IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 10, 2018, accepted May 23, 2018, date of publication June 8, 2018, date of current version July 6, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2845390

Heart Rate Extraction Based on Near-Infrared Camera: Towards Driver State Monitoring

QI ZHANG[1,2], YIMIN ZHOU[2], (Member, IEEE), SHUANG SONG[3], GUOYUAN LIANG[2], AND HAIYANG NI[4]

[1]Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin University of Electronic Technology, Guilin 541004, China
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[3]Department of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China
[4]Tianjin Academy of Traditional Chinese Medicine Affiliated Hospital, Tianjin 300193, China

Corresponding author: Yimin Zhou (ym.zhou@siat.ac.cn)

This work was supported in part by the Shenzhen Science and Technology Innovation Commission Project under Grant JCYJ20160510154736343 and Grant JCYJ20170307165442023 and in part by the Guangxi Key Laboratory of Automatic Detecting Technology and Instruments under Grant YQ18106.

ABSTRACT In this paper, a remote sensing method for heart rate (HR) detection is proposed via near-infrared facial video data, which can be used for the online monitoring of the physiological parameters of the drivers and further analysis so as to enhance the driving safety. The proposed HR detection method is based on an automatic facial tracking algorithm, i.e., the Kanade-Lucas-Tomasi, to transform the facial images into time-series signals. Empirical mode decomposition, bandpass filtering and fast Fourier transformation are applied to the time-series signals for the HR extraction. Indoor experiments in different scenarios are conducted to discuss the involved impact factors, i.e., the distance between the near-infrared camera and the participant, the illumination intensity and the video duration, to acquire the optimal settings for HR measurement. It is more robust and convenient than those of the current commercial devices and webcam-based HR measurement approaches, especially in complex illumination environments, such as wearing hat, glasses or makeup, with measurement accuracy 95%. Field experiments are also performed for the prognosis of physical state of the driver.

INDEX TERMS Driving state prognosis, empirical mode decomposition, heart rate measurement, near-infrared video, non-invasive measurement.

I. INTRODUCTION

Heart rate (HR) monitoring in the healthcare aims to equip providers with meaningful data to monitor the health, prognose symptom and improve the driving safety for the transportation [1]. It is a hot topic for accidents prevention via monitoring the driver's physical condition. However, physiological data collection is a long-term, continuous process that poses various challenges. Human vital signs (i.e., HR, blood pressure, respiration rate and blood oxygen) constitute an essential part in healthcare, where the fluctuation in these data can be regarded as indicators for the health status of the patients. Conventionally, these vital signs are monitored via contact devices. For instance, traditional methods for the HR measurement includes attaching multiple sensing electrodes to the skin to obtain Electrocardiograph (ECG) signals [2]–[4] and blood oxygen saturation measurement via a finger sensor, which is inconvenient to deploy real-time data collection or monitoring continuously. Although some researchers use an electrode integrated into the driving seat to obtain the ECG signal during driving [2], [5], [6], these methods are uncomfortable for the drivers.

In some studies, several non-contact techniques have been proposed to estimate HR of drivers. Lee *et al.* [7] used the continuous-wave Doppler radar to detect the HR of the driver. Bounyong *et al.* [8] used a microwave sensor and template matching algorithm to monitor the HR of the driver. It is well known that a webcam can be used for non-contact measurement of HR and respiration rate via image processing [9]–[12]. The involved technique is based on the photoplethysmograph (PPG), which can detect the blood volume variation by analyzing the transmitted or reacted light [13]. The PPG is a pulsatile physiological waveform attributed to cardiac synchronous variations in the blood volume for each heartbeat which is superimposed on a slowly varying baseline with lower frequency components attributed to respiration, sympathetic nervous system activity and thermoregulation.

Thus using a visual sensor such as a camera to measure the HR is a feasible approach.

A camera-based HR detection is as follows. First, facial video is collected over a certain period. Then the facial images are separated into the three RGB channels, and the resulting two-dimensional images are transformed into one-dimensional time-series signals. Each channel signal is normalized, detrended and processed using the independent component analysis (ICA). Afterwards the signals are filtered via bandpass filtering and fast Fourier transformation (FFT) to extract the HR. Further details can be found in [14]. Monkaresi *et al.* [15] used a machine learning method based on ICA to improve the accuracy of the contactless HR monitoring. Tsouri and Li [16] investigated the accuracy of pulse rate estimation based on single channel processing in alternative color spaces using standard RGB camera.

