Robust real-time heart rate prediction for multiple subjects from facial video using compressive tracking and support vector machine

Lingling Liu
Yuejin Zhao
Lingqin Kong
Ming Liu
Liquan Dong
Feilong Ma
Zongguang Pang

Robust real-time heart rate prediction for multiple subjects from facial video using compressive tracking and support vector machine

Lingling Liu, Yuejin Zhao,* Lingqin Kong, Ming Liu, Liquan Dong, Feilong Ma, and Zongguang Pang

Beijing Institute of Technology, School of Optoelectronics, Beijing Key Laboratory of Precision Photoelectric Measuring Instrument and Technology, Beijing, China

Abstract. Remote monitoring of vital physiological signs allows for unobtrusive, nonrestrictive, and noncontact assessment of an individual's health. We demonstrate a simple but robust image photoplethysmography-based heart rate (HR) estimation method for multiple subjects. In contrast to previous studies, a self-learning procedure of tech was developed in our study. We improved compress tracking algorithm to track the regions of interest from video sequences and used support vector machine to filter out potentially false beats caused by variations in the reflected light from the face. The experiment results on 40 subjects show that the absolute value of mean error reduces from 3.6 to 1.3 beats/min. We further explore experiments for 10 subjects simultaneously, regardless of the videos at a resolution of 600 by 800, the HR is predicted real-time and the results reveal modest but significant effects on HR prediction. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.5.2.024503]

Keywords: remote monitoring; image photoplethysmography; heart rate; compress tracking; support vector machine.

Paper 17228RRR received Jul. 28, 2017; accepted for publication Apr. 30, 2018; published online Jun. 29, 2018.

1 Introduction

Human heart rate (HR) is an important physiological parameter with numerous applications, including clinical diagnosis and health monitoring. The methods to detect HR vary from contact to noncontact,¹ and the noncontact method can be divided into three categories in chronological order: microwave Doppler radar,² thermal imaging,³ and visible imaging algorithms.^{4,5} Among the aforementioned methods, video-based HR estimation using image photoplethysmography (IPPG) methodology is easy to operate, inexpensive, and noninvasive, and the ubiquity of cameras across a myriad of consumer devices makes this technique a focal point of research in biomedicine. People have used multiple methods to identify HR within videos. Poh et al.⁴ proposed the algorithm of HR detection based on independent component analysis (ICA) technique with three participants. ICA is computationally intensive and is at times ambiguous as to the order of the decomposed channels.⁶ In recent years, some researchers have used machine learning methods for HR estimation, followed by the work of Poh et al., and Osman et al.⁷ used a support vector machine (SVM) based approach to detect HR from the face videos, but the predicated HR could be erroneous in frames, where the IPPG signal is masked by artifacts in the video. To visualize pulse in subjects, some researchers have applied Eulerian Video Magnification;⁸ this technique amplifies color differences within a particular frequency spectrum to enable visualization of imperceptible features, which works well on a single subject. However, this is particularly detrimental when there are multiple subjects with different HRs present in the video. Furthermore, Estrada and Stowers⁹ detect HR for multiple subjects by combining detection of HR with amplifying it for visualization, but

this technique cannot be implemented in real-time and will magnify the noise and compression artifacts. To realize HR estimation in real-time, faces tracking was an effective approach. The previous approaches proposed in simultaneous tracking of multiple faces in real-time were mainly based on continuously adaptive mean-shift algorithms (Camshift),¹⁰ which performed well in accurately estimating object locations, but performed poorly in consistently labeling objects over time.¹¹ A lot of studies showed that compress tracking (CT) algorithm¹² could run in real-time and performs favorably against state-of-the-art algorithms on challenging sequences in terms of efficiency, accuracy, and robustness, and this study is primarily aimed at single target tracking and takes into account multitarget tracking issues.

In this paper, we proposed to use the improved CT algorithm and SVM to accurately predicate HR for multiple subjects in real-time. In contrast with the previous work, we accurately and quickly obtained multiple subject's region of interest (RoI) through improving the CT algorithm and filtered out the false IPPG signal through SVM, instead of extracting feature from the signal, we trained the SVM with the IPPG signal directly. This makes it possible for the health monitoring of multiple subjects simultaneously in daily lives and promote the clinical application of IPPG technology.

