A ROBUST REAL TIME SYSTEM FOR REMOTE HEART RATE MEASUREMENT VIA CAMERA

Duc Nhan Tran, Hyukzae Lee, and Changick Kim

Korea Advanced Institute of Science and Technology
Department of Electrical Engineering
{nhantd, hyukzaelee, changick}@kaist.ac.kr

ABSTRACT

Heart rate (HR) is an important indicator of human health status. Traditional heart rate measurement methods rely on contact-based sensors or electrodes, which are inconvenient and troublesome for users. Remote sensing of the photoplephysmography (PPG) signal using a video camera provides a promising means to monitor vital signs of people without the need of any physical contact. However, until recently, most of the literature papers approaching this problem have only reported results from off-line recording videos taken under well controlled environments. In this paper, we propose a method to improve HR measurement accuracy under challenging environments involving factors such as subjects movement, complicated facial models (i.e., hair, glass, beards, etc.), subjects' distance to camera, and low illumination condition. We also build a framework for real-time measuring system and construct a stable model for recording and displaying results for long term heart rate monitoring. We tested our system on challenging dataset, and demonstrated that our method not only deals with real-time, on-line measurement tasks, but also outperforms others' works.

Index Terms— PPG, remote heart rate measurement, vital signs monitoring, real-time system.

1. INTRODUCTION

Remote monitoring of vital signs, including HR, via conventional commercial camera has recently drawn significant attention to the research community due to the advantages it offers. The main benefit of this new method is non-invasive and passive monitoring, which brings convenience and comfort to users. Furthermore, with rapid advancement in technology, which enables development of low-cost digital cameras, non-invasive HR measurement using video camera even attracts more attention. Remote HR monitoring has high potential in wide range of applications. First, it can be used for medical circumstances requiring long term non-invasive HR monitoring. Second, several research works have demonstrated strong correlation between HR and one's physiological and physical state. Therefore, non-invasive HR monitoring can be employed to build a comprehensive affective computing system for analyzing human emotion and behavior.

Recent studies demonstrating how changes in intensity of skin regions during cardiac circle can be captured by ordinary camera have stimulated the utilization of input from video camera to measure HR [1]. However, it remains a challenging problem due to the relatively small signal changes compared to fluctuation caused by various environments and subject factors such as subject's movement. Most of the previous works investigate this problem using face videos recorded under well-controlled environments. For example, subjects are asked to stay still and videos are recorded under bright lighting conditions. However, in real scenarios, especially, during long term health monitoring, movement and illumination changes are inevitable. Furthermore, previous medical studies [2] show that continuous HR information can provide further indication of heart condition. Moreover, a real time system can provide fast and convenient data collection process for users. Despite all of the above benefits, few methods are developed to remotely measure HR in real time using face input video.

In this paper, we propose a novel framework to work on real-time and provide robust performance under challenging environment conditions such as subject's movements, low illumination, etc. The paper is organized as follows: review of previous works in Section 2 to highlight the drawbacks of current approaches, description of our framework to overcome these drawbacks in Section 3 and analysis of experimental results in Section 4. The paper ends with discussion of the remote HR monitoring system applicability and conclusion.

2. RELATED WORKS

The use of PPG, a low cost and unobtrusive method of capturing the changes of cardiovascular pulse (also called blood volume of micro-vascular) during cardiac circles for remote physiological measurement has been an attractive research topic. PPG is based on the principle that when skin region is illuminated by light source, blood absorbs light more than surrounding tissue, so the variation in blood volume affects trans-

mission or reflectance of light. Since blood volume changes due to heart pulse, PPG can be used to measure HR. Traditionally, PPG has been developed using dedicated light sources, i.e., red and/or infra-red wavelengths. However, recent works report that HR measurement can be achieved using commercial cameras with normal ambient light.