Up to now, there is few research to investigate the non-invasive HR measurement via the near-infrared light. PPG technology is quite popular, and specific devices should be worn by the participants which is not convenient and portable. Therefore, with the aid of the image processing techniques, a non-contact HR measurement method could be explored.

The accuracy of the HR measurement depends on the quality of the facial recognition, which could be easily influenced by the light intensity since webcams are sensitive to the illumination. Therefore, acquiring a clear image is the key step for accurate HR obtainment. Furthermore, the involved environmental factors, i.e., the illumination intensity, light direction and light source, are quite complex which have serious impact on the facial detection performance, hence the accuracy of the HR measurement using a webcam is rather limited. On the other hand, an infrared camera is only sensitive to the infrared light, and it could be used for the HR measurement which is highly robust against various types and intensities of the illumination. There are three types of the infrared light: near, mid and far. Here, the near-infrared camera images are used for the HR measurement, which have strong anti-jamming capabilities and they are virtually unaffected by the disturbances, i.e., facial makeup and camouflage. In this paper, the near-infrared facial images are used for the drivers' HR acquirement.

The contribution of the paper is summarized as: (1) A non-invasive HR measurement method is proposed based on near-infrared facial images for online physiological state monitoring. The captured images are processed via Kanade-Lucas-Tomasi algorithm, EMD, filter and FFT operation to acquire the accurate HR value; (2) The impact factors involved in the HR measurement are discussed from various aspects, such as illumination, distance between the near-infrared camera, and physical status of the participants to obtain the optimal settings; (3) The effectiveness of the HR measurement is evaluated from the facial images with disturbances such as makeup, wearing glasses or hat; (4) Field tests are performed when several participants are driving commercial vehicle to verify the feasibility and accuracy of the proposed HR measurement method. It is proved that the non-invasive HR measurement could be used as a indictor to monitor the physical status of the participants continuously.

The remainder of the paper is organized as follows. Section I explains the detailed procedure of the HR measurement using a near-infrared camera. The related impact factors for HR measurement are discussed & analyzed and experiments in various scenarios are performed to verify the effectiveness of the proposed method along with the field tests in Section II. Conclusions are given in Section III.

II. THE PROCEDURE OF THE HR MEASUREMENT

Human skin consists of three primary layers, the epidermis (nearly $100\mu m$), the dermis (nearly $1-4mm$) and the hypodermis (nearly $1-6mm$). When a beam of the light shines on the skin tissue, some of the light are reflected by the epidermis, some pass through the epidermis and dermis and some are absorbed by the hypodermis. The penetration of the light in the human skin tissue is illustrated in Fig. 1.

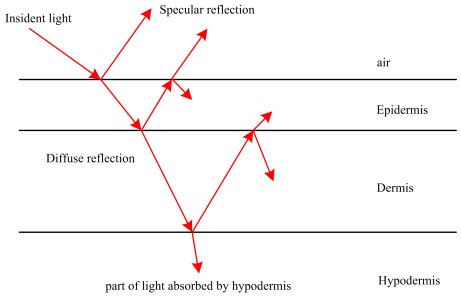


FIGURE 1. Penetration of the light in the human skin tissue.

A. THE REGION OF INTEREST SELECTION

The extent of the light penetration into human tissues is dependent on the wavelength of the light, where the longer wavelength the deeper skin penetration [17]. In order to clearly obtain the blood volume variation, the near-infrared camera is selected to collect the facial images, since the wavelength of the near-infrared light is greater than that of the visible light. It should be mentioned that the human facial blood vessels have complex patterns and human facial skin color varies slightly with the blood circulation. Therefore, the captured image of a human face via a near-infrared camera can be analyzed with the region of interest (ROI) selection for HR detection.

There are many methods for tracking human faces [18], [19]. In this study, the Computer Vision System ToolboxTM of MATLAB is used for facial detection and tracking in order to obtain the coordinates of the facial position. First, a cascade detector, which is a type of facial detector, is designed to locate a face in the first frame of a video. Then the Kanade-Lucas-Tomasi (KLT) algorithm is applied to track the facial image in the subsequent video frames. Further details on the facial detection using the KLT algorithm can be referred [20]. The KLT algorithm can supply the coordinates of the facial image with the defined width and

height of the box around the captured face. The central 80% of such a box is selected as the ROI for subsequent calculation.