2 Method

2.1 Principles of IPPG-Based HR Prediction

For the IPPG-based HR estimation, our objective is to develop an algorithm to predicate the underlying PPG signal $p(t)$ using the recorded video of a person's face. The recorded video is in

*Address all correspondence to: Yuejin Zhao, E-mail: yjzhao@bit.edu.cn

the form of intensity signal $V(x, y, t)$ comprising of sequences of frames ($t = 1, 2, 3, \dots$). Each frame of the video records the intensity level of the light reflected back from the face over a two-dimensional grid of pixel (x, y) in the camera sensor.

In general, the measured intensity of the reflected light can be decomposed into two components: (i) intensity of surface reflection $R(x, y, t)$ and (ii) intensity of subsurface reflection and backscattering $S(x, y, t)$:

$$V(x, y, t) = R(x, y, t) + S(x, y, t). \quad (1)$$

The surface reflection intensity $R(x, y, t)$ determined by the intensity of ambient light falling on the face is generally assumed to remain the same over the PPG estimation window (typically 5 to 60 s), as a large part of the light incident on face gets reflected back from the surface of the skin, and is characterized by the skin's bidirectional reflectance distribution function. Remaining part of the incident light goes underneath the skin surface and is absorbed by the tissue and the chromophores (Hb, HbO₂) in blood. The intensity of $S(x, y, t)$ changes as the volume of blood in the arteries and capillaries changes within a cardiac cycle. Since PPG signal, by definition, is proportional to this cardiosynchronous pulsatile blood volume change in the tissue,¹³ one can predicate PPG signal $p(t)$ by extracting small variations in the subsurface skin reflection and backscattering $S(x, y, t)$:

$$p(t) = \sum_{x=0}^w \sum_{y=0}^h S(x, y, t), \quad (2)$$

where w and h are the width and height of the RoI, respectively.

2.2 Experimental Setup and Measurement Procedure

Our experimental scheme is shown in Fig. 1. Nikon CoolPix L610 was used as the video-capturing unit, and a personal computer (PC) with conventional 3 GHz Intel Core i5-4430 (ASUS Co.) is employed to record the videos in real-time for analysis. All videos including human face were recorded at 30 frames per second (fps) with pixel resolution of 600×800 and saved as AVI format in the PC.

All experiments were performed indoors with a normal ambient light source, videos containing human face are subsequently transmitted to PC equipped with a conventional 3 GHz Intel Core i5-4430. Meanwhile, a finger BVP sensor (Qinhuangdao Contec Medical System Co., Ltd.) is used to

Fig. 1 Experimental setup for the study.

measure relevant signal of the subject at the same time. About 40 subjects (28 males and 12 females, Chinese ethnic group) with an age range from 22 to 35 are enrolled for this study and all of the subjects had no history of cardiovascular disease and did not use drugs that might influence heart rhythm, all subjects gave their written informed consent of participation in the experiment. During the experiment, the participants were allowed to move slightly, hence the collected data contained spontaneous movements.

2.3 Region of Interest Extraction

We first obtain accurate RoI by face detection and tracking algorithm. The face detection algorithm we utilized, which was realized in an Open Computer Vision library (OpenCV), was based on the work of Viola and Jones,¹⁴ as well as the work of Lienhart and Maydt.¹⁵ The face tracking algorithm we improved, which was initially proposed by the authors in Ref. 12, was based on compressive sensing theories.¹⁶ Considering that more than one subject was tracked in our experiment, the geometrical distance between two subjects should be small and variable due to the subjects' movement, four approaches were proposed to enhance the tracking performance. First, we initialized the face tracking algorithm with the face detection results (more than one box) instead of the labeled tracking box, assigning each object a unique track ID, which stays constant throughout the sequences. Second, a variable positive sample radius was proposed to improve the correspondence,¹² because close distance will lead to overlap of one's negative sample region with another's positive sample region, further leading to failure in updating the classifier, especially when there exists occlusion between the adjacent subjects, face tracking was terminated, and face detection was restarted. As indicated in Fig. 2, we partitioned the neighboring space into eight subspaces, as listed in the whole image clockwise from 1 to 8. Suppose the proposed detection box has the parameters x_0, y_0, w_0 , and h_0 , and its adjacent detection box has the parameters x_i, y_i, w_i , and h_i , where i ranges from 1 to 8.

Moreover, we provided the calculation of the negative sample radius of the proposed detection box r_{negative} , as shown in Table 1. We chose the smallest r_{negative} as the negative sample radius in the current frame.