One of the recent attempts to measure HR using video camera is done by Poh *et al.* [3, 4], in which face videos recorded by a web-cam are used as input. They detect the region of interest (ROI) as center of face area, and compute the mean pixel values in 3 channels (R,G,B) for each frame as temporal signals. Then Independent Component Analysis (ICA) is used to separate PPG signal from 3-channel signals. The PPG signal is passed through temporal filter and transformed to the frequency domain to find the peak value within the certain range, i.e., 0.7 to 4 Hz which corresponds to the human HR range. According to previous works [1], green channel contains the strongest plephysmographic signal. Therefore, they compare their result with method using only green channel, and show that using ICA can improve the measurement accuracy.

Lewandowska *et al.* [5] propose a different method to measure HR directly from web-cam by applying principal component analysis (PCA) to the temporal signal recorded from (R,G,B) channels, followed by frequency analysis. Another approach is proposed by Lan wei *et al.* [6] in which nonlinear dimension reduction technique, i.e., Laplacian Eigenmap, is employed into 3-channel signal to extract the PPG.

An approach in [7] by Xu *et al.* is based on the same physical principles as pulse oximeter, which uses red and infrared light absorption characteristics of oxygenated and deoxygenated hemoglobin. They model skin pigmentation as linear combination of melanin and hemoglobin absorbance in the log space. By applying the Lambert-Beer law, they can extract HR-related signals. The result is reported computationally efficient and reasonably accurate.

Another interesting approach is published by Balakrishnan *et al.* [8], in which it is hypothesized that cynical movement of blood from the heart to the head via carotid arteries causes slight head movement at the cardiac frequency. By tracking subtle head oscillation and performing PCA over trajectories, HR signal can be extracted.

To our best knowledge, all of mentioned methods have several limitations considering the robustness and applicability in general scenarios:

1) They do not attempt to solve the subjects' movement problem since no module is designed and subjects are not allowed to move during data collection. In [3], Poh *et al.* accessed this problem and reports that slight movement could affect their system's performance. Similarly, in method [8], which is based on tracking subtle head oscillations, any changes in head position could also disrupt the HR signal. Since subjects tend to move around freely in the long term HR monitoring, these methods are usually not practical.

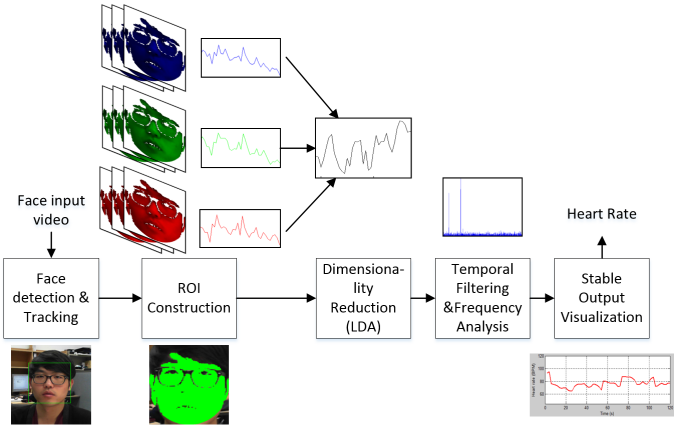

Fig. 1. Proposed real-time system flow for HR measurement using ordinary camera.

2) Few methods are reported to achieve the real-time/online measurement task. One of the main difficulties of this task is to attain low computational complexity requirement. Xu *et al.* [7] mention that their method is simple and computationally efficient; however, in other works, they do not clarify their system's computational complexity. The second difficulty is the short signal interval used for one HR calculation. For pulse-oximeter, a medical device to measure and display HR in real-time using LED placed on fingertips or earlobes, 8-20s signal interval is used. However, in the above methods, HR is obtained by processing 30-60s signal interval, which is not desired for real-time task. Due to limited and probably noise-contained information from short period of time, achieving correct HR value using short signal duration is a challenging task.

In this paper, we propose a novel system (see Fig. 1) that not only can deal with real-time tasks, but also produce robust and highly accurate measurement. Our contribution can be summarized as follows: first, we combine skin detection with face detection/tracking for dealing with subject heads movement, non-rigid facial motion (e.g., smiling, etc.), and various facial texture models (i.e., face with beards, glass or makeup, etc.). Second, we employ linear discriminant analysis (LDA) for utilizing information from three channels (R,G,B) to form HR signal. Lastly, we propose a framework for real time system. The framework contains modules for capturing/measuring and noisy output elimination and visualization using finite state machine.

