixel intensity of the detected face region is obtained in each RGB frame to form three 1D signals. The 1D signals are detrended using a high-pass like filter based on a smoothness priors [15]. The parameter for the detrending filter sets the high pass cutoff frequency. We denote this parameter as λ. We use $\lambda = 100$ which corresponds to $0.021 f_s$ where f_s is the sampling rate (in our case, $f_s = 30$ Hz which is the frame rate of our video sequence). The detrended signal is normalized with z-score normalization to form a zero-mean and unit variance signal. ICA is used on the three 1D normalized signals to separate the HR signal from the other noise signals. The Power Spectrum Density (PSD) is obtained for the three ICA components and used to choose the HR signal. The HR signal is the ICA component which has the highest peak in PSD within the resting HR range (0.7 to 2 Hz). After a 5 point moving average filter ($M = 5$), the signal is bandpass filtered using a 128-point Hamming window (filter order $N_f = 127$) with fixed cutoff frequency. The Low cutoff frequency (f_l) and high cutoff frequency (f_h) are 0.7 and 2 Hz respectively. Next, the bandpass signal is interpolated to a higher sampling frequency $f_{s_{new}} = 256$ Hz using cubic spline interpolation. In the last block of Figure 1, the Inter-Beat-Interval (IBI), is the time interval between two peaks in units of seconds, is determined to estimate the HR. The IBI detects peaks and finds the interval between them to estimate the HR. Finally, the IBI signal is filtered using the NC-T filter [16] in the IBI block with fixed parameters $u_n = 0.4$ and $u_m = 1.0$ Hz. This filter can remove the unstable HR estimation by filtering the rapidly changing values.

Our previous method, AFR, used an adaptive cutoff frequency range instead of a fixed cutoff frequency range for the bandpass filter [7]. AFR selects frequencies by targeting those within the typical range of human heart rate and by ignoring oscillatory signals that may be present in the background signal (i.e., lighting, camera vibration). AFR begins with face and background region detection. Two sets of 1D RGB signals from both regions are detrended and normalized. The ICA and PSD processes are the same as the Picard method. Adaptive cutoff frequencies are determined for the bandpass filter in two steps. First, AFR forms frequency clusters in each PSD from both face and background regions which are the candidates for the cutoff frequencies for bandpass filter. A frequency cluster is a range of neighboring frequencies that are determined by thresholding the PSD (thresholds $t_r = 0.1$, $t_n = 0.1$ are determined empirically). Second, background removal was done by comparing frequency clusters from the face and background regions using the Sum of Absolute Differences (SAD) in the clusters. If the SAD between two clusters is less than a threshold ($t_m = 0.4$), two clusters are classified as a similar pair. The threshold was empirically determined [7]. Starting from the highest energy cluster, AFR selects the face region cluster that does not match with the background cluster. The new cutoff frequencies for the bandpass filter are derived from the selected cluster's range. Once the signal is bandpass filtered, the IBI computation of Picard is done.

Proposed Method

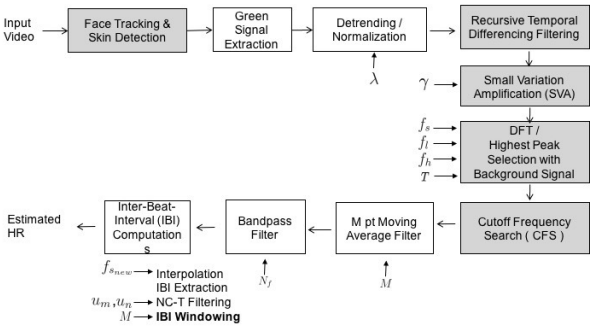


Figure 2: The block diagram of the proposed method.

Figure 2 shows the block diagram of the proposed method. Gray colored blocks are the extensions of the Picard [3] and AFR [7] methods.