Third, we detected the subjects' position on the video to avoid incomplete RoI, as the smaller is the RoI, the bigger is the HR estimation error. Finally, but equally important, the output of the Bayes classifier, which also indicates the feature matching degree, was used as a flag of operating mode. That is to say, if the flag was lower than the threshold we set due to the drift of the tracking box or part of the subjects' face moving

Fig. 2 Eight possible positions of the detection box's neighboring box.

Table 1 Different negative sample radii of the detection box based on adjacent boxes.

Position	Criterion	Positive sample radius
1	$\begin{cases} x_i + w_i \leq x_0 \\ y_i + h_i \leq y_0 \end{cases}$	$r_{\text{negative}}^2 < (x_i + w_i - x_0)^2 + (y_i + h_i - y_0)^2$
2	$\begin{cases} y_i + h_i < y_0 \\ x_i + w_i > x_0 \\ x_i < x_0 + w_0 \end{cases}$	$r_{\text{negative}} < y_0 - (y_i + h_i)$
3	$\begin{cases} y_i + h_i \leq y_0 \\ x_i \geq x_0 + w_0 \end{cases}$	$r_{\text{negative}}^2 < (x_0 + w_0 - x_i)^2 + (y_i + h_i - y_0)^2$
4	$\begin{cases} y_i + h_i > y_0 \\ x_i > x_0 + w_0 \\ y_i < y_0 + h_0 \end{cases}$	$r_{\text{negative}} < x_i - (x_0 + w_0)$
5	$\begin{cases} y_0 + h_0 \leq y_i \\ x_0 + w_0 \leq x_i \end{cases}$	$r_{\text{negative}}^2 < (x_0 + w_0 - x_i)^2 + (y_0 + h_0 - y_i)^2$
6	$\begin{cases} y_0 + h_0 < y_i \\ x_0 < x_i + w_i \\ x_0 + w_0 > x_i \end{cases}$	$r_{\text{negative}} < y_i - (y_0 + h_0)$
7	$\begin{cases} y_0 + h_0 \leq y_i \\ x_0 \geq x_i + w_i \end{cases}$	$r_{\text{negative}}^2 < (x_i + w_i - x_0)^2 + (y_0 + h_0 - y_i)^2$
8	$\begin{cases} y_0 < y_i + h_i \\ x_0 > x_i + w_i \\ y_0 + h_0 > y_i \end{cases}$	$r_{\text{negative}} < x_0 - (x_i + w_i)$

out of the video, then face tracking was terminated and face detection was restarted. Each face was modeled as a rectangle with parameters x, y, w , and h , where x and y are the origin coordinates of the detection box, and w and h are the width and height of the detection box, respectively.

The RoI was extracted after two steps, as shown in Fig. 3. First, we obtained the parameters of the face tracking box. Second, the rectangle of the RoI was selected based on Eq. (3), where $h1$ is the height of the tracking box and $h2$ is the height of the RoI, from a database of about 400 frontal face images created in our lab, a study on the relationship between the two parameters were made as follows:

$$h2/h1 = 0.6. \quad (3)$$

2.4 Data Analysis

Afterward, we obtain the raw IPPG data by calculating the gray-scale average value of the RoI with hue space in HSV

Fig. 3 Definition of RoI utilized in the paper.

(hue, saturation, and value) instead of using the most common RGB color scheme, as HSV scheme is invariant to viewing orientation and illumination intensity.¹⁷ In order to inhibit the random noise for every RoI's H channel, the pixel value in each frame was added and normalized as follows:

$$x'_i = x_i - \mu. \quad (4)$$

For each $i = 1, 2, 3, \dots$, which means the frame number within 2 s, and μ are the mean of x_i in the data list. The normalization transforms x_i to x'_i , which is zero mean and corresponds to Eq. (1) in Sec. 2.1. Finally, to eliminate potential noise and increase the robustness of the algorithm, the list of x'_i was preprocessed with a bandpass filter (128-point Hamming window, 0.6 to 3 Hz) as well as a time-frequency method wavelet analysis in Ref. 18. Nevertheless, the preprocessing of the raw IPPG data faced with several technical challenges. First, the false positive¹⁹ in face detection, as the gathering of the sequence of gray value is in real-time and consecutive, when the presently detected subject belongs to the background as a result of the false positive of the detection, the raw IPPG data were valueless and made the noise reduction and HR estimation worthless, almost no research is applicable on this problem as far as our knowledge. Second, the light variations and motion artifacts influence the accuracy of HR estimation as the noise reduction method cannot completely reduce the noise caused by the variation of reflected light on the subject's face. To solve this problem, a widely known type of SVM called SVC was used as a binary classifier. The operation is shown in Ref. 20, given the training database:

$$z = \{(x_i, y_i) | x_i \in R^n, y_i \in \{-1, 1\}, i = 1, 2, \dots, m\}. \quad (5)$$

where x_i is a vector data and y_i is the corresponding expected output. Considering the number of feature and training examples, we use the following type of logistic regression kernel in this work:

$$\text{Sigmoid: } \kappa(\mathbf{U}, \mathbf{V}) = \tanh(g\mathbf{U}^T \mathbf{V}). \quad (6)$$