FIGURE 2. The captured near-infrared image.

B. SIGNAL TRANSFORMATION

The images in the ROI can be separated into the three RGB channels, and the collected near-infrared image of the face is illustrated in Fig. 2. Since the visible wavelengths have been eliminated in each channel, the images in the red, green and blue channels appear as gray-scale images. Here, the green channel images are used for further analysis.

The average image value of all the pixels in the green channel of the ROI can be calculated as,

$$V_i = \frac{\sum_1^n \sum_1^m I_i}{nm} \tag{1}$$

where V_i is the average value of all the pixels of the i^{th} frame in the ROI; n and m denote the row and column of the pixels, respectively. I_i represents the pixels of each frame and V_i can be normalized as,

$$X_i(t) = \frac{V_i(t) - u}{\partial} \tag{2}$$

where $X_i(t)$ is the normalized signal; u and ∂ are the mean and standard deviation of the $V_i(t)$, respectively.

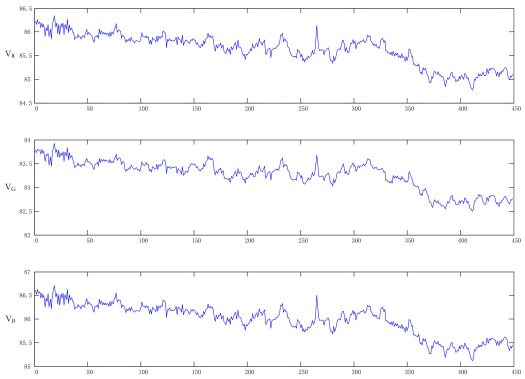


FIGURE 3. The average pixel value of each channel in 30s.

The average pixel values of each channel over 30s are shown in Fig. 3, and the signals of each channel in the frequency domain after FFT operation are depicted in Fig. 4. The HR of an average relax adult is 50-95 beat/min, corresponding to the frequency range 0.83-1.58 Hz. From the

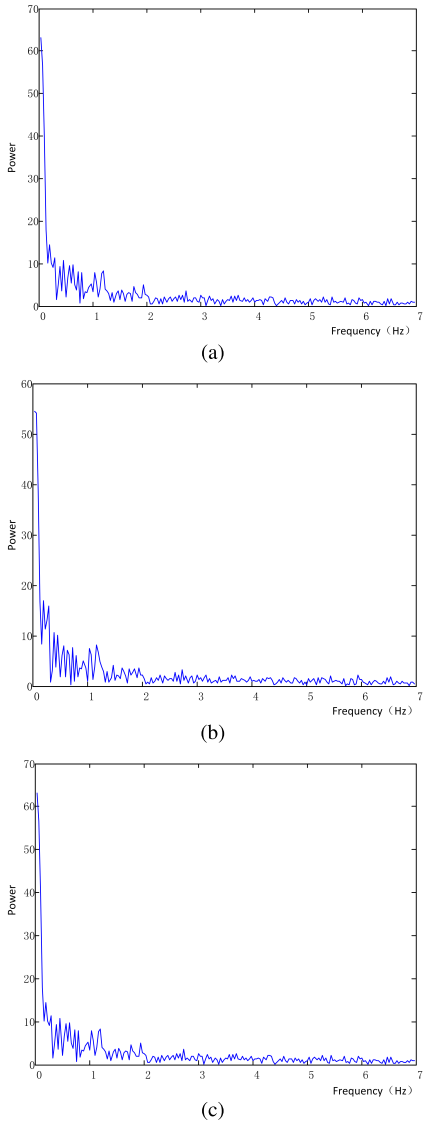


FIGURE 4. Each channel signal after FFT in the frequency domain. (a) R channel signal in the frequency domain. (b) G channel signal in the frequency domain. (c) B channel signal in the frequency domain.

obtained spectrum in Fig. 4, it can be observed that the HR signal is corrupted by the noise. So useful HR information has to be extracted from the original signals.