3. PROPOSED FRAMEWORK

Our framework is composed of five steps as in Fig. 1. In the first step, we extract the preliminary ROI from raw video by combining the Viola and Jones's [9] face detector and the KLT tracker [10] to track a face region. Then we refine the ROI by eliminating unnecessary region in the face box from

step 1. In the third step, from the three-dimensional temporal signal, corresponding to (R,G,B) channels signal recorded for a time period, we apply a dimension reduction technique (i.e., LDA) to achieve a temporal HR signal. The HR signal is then passed through temporal filter and frequency analysis in the step 4, respectively, to reduce noise and detect the peak value in spectral power density corresponding to HR value. The last step is to identify noisy result (i.e., result that have noticeable difference from previous stable values) and display output. Details of each step are explained in the following subsections.

3.1. Face detection and tracking

Previous works [3], [4], [6], use the Viola-Jones face detector [9] to detect faces region in every frame. However it sometimes produce false positives and false negatives, which is not desirable to extract input. Furthermore, since the Viola-Jones method does not work well on non-frontal faces, subjects with face rotation and movement during measurement process can worsen the problem. In addition, face tracking allows us to update the system with multiple subjects' HR measurement function in an efficient way. For these reasons, we apply the Viola-Jone face detector in the first frame and then use the KLT tracker [10] to track the face. According to Baker *et al.* [10], the goal of the Lucas-Kanade algorithm is to find the match that minimizes the following function with respect to p:

$$\sum_{x}[I(W(x;p)) - T(x)]^2, \tag{1}$$

where T is the template image patch, I is the current frame, x is the position of the template window, W is the set of transformations between the last image frame and the current one and p is the parameter represents for the transformation. We apply the inverse compositional algorithm [10] to reduce computation load while maintaining performance as the traditional KLT algorithm produces.

3.2. ROI construction

The goal of this step is to construct precise ROI from the preliminary face region. We propose a simple, yet robust method to define efficient ROI by combining the first step with a skin detection method. Defining ROI using this process can alleviate the following problems in real world scenarios: subject's movement and non-rigid facial motion (e.g., smile), complex facial models (i.e., people with beard, glasses, etc.).

We utilize one of the skin detection method mentioned in [11], which applies spatial thresholding in normalized Y, Cb, Cr component. Denote S as the set of pixels classified as skin pigmentation, obtaining from the skin detection, and F as the set of pixels belong to the face region, then we can construct ROI by a set R of pixels:

$$R = S \cap F. \tag{2}$$

3.3. Dimensionality reduction

By calculating the mean values of the pixels in the ROI with three channels (R,G,B) overtime, we obtain the three temporal signals, which can be seen as an unified three dimensional signal by time. We observe that the fluctuations in (R,G,B) traces due to heart pulse have strong correlation between each channels. To construct the HR signal, we apply LDA, which is computationally light. Denote $x_1(t), x_2(t), x_3(t)$ as temporal signals corresponds to 3 channels, we first normalize them as follows:

$$y_i(t) = \frac{x_i(t) - \mu_i}{\sigma_i}, i \in \{1, 2, 3\}, \tag{3}$$

where μ_i and σ_i are the mean and the standard deviation of $x_i(t)$. To apply the LDA method, we need to construct class values from one out of three channel, and build the data from the other two dimensions with the corresponding class values. We notice that in [1], the authors argue the second channel (i.e., green signal) represents the strongest plephysmographic signal. Moreover, previous works and medical devices such as pulse oximeters use infrared light as the ambient light source, in which the significant component is red, to observe the PPG changes during cardiac circle. They implicitly indicate that the red channel represents for blood concentration. Therefore, we quantize the red channel signal $y_1(i)$ to form the class value corresponding to 2-dimensional sample $(y_2(i), y_3(i))$ with respect to i-th frame in time domain. Note that now the quantized $y_1(i)$ value represents for a class and $(y_2(i), y_3(i))$ represent for a 2-D sample in that class. The quantization level is determined by the number of classes from our system design. Using LDA with the class value formed by above method can exploit both the strongest PPG signal in green channel and additional information from blue channel. Denote $Y(t) = \big(y_2(t), y_3(t)\big)$ with $Y \in R^{t\times 2}$, our goal is to find the transformation $W \in R^{2\times 1}$ that maps $Y \in R^{t\times 2}$ to $Z \in R^{t\times 1}$, which is the desired temporal HR signal. W can be obtained by solving:

$$W^* = \underset{W}{\operatorname{argmin}} \frac{W^T S_B W}{W^T S_W W}, \tag{4}$$

where S_B is intra-class scatter matrix, and S_W is inter-class scatter matrix.

3.4. Temporal filtering and frequency analysis

Because HR frequencies fall in the range 0.7 to 4 Hz, which corresponds to 42-240 beats per minute (BPM), Hamming window based band-pass filter is used for removing non-related HR frequencies. One problem arisen with short processing signal interval is sparse frequency resolution; therefore, we apply the zero-padding technique in the temporal domain to mitigate this problem. By detecting f_{HR} corresponding the peak value in power spectral density in the frequency domain, HR can be measured as: $HR = 60 f_{HR}$.

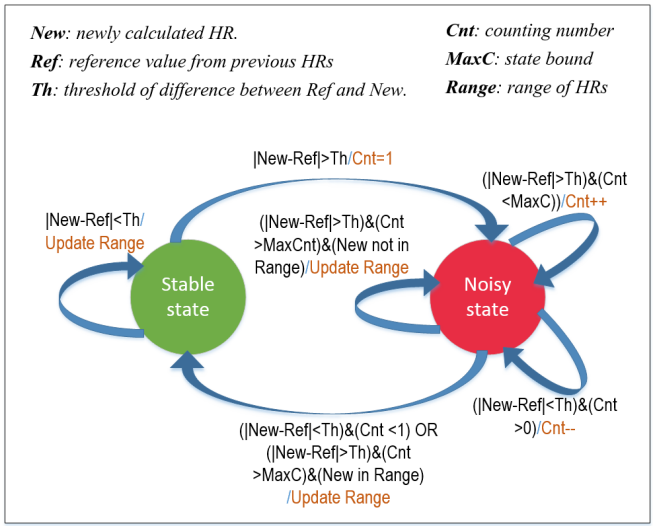

Fig. 2. Noisy output detection using finite state machine.

3.5. Real-time approach

In order to verify our system's computation time, we conducted an experiment on our collected dataset. We measure the algorithm's computing duration for each video in the dataset, and calculate the average of processing time for one frame. For comparison purpose, we carry the same experiment on Poh's [3] method, while Xu's [7] and green channel algorithm are not suitable for this analysis since those algorithms does not include similar modules. Table 1 and Table 2 demonstrate that our algorithm could process approximately 10 frames/second therefore, our modules are light enough to enable our system working on real-time. As a result, we can design framework to compute HR in real-time/on-line manner. We record signals for suitable period (6-10 seconds) and then compute HR at the end of each duration. This process is repeated with the input coming from new signals. Note that we use several previous frames, i.e., frames in previous 4-8 seconds, as an overlapping region signal. This scheme is logical for real-time measurement since the interval is reasonably short.

The problem with real-time systems is that signals can be noisy due to environment circumstances and limited information in short signal. To deal with this issue, we propose a finite state machine in order to detect noisy output. We assume that the output value difference between two consecutive HR measurements, which is between 2-4 secs in the proposed scheme, cannot be too large. When the output difference is greater than a threshold value, its state is set as noisy, and stay as noisy unless the difference gets back smaller than the threshold. When the noisy output is detected, the HR value is set as the previous result, i.e., the value in 2-4 seconds ago. Figure 2 explains details of our method.