In the first block we track the facial region of the subject (to allow for HR estimates in both motion and no-motion conditions). This step is distinct and is not used in the Picard and AFR methods. Once the facial region is identified, then skin pixels are detected within the facial region. Next, only the green channel signal is used (Picard and AFR used 3 1D RGB signals). The reason why we use only the green channel image is that it has the most information relevant to HR. Kumar *et al.* [8] discussed that hemoglobin (Hb), which is related to blood volume change and eventually to HR, has a high absorption coefficient in the green wavelength. The average green pixel intensity signal is obtained in each frame for the skin region to form a 1D signal. As in AFR, we use the background (non-skin) signal but compare the highest peaks of the background and skin signals instead of clustering. Detrending and normalization is the same as Picard and AFR. Next temporal differencing filter and small variation amplification (SVA) is used instead of ICA. The idea here is the amplification of the small color variations and the attenuation of the large variations in the skin regions to estimate HR. PSDs are obtained next for the amplified green signal from the skin and background regions. The highest peak is determined by comparing the the highest peak frequencies from skin and background regions. Equation for similarity is provided in the following subsection. The highest peak is then used in the Cutoff Frequency Search (CFS) to find the

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-159

cutoff frequencies for the bandpass filter. After a M-point moving average filter ($M = 5$), the signal is bandpass filtered with the cutoff frequencies from CFS. Our method's IBI is almost same as the IBI of Picard and AFR. The only difference is the use of IBI windowing (boldface in Figure 2). IBI windowing uses a M point moving average filter ($M = 5$) on the NC-T filter output in order to remove the outliers. We will describe in detail Tracking/Skin Detection, Temporal Filter/Amplification and Peak Selection/CFS in the following sections.

Face Tracking and Skin Detection

The initial bounding box for tracking is obtained by face detection using a Haar Cascade Classifier [17]. For tracking, we derived a reference color model from the initial bounding box of the face region. For the color model, each RGB color space is quantized from the original 256 bins to 16 bins and is mapped into 1D 16^3-bin histogram. The sum of this histogram is then normalized to one. Particle filter tracking is used to find the corresponding face region in each frame [18]. Denoting the hidden state and the data at time t by x_t and y_t respectively, the probabilistic model we use for tracking is

$$p(x_{t+1}|y_{0:t+1}) \propto p(y_{t+1}|x_{t+1}) \int_{x_t} p(x_{t+1}|x_t) p(x_t|y_{0:t}) dx_t \quad (1)$$

where $p(y_{t+1}|x_{t+1})$ is the likelihood model of the data, and $p(x_{t+1}|x_t)$ is the transition model of the second-order autoregressive dynamics [18]. We define the state at time t as the location in the 2D image represented as pixel coordinates. For obtaining the likelihood $p(y_t|x_t)$, we use the distance metric $d(y) = \sqrt{1-\rho(y)}$ where $\rho(y)$ is the sample estimate of the Bhattacharyya coefficient between the reference color model and the candidate color model of each particle at position y [19].

Given that our target signal includes only small color variations, even small levels of noise may significantly impact our HR estimates. Therefore, in an attempt to minimize noise, skin detection is use to remove hair and eye regions (which do not contain HR information). A bayesian classifier using non-parametric density estimation is used for skin detection [20].

Recursive Temporal Differencing Filter and Small Variation Amplification (SVA)

The basic idea is that we only amplify the small changes in time and suppress the large changes, because the HR signal is the small color changes in the skin region caused by cardiac activity. To achieve, a first order temporal recursive differencing filter is used on the detrended green channel signal:

$$g[n] = g[n+1] - g[n] \quad (2)$$

where $g[n]$ is the detrended green signal from skin pixels and n is the frame index.

Small variation amplification (SVA) is then used (Equations 3 and 4):

$$g_{amp}[n] = \alpha g[n] \quad (3)$$

$$\alpha = |g[n]|^{\gamma} \quad (4)$$

where α is the amplification factor based on the difference value of green signal. We choose $\gamma = -0.1$ empirically. From these blocks, we can suppress the large temporal variations in the signal and amplify the HR signal reflected in small temporal changes.