1) EMPIRICAL MODE DECOMPOSITION

Poh *et al.* [14] used an ICA method to denoise the signal. Here the Empirical mode decomposition (EMD) is used for the signal preprocessing, which can not only eliminate the signal trend item but also filter out the high-frequency noise.

EMD, proposed by Huang *et al.* [21], is an adaptive multi-component signal decomposition method that can yield intrinsic mode functions (IMFs). These IMFs should satisfy two conditions: 1) The number of the extrema and the zero-crossings in the entire data sequence must equal or differ at most by one; 2) At any instant, the mean of the envelopes

of the local maxima and minima must be zero. The first condition is similar to the narrow-band requirement for a stationary Gaussian process, which can ensure that the local maxima and minima of the data series are always positive and negative, respectively. The second condition is to transfer a global requirement into a local one. The EMD algorithm can be summarized as follows [21]:

Step 1: To identify all the extrema (maxima and minima) of the signal, $x(t)$;

Step 2: To apply cube spline interpolation to generate the upper and lower envelopes $e_{min}(t)$ and $e_{max}(t)$, respectively;

Step 3: To compute the average value as, $m(t) = (e_{min}(t) + e_{max}(t))/2$;

Step 4: To extract the detail, $d(t) = x(t) - m(t)$;

Step 5: To iterate over the residue $r(t)$.

In practice, the above procedure has to be refined by a sifting process, which occurs after the first iteration of step1-step4 by refining the signal $d(t)$ with zero-mean according to a termination criterion SD (standard deviation, computed from the two consecutive sifting results),

$$SD = \sum_{t=1}^{N} \left[\frac{\left| d_{j(k-1)}(t) - d_{jk}(t) \right|^2}{d^2_{j(k-1)}(t)} \right] \leq \xi \tag{3}$$

where N is the total data length and ξ is the threshold, here, $\xi = 0.25$; d_{jk} is the k^{th} sifting result of the j^{th} mode, considered as the j^{th} effective mode. The corresponding residue $r_j(t)$ can be computed applied by the above step 5.

Thus the input signal $x(t)$ is decomposed into p IMFs,

$$x(t) = \sum_{j=1}^{p} c_j(t) + r_p(t) \tag{4}$$

where $r_p(t)$ denotes a residue and $\{c_j(t), j = 1, 2, ...p\}$ denote p IMFs. The EMD decomposition result of the G channel signal is demonstrated in Fig. 5. The signal was decomposed into 7 IMF components and a residue based on the applied scenario and human experience, where the frequencies in each IMF are arranged in descending order. Thus the first IMF component is in the highest frequency band and the residue is in the lowest frequency band.

Fig. 6 shows the G channel signal and the residual signal. The blue line represents the G channel signal, and the green line represents the residual signal after EMD decomposition. The residual signal is the trend item, which is the last IMF component in the lowest frequency band. The first IMF component is in the highest frequency band, which is the high-frequency noise that should be eliminated.

2) SIGNAL RECONSTRUCTION

After the IMF components obtainment, the signal can be reconstructed as,

$$x_r(t) = \sum_{i=m}^{n} IMF[i] \tag{5}$$

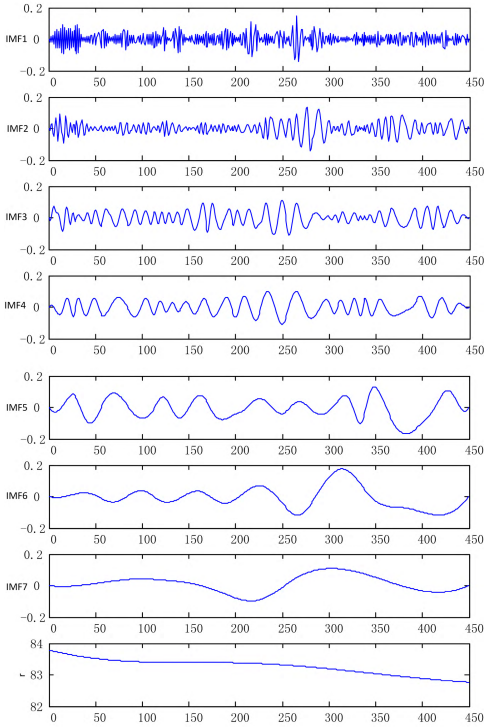


FIGURE 5. EMD decomposition results of the G channel signal.