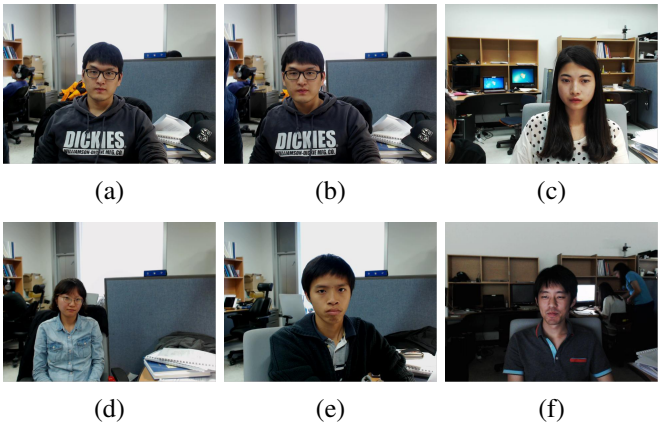

Fig. 3. Sample screen of videos collected in our challenging dataset. (a) and (b) are frame 11 and 111 of a video with subject's movement, (c), (d), (e), (f) are frames in videos of subjects with close distance to camera, far distance to camera, bight illumination condition and low illumination condition, respectively.

4. EXPERIMENTAL RESULT

4.1. Data collection

We used Logitech C920 webcam connected to desktop to record videos, and finger pulse-oximeter (Choicemmed Handheld MD300-K1E) to measure ground truth values. Face videos were recored while subjects were sitting in front of computer and ground truth values of HR were collected simultaneously every two seconds in real-time manner. Each video is RGB color with 640×480 pixel frames and 66 seconds long. As mentioned above, since real-time on-line measurement is our desired system, we recorded the videos with 10 frames per seconds, i.e., the sampling rate is 10Hz. Since frequencies corresponding to HR fall in the range of 0.7-4Hz, the sampling rate is 2 times higher than the signal frequencies. Therefore by Nyquist theorem, 10Hz is the sufficient sampling rate. The total number of subjects is 10 (3 females and 7 males), ranging from the age of 20-35 with various face appearances (i.e., glass and facial makeup). To compare our algorithm with others', we setup several datasets, including: fine dataset and challenging dataset. The fine dataset is constructed under well-constrained environment such as: no subject's movement, bright illumination condition with ambient light. The challenging dataset, on the other hand, is composed of several sub-datasets that reflect real world scenarios: movement dataset (subjects are asked to move continuously, with different degrees from small to big and different types (head rotation and body movement)), distance to camera dataset (from short (0.5-0.6 meter) to long (1-1.2 meters)), and illumination condition dataset (from bright to low). Figure 3 shows several screen examples of our challenging dataset. In

Table 1. HR measurement performance comparison with the fine dataset using long period of signal, i.e., 30-40 seconds signal interval. Average computation time for one frame on the whole dataset is also included.

Method	ME (bpm)	RMSE (bpm)	MPE (%)	Comp. time (s/frame)
Green channel	-2.08	3.44	3.09	-
Poh *et al.*	-1.38	2.56	2.59	0.047
Xu *et al.*	**-0.88**	1.67	**1.49**	-
Ours	**-0.88**	**1.53**	1.58	0.068

total, there are 10 videos in the fine dataset and 29 videos in the challenging dataset collected under the conditions mentioned above, and in each video, 33 HR ground truth values are also collected.

4.2. HR measurement with the fine dataset

The purpose of this experiment is to check the performance of several approaches with the fine dataset. We re-implemented two previous methods: Poh *et al.* [3] and Xu *et al.* [7], and another method using only the raw green channel signal. In order to have fair comparison, we apply the same temporal and frequency analysis step since previous works also used similar filter and peak detection technique. Several statistical analysis criteria were used to evaluate the accuracy of the HR measurement methods, i.e., mean error (ME), root mean square error (RMSE) and mean percentage error (MPE). For the first experiment, we use a long period of signal, i.e., 30-40s interval to estimate HR. This experiment is done with 10 measurements on 10 subjects and the result is shown in Table 1. We observe that with sufficiently long signal interval and well conditioned data, we can obtain the significantly low error performance for all of the above methods. Our proposed algorithm and Xu *et al.*'s method achieved comparable results, with slightly better performances than Poh *et al.*'s and the algorithm using only green channel. However, consuming 30-40s signal interval is not ideal for real-time, continuous HR measurement. In the remaining section, we compare our approach using the shorter period of signal, i.e., 8 seconds.