Peak Selection and Color Frequency Search (CFS)

Peak Selection begins with finding the highest peak in the PSD for the skin and background regions. If the highest peak from the skin region is similar to the highest peak from the background region, then this is a strong periodic noise signal from factors such as lighting. The similarity between highest peaks is determined by:

$$d_f = |f_1 - f_2| \quad (5)$$

where f_1 is the highest peak frequency from the skin region and f_2 is the highest peak frequency from the background region. If $d_f < T$ (we empirically choose threshold $T = 0.1$), we then find the next highest peak in the skin PSD. The selected highest peak is used as the starting point for the CFS which eventually determines tighter cutoff frequencies for the Bandpass Filter (BPF).

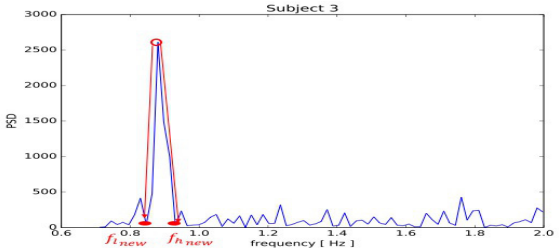


Figure 3: An example of CFS in PSD.

As shown in Figure 3, a gradient search is done on the skin region PSD signal (starting from the highest peak). The points that have a sign change are determined as the new cutoff frequencies ($f_{l_{new}}$ and $f_{h_{new}}$) for BPF. Tighter cutoff frequencies were achieved using CFS for BPF and this supports more stable estimations in the IBI computation.

Experimental Results

In our experiments, we acquired two different datasets. Both were collected in the same room which had windows with semi-transparent blinds and lighting on the ceiling as shown in Figure 4. Camera 1 in Figure 4 was used to record the subject and camera 2 was used to record the pulse oximeter for the ground truth HR. All videos had a spatial resolution of 1920×1080, 30 fps and 60 seconds length.

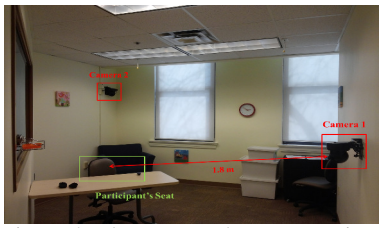
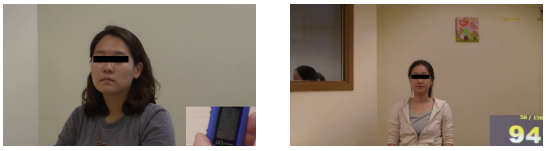


Figure 4: The room and camera setting.

The ground truth HR was measured using a Pulse Oximeter attached to the finger of the subject. A Nonin GO_2 Achieve Finger Pulse Oximeter was used in Dataset 1 and CE & FDA Approved Handheld Pulse Oximeter was used in Dataset 2. The pulse oximeter HRs were manually recorded from the video once per second in both datasets. Dataset 1 included 22 subjects (12 females and 10 males) and Dataset 2 included 18 subjects (9 females and 9 males).

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-159

The distance between the subject and the camera was approximately 1.8 m in both datasets, the zoom was manually adjusted to focus only on the upper torso and face in Dataset 1. In Dataset 2, the zoom was manually adjusted to show entire upper body of the subject. Examples of videos from Dataset 1 and 2 are provided in Figure 5.



(a) Dataset1 Video Setting. (b) Dataset2 Video Setting.

Figure 5: Video Setting Examples.

Dataset 1 included no-motion videos and Dataset 2 included both non-random motion and no-motion videos. In the no-motion videos, the subjects were seated and were asked to look toward the camera. In the non-random motion videos, the subjects were asked to move their head from left to right repeatedly while keeping their faces toward the camera.

In our experiments, we set the parameters for Picard [3] and AFR [7] as described in the Picard and AFR Methods Section. A summary of the parameters used in the three methods are provided in Table 1. All the parameters are chosen empirically except the video frame rate f_s.