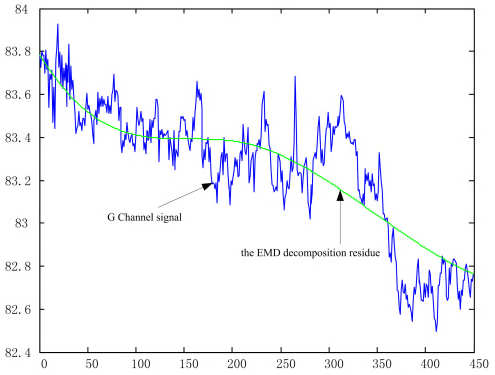


FIGURE 6. G channel signal and the EMD decomposition residue.

where $x_r(t)$ is the reconstructed signal, i is the index of the IMF components, and m and n are the first and last selected indices of the IMF components, respectively.

For instance, there are eight IMF components in Fig. 5. Six IMF components are selected to reconstruct the signal, with $m = 2$ and $n = 7$. Fig. 7 shows the G channel signal and the result of the reconstruction. It can be seen that the reconstructed signal is denoised and detrended, while the reconstructed signal is filtered using a bandpass filter afterward.

3) EXTRACTION OF HR

After the signal has been processed using EMD, the frequency computation may be occasionally affected by the

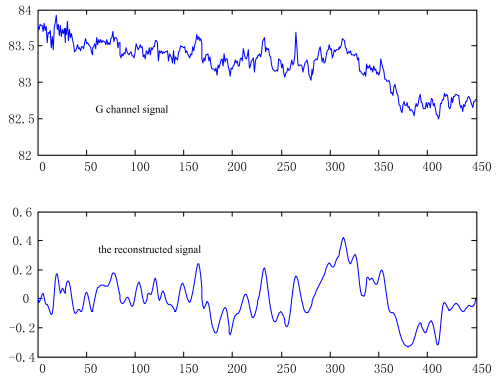


FIGURE 7. G channel signal and the reconstructed signal by EMD.

noise. Therefore, the processed signal is filtered using a band-pass filter (128-point Hamming window; 0.8-3.2 Hz; HR, 48-192). Finally, FFT (in MATLAB with the function 'fft(S,N)') is applied to the filtered signals to obtain the HR value,

$$HR = 60 \times f_h \tag{6}$$

where f_h is the extracted HR frequency; S and N parameters in 'fft' function are the dealt IMF components and the points, respectively; here, N=449. The extracted HR frequency with FFT is shown in Fig. 8.

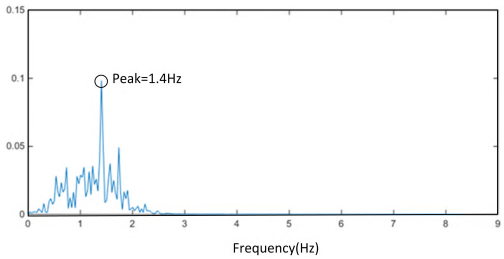


FIGURE 8. Result of the FFT after bandpass filter.

During the signal processing, a large measurement error would be generated if the FFT operation is performed directly without noise processing. Hence, the noise should be removed by the filter first. Only a small portion of the facial image data is intercepted for processing, which could guarantee the real-time requirement for HR measurement. Detailed information for the real-time data processing can be referred in [22].

In summary, the schematic diagram of the non-invasive HR measurement via the near-infrared images is illustrated in Fig. 9. First, the facial video images are collected when a participant sits directly in front of the near-infrared camera in certain distance and keeps the pose as still as possible for 30s, which is similar to the driving state of the driver. Then the facial portion is determined and recognized during the signal preprocessing step. The identified facial images are separated into three channels, where each channel image is transformed into a serial of one-dimensional signal and