We implemented the SVM in MATLAB version 8.4 (R2014b) using the libsvm library.²¹ The obtained database was divided into SVC training (70%) and testing (30%), maintaining the unaltered unbalance among classes. A grid search was performed with the training database to determine the best parameters (c and g) for the SVC classification. We used cross-validation with 10-fold for each case. The performance of the classifier was evaluated using the following metrics: sensitivity—true positive rate (TPR), specificity—true negative rate (TNR), true negative rate, and accuracy (ACC):

$$\begin{aligned} \text{FPR} &= \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{\text{FP}}{N}; \\ \text{PPR} &= \frac{\text{TP}}{\text{TP} + \text{TN}} = \frac{\text{TP}}{P}; \\ \text{ACC} &= \frac{\text{TP} + \text{TN}}{P + N}. \end{aligned} \quad (7)$$

PPG signal was highly correlated with HR signal²² and HR signal was classified as fractal signal in Ref. 23, which was different from the noise (random signal). Thus, in run time, the trained SVM model is used as a sliding window; once the pre-processing of the raw IPPG data finished, the data list with

Fig. 4 Boxes marked as N illustrate the location, where negative samples were extracted. The box in the subject's face marked as P illustrates the location, where the positive samples were extracted.

the length of 64 (equals to about 3.5 s) was used as the feature data list directly without further feature extraction. In our experiment, the training database contained 230 positive data lists and 92 negative data lists, and the validation dataset consisted of 100 positive and 38 negative samples, we train the model with different parameters until $ACC > 80\%$. Figure 4 illustrates the training samples extracted from the video, the false detected box and the detected box with large drift provided negative samples.

2.5 HR Computation

At last, the data lists which were classified as the positive data were handled with the fast Fourier translation to compute the HR, and the spectrum with the maximum peak across 0.75 and 3.0 Hz (corresponding to 45,180 bpm) was selected to calculate the HR.

3 Results

In order to validate the performance of improved CT algorithm, we perform the experiment on 10 participants, as shown in Fig. 5. The predicated HR data and the corresponding ID were displayed in real time on the video and the detectors observed the results conveniently. From the picture, we can directly recognize the effective improvement using the modified CT algorithm [Fig. 5(b)] in contrast with the unmodified CT algorithm [Fig. 5(a)].

Fig. 5 Comparison of tracking effect for ten subjects using (a) unimproved CT (Video 1) and (b) improved CT (Video 2). The ID in the left was obtained from the face detection algorithm, the HR in the middle was the HR of the corresponding subject (Video 1, MP4, 2.3 MB [URL: <https://doi.org/10.1117/1.JMI.5.2.024503.1>]; Video 2, MP4, 2.18 MB [URL: <https://doi.org/10.1117/1.JMI.5.2.024503.2>]).

Table 2 Comparison of RMSE and CC using the improved CT and unimproved CT.

Method	RMSE	CC	fps
Unimproved CT	5.0251	0.2271	20
Improved CT	2.7685	0.8866	17
Camshift	×	×	17

The root mean square error (RMSE) and concordance correlation coefficient (CC) between the predicated HR and the ground truth HR recorded by the oximeter were used to further evaluate the performance of our method, as shown in Table 2, also, the tracking speed of the Camshift algorithm and CT algorithm was compared. It should be noted that the results reported in Table 2 are averaged across all 10 subjects within 2-min intervals. From Table 2, we can see that the improved CT algorithm has almost the same tracking speed of 17 fps compared with the Camshift algorithm. Overall, the proposed method gives an obvious improvement contrast with the unimproved CT algorithm, where the root mean square error has reduced to 2.7685 BPM and the correction coefficient between the predicated HR and the actual HR increased from 0.2271 to 0.8866 at the cost of another 8 ms in every frame. It should be noticed that, we have not recorded the results using the Camshift algorithm, as the algorithm cannot consistently label objects over time, which will lead to invalid IPPG data.