Parameter	Picard [3]	AFR [7]	Proposed
λ	100		
f_s	30		
(f_l, f_h)	(0.7,2)		
$f_{s_{new}}$	256		
M	5		
N_f	127		
(u_n, u_m)	(0.4,1)		
t_r	-	0.1	-
t_n	-	0.1	-
t_m	-	0.4	-
γ	-	-	-0.1
T	-	-	0.1

Table 1: A summary of parameters used in the three VHR methods.

The initial facial region was detected in the first frame of the video [17]. 80% of the detected face region's width and height were used with all three methods. Skin detection was trained from the SFA image database [21]. The background region is selected on the wall behind the subject (with the size of 80% of face region).

The proposed method shows better overall performance, compared to the Picard [3] and AFR [7]. For example, Figure 6 shows one of the final HR estimation from all three methods for the motion video of Subject 9 in Dataset 2. The red line shows manually annotated ground truth HR from the pulse oximeter. The estimated HR from Picard, AFR and the proposed method are labeled as black, green, and blue respectively. Our proposed method shows the closest estimation to true HR in this case.

To evaluate the overall performance, we defined the error metrics as:

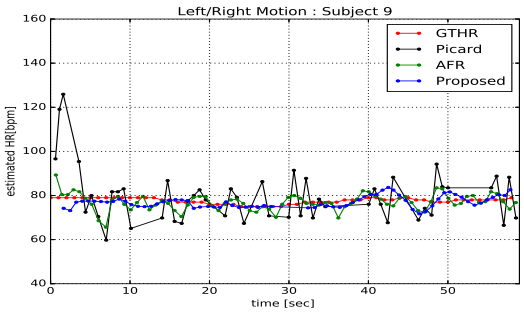


Figure 6: A comparison of the three VHR methods - Subject 9 motion in Dataset2 : Estimated HR [bpm] vs Time [sec].

$$e_1 = \begin{cases} \frac{1}{N}\sum_{n=1}^{N} \left| \frac{HR_{est}[n] - HR_{true}[n]}{HR_{true}[n]} \right| * 100 \quad [\%] \\ \\ \frac{1}{N}\sum_{n=1}^{N} |HR_{est}[n] - HR_{true}[n]| \quad [bpm] \end{cases} \tag{6}$$

where $HR_{est}[n]$ is the estimated HR in units of beat per minute (bpm), $HR_{true}[n]$ is the manually recorded HR in units of bpm from the pulse oximeter and n is the time domain index when the HR estimations exist. Comparison results for the three VHR methods in terms of Error Rate e_1 in both units [bpm and %] are shown in Table 2, 3.

In Dataset 1, the proposed method has the lowest average error rate across all 22 subjects (3.67 % and 4.71 bpm) as shown in Table 2a. The proposed method outperforms Picard and AFR in most of the videos except three cases. The overall error rate in Dataset 1 is lower than the one in Dataset 2 because Dataset 1 has more pixels in the face region due to the manual zoom focus on the upper torso and face (compared to the whole upper body in Dataset 2).

In Dataset 2, the proposed method has the lowest average error rate across all 18 no-motion videos (7.09 % and 9.48 bpm) as shown in Table 3a. The proposed method outperforms Picard and AFR in most of the videos except four cases. Even though the proposed method outperforms the previous methods in Dataset 2, the overall error is still higher than Dataset 1 due to the less number of pixels in face region.

In non-random motion videos, all three methods still have a large error rate compared to the no-motion scenario. The proposed method has the lowest average error rate as measured by percent across all 18 subjects (16.89 %), while the Picard method has the lowest average error rate as measured by bpm across all 18 subjects (13.36 bpm) as shown in Table 3b. The proposed method outperforms Picard and AFR in most of the videos except five cases. However, when the highest peak is not matched with the true HR, the proposed method has the tendency to propagate the error because of CFS.