4.3. Real-time HR measurement with the challenging dataset

The purpose of the second experiment is to test the above algorithms' performance on real world-like scenarios. We use 8-seconds interval for one HR measurement, and subsequent measurements were performed using 75% signal overlap (6 seconds) to calculate HR every 2 seconds. Eight second interval is reasonable since several medical devices measuring HR in real-time mention to use this period length. Because each video is 66-second long and we start measuring HR at 8th second, in total we have 30 outputs per video.

Table 2. HR measurement performance comparison with the challenging dataset using short period of signal, i.e., 8 seconds signal interval. (1) and (2) are denoted for the movement sub-dataset and the whole challenging dataset, respectively.

Data-set	Method	ME (bpm)	RMSE (bpm)	MPE (%)	Comp. time (s/frame)
1	Green chn.	-2.66	12.50	13.05	-
	Poh *et al.*	**-0.62**	8.88	7.89	0.055
	Xu *et al.*	3.90	14.21	13.82	-
	Ours	-1.14	**5.38**	**4.05**	0.083
2	Green chn.	-0.81	12.64	12.69	-
	Poh *et al.*	-0.81	10.08	9.23	0.056
	Xu *et al.*	2.83	11.72	11.16	-
	Ours	**-0.54**	**5.72**	**4.89**	0.086

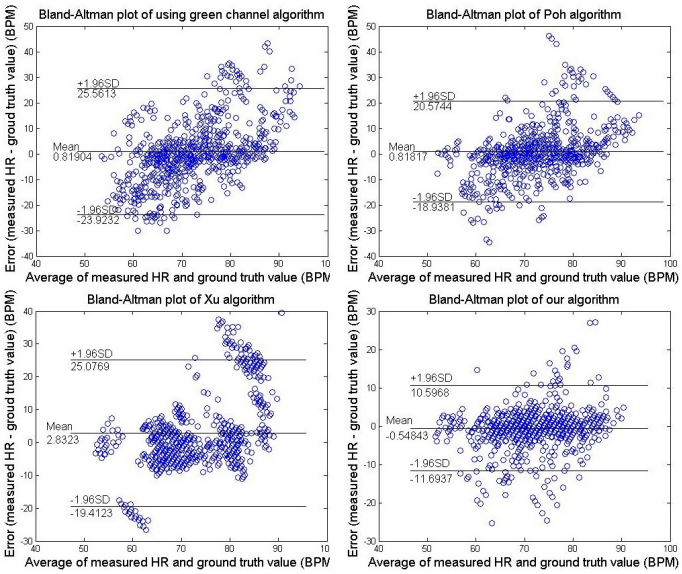

Fig. 4. Bland-Altman plots of the measurement error of algorithm using green signal only, Poh *et al.*'s, Xu *et al.*'s and our proposed scheme in upper-left, upper-right, lower-left, lower-right figure respectively. The analysis was done with total 870 measurements on our challenging video dataset.