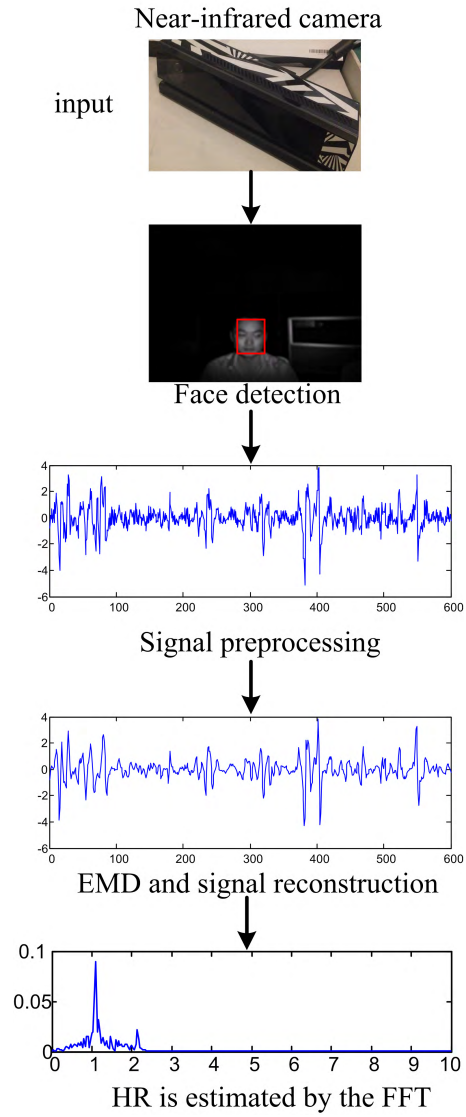


FIGURE 9. Schematic diagram of the HR detection using a near-infrared camera.

processed via EMD. Next, the bandpass filtering and FFT are applied to extract the useful signal and HR value.

There would be more constraints if the HR measurement is obtained with only webcam, such as the normal daylight illumination [22]. It is noted that the illumination variation has heavy influence on the measurement from the color camera. Considering the night situation, therefore, the near-infrared camera is used to avoid the illumination impact on the HR measurement.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL SETTINGS

In order to verify the feasibility of the proposed algorithm, the experimental platform is setup in a normal indoor environment. A Kinect 2.0, i.e., a motion sensing input device developed by Microsoft, can simultaneously obtain the color

video and near-infrared video images used for the experiments. It can guarantee that both types of the facial video images are collected under the same conditions. During the experiments, Kinect 2.0 is fixed right in front of the participant and the distance between the participant and the Kinect 2.0 is approximately 60cm. The participant is asked to breath spontaneously, face the Kinect 2.0 for 30 seconds, and remain as still as possible.

A RGB image and the corresponding near-infrared image obtained using the Kinect 2.0 simultaneously with sampling frame rate 15 fps are shown in Fig. 10. The pixel resolution of the RGB video image is 1920 × 1080, and the near-infrared video image is 512 × 424. Both images are saved as AVI format in a computer. A commercial finger pulse oximeter (YUYUE, YX301) is applied for the HR measurement of the participant at the same time for comparison. The KLT algorithm is used for facial recognition and tracking in order to localize and extract the face position from the collected video images. Then the facial images are separated into the three RGB components. Each component is calculated and processed using EMD, bandpass filtering and FFT to obtain the HR of the participants.

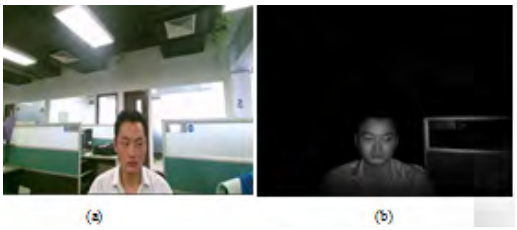


FIGURE 10. Facial video data acquisition: (a) RGB image; (b) near-infrared image.

B. THE EFFECT OF THE KINECT SETTING

The mean value and the root-mean-squared error (RMSE) are used to evaluate the HR measurement performance,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{8}$$

where x_i is the HR value and $\bar{x}$ is the mean value. In order to examine the effect of the illumination intensity, recording time interval and distance between the camera and the participant on the HR measurement, a series of experiments are performed using the proposed method.

1) OPTIMAL MEASUREMENT DISTANCE

In the first group of the experiments, the distance is varied between the face of the participant and Kinect 2.0. The luminous intensity is 40 lx and the video recording duration is 30s. The RMSEs of the HR of different distances are illustrated in Table 1. The best performance can be achieved when the distance between the participant and the camera is ranged between 55cm and 70cm. When the distance is less than 50cm or more than 70cm, the near-infrared light reflection would become too strong or too weak, which could adversely affect the performance. Here, 60cm is selected as the optimal measurement distance.