Statistical Analyses

We used a series of the paired samples t-tests [22] to compare the proposed method to the other two methods. The outcome variable was the percentage error rate $e_1[\%]$ for no motion and non-random motion videos. Two-tailed tests were used with an

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-159

ID	Picard [3]		AFR [7]		Proposed	
	[bpm]	[%]	[bpm]	[%]	[bpm]	[%]
1	5.53	8.00	1.49	2.16	1.58	2.29
2	3.58	6.32	2.97	5.27	1.81	3.21
3	5.50	10.44	1.89	3.59	1.19	2.27
4	6.47	7.23	2.75	3.08	2.44	2.73
5	11.29	21.21	2.69	5.03	2.26	4.21
6	12.37	13.67	4.47	4.92	3.15	3.46
7	8.36	10.66	2.89	3.69	2.36	3.01
8	7.12	10.00	2.95	4.12	1.93	2.69
9	5.95	7.91	2.39	3.17	2.10	2.78
10	6.57	6.99	2.85	3.03	15.15	16.05
11	5.03	6.37	1.43	1.82	1.91	2.44
12	5.34	9.82	4.28	7.82	2.47	4.49
13	16.49	22.11	16.25	21.16	14.02	18.17
14	5.77	7.26	2.88	3.63	1.84	2.32
15	9.29	12.96	5.05	6.88	4.15	5.65
16	8.56	12.25	6.58	9.35	3.03	4.33
17	4.38	6.55	2.76	4.12	2.29	3.44
18	4.79	6.32	1.73	2.29	1.23	1.62
19	8.00	8.80	2.94	3.21	1.54	1.68
20	6.24	7.52	2.58	3.11	1.96	2.37
21	9.87	11.24	6.19	7.04	8.84	9.74
22	7.98	10.33	4.00	5.17	3.57	4.63
Avg.	7.48	10.18	3.82	5.17	**3.67**	**4.71**

(a) No-motion videos.

Table 2: A comparison of three VHR methods in Dataset 1 : Error Rate e_1 [bpm , %]

alpha level of 0.05. Descriptive statistics, including mean error rate (M) and standard deviation (SD) are presented in Table 4.

For the no-motion videos (N = 40, Table 4a), the proposed method had a lower mean error rate (M) than both other methods, but was only significantly lower than the Picard method ($p < 0.05$), not AFR ($p = 0.37$). The AFR error was also significantly lower than Picard ($p < 0.05$). Thus both the proposed and AFR methods outperformed the Picard method. For the non-random motion videos (N = 18, 4b), the same pattern of results emerged; however, method differences were not statistically significantly different.

Conclusion

We presented a new VHR method, which includes advances such as small amplification variation instead of ICA and a new cutoff frequency search method for BPF. We conducted empirical testing to evaluate the difference between our proposed method and two prior methods, the Picard method and our previous AFR method. Our proposed method had the lowest average error rate and the Picard method had the highest average error rate. All three methods were less accurate in the motion conditions, compared to no-motion conditions, highlighting the need for further advances in motion-sensitive VHR. Future work includes developing methods to correct for the motion artifacts.

ID	Picard [3]		AFR [7]		Proposed	
	[bpm]	[%]	[bpm]	[%]	[bpm]	[%]
1	21.44	32.28	13.06	20.04	6.91	10.51
2	10.90	15.38	7.62	10.58	6.50	9.12
3	10.67	13.10	9.89	12.15	9.58	11.75
4	19.40	26.85	3.86	5.29	36.96	50.81
5	10.11	10.44	5.93	6.09	6.13	6.33
6	15.94	21.34	14.84	19.88	2.49	3.32
7	8.31	10.04	3.40	4.10	3.71	4.46
8	16.39	23.00	10.31	14.30	7.70	10.65
9	11.17	14.35	9.09	11.69	5.85	7.53
10	9.29	13.78	3.83	5.64	1.46	2.17
11	17.41	19.07	28.60	31.39	8.13	8.91
12	12.70	18.55	8.15	11.79	4.24	6.11
13	15.24	16.86	23.02	25.54	3.26	3.60
14	7.46	8.84	5.06	6.01	4.07	4.85
15	7.99	11.08	4.66	6.48	3.02	4.16
16	8.99	11.83	5.94	7.87	6.66	8.86
17	20.05	38.98	6.22	11.16	6.09	10.87
18	6.49	9.14	4.87	6.75	4.77	6.59
Avg.	12.77	17.50	9.35	12.04	**7.09**	**9.48**