The parameters of Pearson’s correlation coefficient (CC), average absolute error (AAE), average error percentage (AEP) and average simulation time per frame (ASTPF) are used to evaluate the performance of SVM, as shown in Table 3. The bigger value of CC, the smaller value of AAE and AEP, the better performance of SVM in prediction will be obtained.

Table 3 Evaluation parameters for SVM.

CC	AAE	AEP	ASTPE
0.8786	0.386	7.29%	10 ms

Fig. 6 Bland–Altman plots analysis of the agreement between actual HR using oximeter and predicted HR for multiple subjects using (a) unimproved CT, (b) improved CT, and (c) improved CT and SVM.

Fig. 7 Single face tracking with CT (Video 3, MP4, 0.28 MB [URL: <https://doi.org/10.1117/1.JMI.5.2.024503.3>]).

Fig. 8 Multiple faces tracking with Camshift (Video 4, MP4, 1.38 MB [URL: <https://doi.org/10.1117/1.JMI.5.2.024503.4>]).

Figure 6 shows the agreement between the predicted HR and the actual HR, which was tested by Bland–Altman analysis between 510 pairs of measurements from 40 subjects. The mean bias d was -3.6 bpm with 95% limits of agreement from -18.5 to 11.3 bpm [Fig. 6(a)] without improving CT. The mean bias d was -2.4 bpm with 95% limits of agreement from -13.4 to 8.6 bpm [Fig. 6(b)] using the improved CT. By contrast, the mean bias d was -1.3 bpm with 95% limits of agreement from -8.6 to 6 bpm [Fig. 6(c)] using the proposed method with improved CT and SVM. In the figure, the different values exhibited large variations for midrange average values (around 65 to 75) and minimal variations for small and large average values. Thus, more people in this range and fewer people at two extremes were observed. In order to complete the experimental results, we also show the single face tracking result with CT in Fig. 7 and the multiple faces tracking result with Camshift in Fig. 8.

4 Discussion

Among all of the subjects, the range of HR for the subjects has varied from 58 to 94 bps, and the intrasubject variability of HR has ranged from 5 to 10 bps during the measuring time. Our tracker can handle complete occlusion and fast movements well as its appearance model is discriminatively learned from

target and background with a data-independent measurement, thereby alleviating the influence from background, but in consideration of the accuracy of the HR measurement results, the range of spontaneous movements was less than about 20 deg, and the occlusion was avoided during the measurement.

In the real scenario, if the HR measurement environment changed, we should train the grayscale in the RoI to obtain feature data, and if the HR measurement environment has not changed, we have the feature data in advance, as the HR measurement accuracy largely depends on the measurement environment. That is to say, the limitation of this technique is that the HR measurement environment decides whether we need to train data or use the data in advance.

5 Conclusion

In conclusion, we have proposed methods to improve HR estimation speed and accuracy of a noncontact HR monitor system for multiple persons through facial video. The video was obtained from phones or webcams directly. The RoI was tracked rapidly and accurately using the improved CT algorithm via four approaches: fixed tracking ID, variable sample radius (avoiding occlusion), detecting the position of the subjects, and setting a threshold for the output of the classifier. The H channel of the HSV color model was used to extract the raw IPPG data, which

were classified by the SVM after preprocessing. The noise and signal with low SNR due to false detection and motion artifact were further excluded. The experimental results have shown that the HR estimation reached 17 fps on a modestly equipped H channel of the HSV color model was used to extract the raw IPPG data, which were classified by the SVM after preprocessing. The noise computer and the estimation error have almost reduced to half compared with unimproved method. The proposed method could be variable sample radius (avoiding occlusion), detecting the position of the subjects, and setting a threshold for the output of the classifier. Furthermore, we intend to exploit Android OS application software with our proposed method to realize the pulse rate monitoring daily.

Disclosures

The authors have no financial interests or conflicts of interest in this manuscript.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 61705010, 61301190, and 61475018).