TABLE 1. The effect of the distance on the HR measurement.

Distance (cm)	40	45	50	55	60	65	70	75	80
RMSE	9.42	5.98	3.87	3.57	2.08	2.78	2.92	3.02	4.98

2) OPTIMAL MEASUREMENT DURATION

In this case, the participant is asked to sit in front of the Kinect 2.0 at a distance of 60cm and remain still while the recording time varies. The other settings remain the same. The purpose of the experiments is to determine the optimal recording time, which would be sufficiently short to reduce the image processing time and provide optional condition for an accurate and efficient system. The results are shown in Table 2, and it can be seen that the proposed method shows the best performance when the video duration is close to 30s. Furthermore, the duration could be shortened to 10s if the RMSE is limited around 4s, which is still within the acceptable level.

TABLE 2. The determination of the optimal measurement duration.

Record time (s)	5	10	15	20	25	30	35	40	45	50
RMSE	6.12	3.87	3.21	3.75	3.92	2.75	2.87	2.72	3.12	2.78

TABLE 3. The effect of the illumination intensity.

Illumination intensity	RMSE	
	Webcam video	Near-infrared video
2	6.12	2.51
10	3.22	2.53
20	2.51	2.51
30	2.51	2.52
40	2.50	2.50
50	2.52	2.56
60	2.51	2.53
70	2.50	2.50
80	2.53	2.51

3) THE EFFECT OF THE LUMINOSITY

During the experiments, the participant is asked to sit in front of the Kinect 2.0 at 60cm distance and remain still while the luminous intensity is varied from 0 lx to 80 lx. The other settings remain the same. The RMSEs of the test results at various luminous intensities are shown in Table 3, where the luminosity is detected via a digital light meter (TASI TA8123) and adjusted using a LED lamp. It can be seen that the HR measurement based on both color video and near-infrared video has higher performance when the luminosity is less than 10 lx. When it is lower than 20 lx, the webcam video

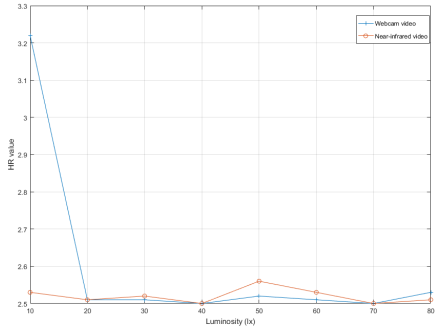


FIGURE 11. The comparison of the HR under different illumination intensities.

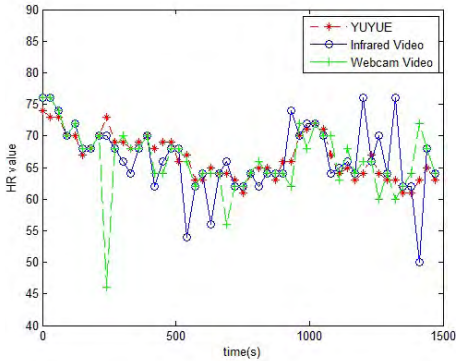


FIGURE 12. HR detection results over a continuous 1500s period (in relaxed situation).

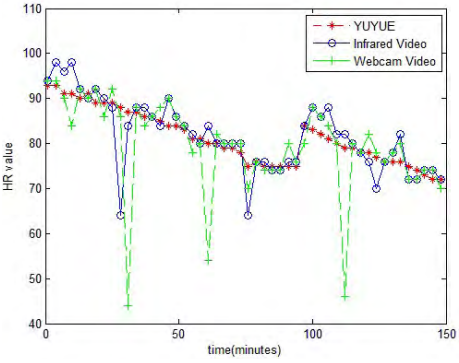


FIGURE 13. HR measurement after a continuous 150 minutes exercise period.

quality would be degraded substantially whereas the near-infrared video quality remains relatively the same. Then the color camera could not be used for the HR measurement under dark environment, which reflects the advantage of the near-infrared camera. The HR results under statistical tests when the luminosity is above 10 lx are shown in Fig. 11. Here, 40 lx is selected which is close to the normal driving situation.

From the experiments described above, it is shown that many factors could influence the HR measurement accuracy. In order to obtain accurate HR measurement using the proposed method, it is necessary to consider the distance, illumination intensity and recording time three involved main factors.