Table 2 and Figure 4 show statistical analysis of the mentioned algorithms' results. We observe that with the movement sub-dataset, while our proposed scheme maintain fairly small error results, other algorithms are severely affected due to noisy input. Similarly, the algorithm using green channel produces the worst output on our total challenging dataset, while Xu *et al.*'s and Poh *et al.*'s still produce relatively large error. The average computation time demonstrate that both

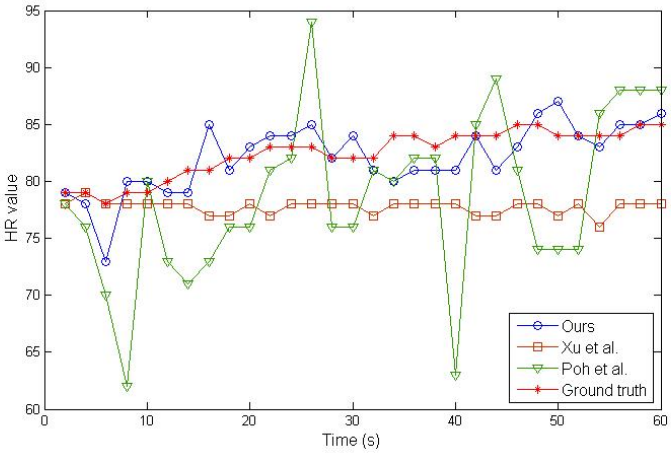

Fig. 5. HR measurement sample results of Poh *et al.*'s algorithm, Xu *et al.*'s algorithm and ours compare to the ground truth values for one subject during 60 seconds. HR values are measured continuously every 2 seconds.

Poh *et al.*'s algorithm and ours could work in real time, however, our system are able to produce more correct results. The Bland-Altman analysis showed our 96% confident interval for error is significantly improved, (-11.69 BPM to +10.59 BPM) compared to other approaches'. Figure 5 shows sample results of a video in the movement dataset. While our algorithm is able to capture small fluctuation of HR, Poh *et al.*'s and Xu *et al.*'s algorithm were unable to follow the ground truth values. To sum up, our method demonstrate evident improvement under challenging conditions with short interval signal over those previous approaches.

5. CONCLUSIONS

In this paper, we have proposed a real time measuring system for human HR using ordinary cameras, which is reliable and robust in various environment conditions. Our scheme is built from the combination of face detection/tracking and skin detection for precise ROI construction, utilization of three-color channel signals (R,G,B) to form the HR signal using LDA, and the finite state machine to detect noisy output in unforeseen conditions. The method has shown reasonable performances under various environments and performed better than several currently available methods. We expect that this system can replace traditional pulse oximeters and be applied to various practical fields such as remote health care monitoring and humans physiological analysis.

6. REFERENCES

[1] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson, "Remote plethysmographic imaging using ambient light," *Optics express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[2] Vanesa España-Romero, Rajna Golubic, Kathryn R Martin, Rebecca Hardy, Ulf Ekelund, Diana Kuh, Nicholas J Wareham, Rachel Cooper, Soren Brage, et al., "Comparison of the epic physical activity questionnaire with combined heart rate and movement sensing in a nationally representative sample of older british adults," *PloS one*, vol. 9, no. 2, pp. e87085, 2014.

[3] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, no. 10, pp. 10762–10774, 2010.

[4] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2011.

[5] Magdalena Lewandowska, Jacek Ruminski, Tomasz Kocejko, and Jedrzej Nowak, "Measuring pulse rate with a webcama non-contact method for evaluating cardiac activity," in *Proc. Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2011, pp. 405–410.

[6] Lan Wei, Yonghong Tian, Yaowei Wang, Touradj Ebrahimi, and Tiejun Huang, "Automatic webcam-based human heart rate measurements using laplacian eigenmap," in *Proc. Asian Conference on Computer Vision (ACCV)*, pp. 281–292. 2013.

[7] Shuchang Xu, Lingyun Sun, and Gustavo Kunde Rohde, "Robust efficient estimation of heart rate pulse from video," *Biomedical optics express*, vol. 5, no. 4, pp. 1124–1135, 2014.

[8] Guha Balakrishnan, Fredo Durand, and John Guttag, "Detecting pulse from head motions in video," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3430–3437.

[9] Paul Viola and Michael J Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[10] Simon Baker and Iain Matthews, "Lucas-kanade 20 years on: A unifying framework," *International journal of computer vision*, vol. 56, no. 3, pp. 221–255, 2004.

[11] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recognition*, vol. 40, no. 3, pp. 1106 – 1122, 2007.