(a) No-motion videos

ID	Picard [3]		AFR [7]		Proposed	
	[bpm]	[%]	[bpm]	[%]	[bpm]	[%]
1	10.86	14.38	9.92	13.13	4.49	5.94
2	20.78	29.57	20.54	29.27	6.82	9.68
3	13.22	16.99	11.27	14.87	5.88	7.81
4	14.56	18.86	15.53	19.89	27.74	35.65
5	19.52	20.55	27.39	28.93	34.08	36.01
6	23.98	26.35	31.15	34.23	18.81	20.67
7	16.34	16.71	17.83	18.18	14.62	14.88
8	10.73	15.12	15.00	21.03	11.28	15.79
9	8.96	11.49	2.66	3.42	2.08	2.66
10	9.70	13.32	9.68	13.30	20.85	28.56
11	14.82	15.96	14.82	15.96	21.33	22.95
12	7.79	11.79	4.74	7.19	2.76	4.17
13	13.29	15.38	35.47	40.69	25.20	28.90
14	12.59	15.68	14.63	18.24	26.58	33.08
15	13.87	23.01	11.75	19.36	11.24	18.53
16	8.33	10.35	4.92	6.08	3.57	4.44
17	11.75	19.75	6.44	10.78	2.64	4.42
18	9.39	11.76	8.20	9.83	8.14	9.89
Avg.	**13.36**	17.06	14.55	18.02	13.78	**16.89**

(b) Non-random motion videos

Table 3: A comparison of three VHR methods in Dataset 2 : Error Rate e_1 [bpm , %]

References

[1] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, pp. R1–R39, March 2007.

[2] M. Poh, D. McDuff, and R. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, no. 10, pp. 10 762–10 774, May 2010.

[3] M. Poh, D. McDuff, and R. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, January 2011.

[4] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[5] H. Monkaresi, R. Calvo, and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, pp. 1153–1160, November 2013.

[6] D. McDuff, S. Gontarek, and R. Picard, "Improvements in remote cardio-pulmonary measurement using a five band digital camera,"

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-159

Pair	Methods	Mean	SD	p-value
1	Proposed	6.8545	8.13569	
	AFR	8.2603	6.79649	0.365
2	Proposed	6.8545	8.13569	
	Picard	13.4718	7.35702	< 0.05
3	AFR	8.2603	6.79649	
	Picard	13.4718	7.35702	< 0.05

(a) No-motion videos, Number of Sample : 40

Pair	Methods	Mean	SD	p-value
1	Proposed	16.8906	11.62436	
	AFR	18.0211	9.96929	0.635
2	Proposed	16.8906	11.62436	
	Picard	17.0567	5.22310	0.952
3	AFR	18.0211	9.96929	
	Picard	17.0567	5.22310	0.602

(b) Non-random motion videos, Number of Sample : 18

Table 4: Paired Sample T- Test Result with Error Rate e_1 [%]

IEEE Transactions on Biomedical Engineering, vol. 61, no. 10, pp. 2593 – 2601, October 2014.

[7] J. Choe, D. Chung, A. J. Schwichtenberg, and E. J. Delp, "Improving video-based resting heart rate estimation: A comparison of two methods," *Proceedings of the IEEE 58th International Midwest Symposium on Circuits and Systems*, pp. 1–4, August 2015, Fort Collins, CO.