References

1. Y. Sun and N. Thakor, "Photoplethysmography revisited: from contact to noncontact, from point to imaging," *IEEE Trans. Biomed. Eng.* **63**(3), 463–477 (2016).
2. K. M. Chen et al., "An X-band microwave life-detection system," *IEEE Trans. Biomed. Eng.* **33**(7), 697–701 (1986).
3. J. Fei and I. Pavlidis, "Thermistor at a distance: unobtrusive measurement of breathing," *IEEE Trans. Biomed. Eng.* **57**(4), 988–998 (2010).
4. 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**(10), 10762–10774 (2010).
5. W. Verkrusysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express* **16**(26), 21434–21445 (2008).
6. S. Xu, L. Sun, and G. K. Rohde, "Robust efficient estimation of heart rate pulse from video," *Biomed. Opt. Express* **5**(4), 001124 (2014).
7. A. Osman, J. Turcot, and R. El Kaliouby, "Supervised learning approach to remote heart rate estimation from facial videos," in *11th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. **1**, pp. 1–6 (2015).
8. H. Y. Wu et al., "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graphics* **31**(4), 65 (2012).
9. M. Estrada and A. Stowers, "Amplification of heart rate in multi-subject videos," Stanford California.
10. D. Comaniciu and P. Meer, "Mean shift: a robust approach towards feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 603–619 (2002).
11. K. Bernardin and R. Stiefelhausen, "Evaluating multiple object tracking performance: the CLEAR MOT metrics," *EURASIP J. Image Video Process.* **2008**, 246309 (2008).
12. K. Zhang, L. Zhang, and M. Yang, "Real-time compressive tracking," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 866–879, IEEE (2012).
13. J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.* **28**(3), R1–R39 (2007).
14. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conf. on Vision and Pattern Recognition*, p. 511, IEEE (2001).
15. R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proc. of IEEE Conf. on Computer Image Processing*, pp. 900–903, IEEE (2002).
16. H. Li and C. Shen, "Real-time visual tracking using compressive sensing," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1305–1312, IEEE (2011).
17. K. N. Plataniotis and A. N. Venetsanopoulos, *Color Image Processing and Applications*, Springer, Berlin (2000).
18. N. Hess-Nielsen and M. V. Wickerhauser, "Wavelets and time-frequency analysis," *Proc. IEEE* **84**, 523–540 (1996).
19. A. A. Abin, M. Fotouhi, and S. Kasaei, "Real-time multiple face detection and tracking," in *Int. CSI Computer Conf.*, pp. 379–384 (2009).
20. G. Y. Shi and S. Liu, "Model selection of RBF kernel for C-SVM based on genetic algorithm and multithreading," in *Int. Conf. on Machine Learning and Cybernetics*, pp. 382–386 (2012).
21. C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011).
22. Y. D. Lin, "Studies on the correlation between the features extracted from the pulse and PPG signals," *Ann. Tradit. Chin. Med.* **28**(1), CCMP96-TP-201 (2010).
23. M. Akay, *Biomedical Signal Processing*, Academic Press, San Diego, California (1994).

Lingling Liu is pursuing a master degree of instrument science and technology in Beijing Institute of Technology. She received her BS degree in optical engineering from Nanjing University of Posts and Telecommunications in 2008. Her interests involve applying quantitative and machine learning methods to noninvasive physiological monitoring. She is a member of SPIE.

Yuejin Zhao is a specially appointed professor of instrument science and technology in Beijing Institute of Technology, one of the directors of China Optical Society, associate director of optical committee in China Ordnance Society, and vice president of Optical-electromechanical integration technology chapter of China Instrument and Control Society. His research areas include MEMS infrared imaging technology, terahertz imaging and spectrum, human physiological parameters with optical detection, space optics, and intelligent photoelectric instruments. He is a member of SPIE.

Lingqin Kong is currently a lecturer and the master tutor in Beijing Institute of Technology. She received her BS and MS degrees in physical electronics from Taiyuan University of Technology in 2006 and 2009, respectively, and her PhD degree in instrument science and technology from the Beijing Institute of Technology in 2014. Her current research interests include video processing for vital signs, optical system design, and high-speed aero-optic effects. She is a member of SPIE.

Ming Liu is an associate professor of instrument science and technology in Beijing Institute of Technology. He received his BS degree in optics and Electronics from Shan Dong University, 2002, and his MS degree in optical engineering and PhD degree in Instrumental Science and Technology from Beijing Institute of Technology in 2005 and 2009, respectively. His research areas are mainly focused on applications of deep learning on novel optics design and image processing. He is a member of SPIE.

Liquan Dong is an associate professor of instrument science and technology in Beijing Institute of Technology. He received his BS and PhD degrees in physics from the University of Beijing Institute of Technology in 2000 and 2007, respectively, his current research mainly include the optical instrument, digital signal and image processing, computational imaging, infrared and THz imaging. He is a member of SPIE.

Biographies for the other authors are not available.