[8] M. Kumar, A. Veeraraghavan, and A. Sabharval, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomedical Optics Express*, vol. 6, no. 5, pp. 1565–1588, 2015.

[9] Y. Yan, X. Ma, L. Yao, and J. Ouyang, "Noncontact measurement of heart rate using facial video illuminated under natural light and signal weighted analysis," *Bio-Medical Materials and Engineering*, vol. 26, no. s1, pp. 903–909, 2015.

[10] V. J. G. de Haan, "Robust pulse rate from chrominance-based rppg," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.

[11] G. de Haan and A. Leest, "Improved motion robustness of remote-ppg by using the blood volume pulse signature," *Physiological Measurement*, vol. 35, no. 9, pp. 1913–1926, 2014.

[12] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3430–3437, June 2013, Portland, OR.

[13] H. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 65:1–8, July 2010.

[14] N. Wadhwa, M. Rubinstein, F. Durand, and W. Freeman, "Phase-based video motion processing," *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 80:1–9, July 2013.

[15] M. Tarvainen, P. Ranta-aho, and P. Karjalainen, "An advanced detrending method with application to hrv analysis," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 2, pp. 172–175, February 2002.

[16] J. Vila, F. Palacios, J. Presedo, M. Fernandez-Delgado, P. Felix, and S. Barro, "Time-frequency analysis of heart-rate variability," *IEEE Magazine on Engineering in Medicine and Biology*, vol. 16, no. 5, pp. 119–126, September/October 1997.

[17] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. I–511–I–518, December 2001, Kauai, HI.

[18] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," *Proceedings of the 7th European Conference on Computer Vision*, pp. 661–675, May 2002, Copenhagen, Denmark.

[19] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[20] D. Chai, S. Phung, and A. Bouzerdoum, "A bayesian skin/non-skin color classifier using non-parametric density estimation," *Proceedings of the International Symposium on Circuits and Systems*, vol. 2, pp. II–464–II–467, May 2003.

[21] J. Casati, D. Moraes, and E. Rodrigues, "Sfa: A human skin image database based on feret and ar facial images," *Proceedings of the IX workshop de Visao Computational*, June 2013, rio de Janeiro, Brazil.

[22] J. McDonald, *Handbook of Biological Statistics : Paired T-Test*. Baltimore, Maryland: Sparky House Publishing, 2014, vol. 3.

Author Biography

Dahjung Chung is currently the Ph.D students of Electrical and Computer Engineering at Purdue University. She received her BS in Electronic and Information Engineering from Ewha Womans University and MS in Electrical and Electronic Engineering from Yonsei University. Her research interests include image/video processing, computer vision and machine learning.

Jeehyun Choe is currently a PhD student of Electrical and Computer Engineering at Purdue University. She received her BS in Electrical and Electronic Engineering from Yonsei University in Seoul, Korea and MS in Robotics Program from KAIST, Daejeon, Korea. Her research interests include image/video processing and computer vision.

Marguerite O'Haire received her BA in psychology from Vassar College(2008) and her PhD in psychology from The University of Queensland (2014). Since then, she has worked as an Assistant Professor of Human-Animal Interaction in the Center for the Human-Animal Bond, at Purdue University. Her research focuses on the unique biopsychosocial effects of interacting with animals for humans, including individuals with autism spectrum disorder and posttraumatic stress disorder.

A.J. Schwichtenberg is an assistant professor at Purdue University in the Department of Human Development and Family Studies (HDFS). She received her Ph.D. in HDFS from the University of Wisconsin, Madison and completed her postdoctoral training at the M.I.N.D. (Medical Investigation of Neurodevelopmental Disorders) Institute at the University of California, Davis. Dr. Schwichtenberg's research focuses on the role(s) of physiological regulation in early at risk development.

Edward J. Delp was born in Cincinnati, Ohio. He is currently The Charles William Harrison Distinguished Professor and Professor of Biomedical Engineering at Purdue University. His research interests include image and video compression, multimedia security, medical imaging, multimedia systems, communication and information theory.