C. THE EFFECT OF THE PHYSICAL CONDITION OF THE TESTER ON THE HR MEASUREMENT

A series of experiments are performed in various scenarios with participants in different physiological states. The participants include males and females from different age groups, where some participants are in relaxed situation, while others have just finished exercise strenuously. The case of multiple participants in the same video is also considered. All the participants are tested while they sit in front of the Kinect 2.0 with an oximeter on one finger. Three sets of HR data are obtained for each test: RGB images, the near-infrared images and oximeter readings.

1) HR MEASUREMENT IN RELAXED SITUATION

First, the HR is measured continuously while the participant is in the relaxed state, and the HR results over a continuous period of 1500s is depicted in Fig. 12, where the sampling interval is 30s. The results of the proposed method are found to be consistent with those obtained using the YX301, with RMSE 2.23. Under the same condition, the near-infrared and RGB videos yielded nearly identical results using the same algorithm. It can be seen that the HR measurement with the proposed algorithm via the near-infrared video are more accurate than those obtained using the YX301. The measurements points with larger errors are caused by the face chattering during the measurement.

2) HR MEASUREMENT OF THE PARTICIPANT AFTER STRENUOUS EXERCISE

HR measurements can be obtained after the participants engaged in strenuous exercise over a continuous period of 150 min. Three minutes video is recorded for the tester and 30s video is intercepted for the HR measurement, and the resultant HR values are shown in Fig. 13. The results of the proposed method are found to be roughly the same as those obtained via the YX301, with RMSE 2.77. The measurement values with larger errors are caused by the face chattering during the measurement. Under the same condition, the HR measurement with the near-infrared video yielded better performance than that with the RGB video.

3) HR MEASUREMENT WITH MULTIPLE PARTICIPANTS

50 participants (including males and females from different age groups) are invited for the HR measurement using the proposed method. The HR measurement results of 20 of these participants (15 males and 5 females; 20-40 years old) are shown in Fig. 14, where the horizontal axis indicates the number of the selected participants. The RMSE between the oximeter and the proposed algorithm is approximately 2.17. Further, small fluctuations can be observed in the three sets of HR measurements and the measurement values with larger errors are caused by the face chattering during the

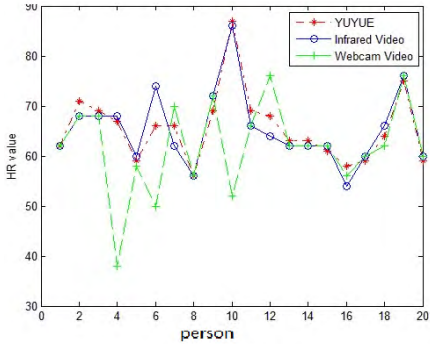


FIGURE 14. Results of the HR detection for multiple participants.

measurement. The main reason for these fluctuations could be that the heads of the participant unconsciously tend to move slightly throughout the experiments.

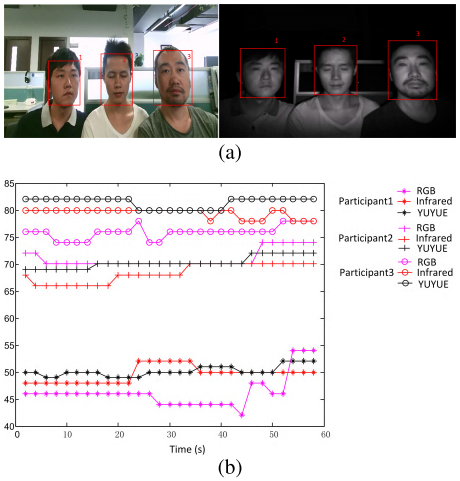


FIGURE 15. EMD decomposition results of the G channel signal. (a) The simultaneous collection of the RGB and near-infrared videos. (b) HR measurements from simultaneous processing of multiple participants in the same scene.

4) HR MEASUREMENT WITH MULTIPLE PARTICIPANTS IN THE SAME SCENE

In this case, videos of three participants within the same scene were recorded to demonstrate the capability of the proposed method to perform simultaneous HR measurements of multi-participants. The RGB images (left) and the resulted near-infrared (right) images are illustrated in Fig. 15(a). Fig. 15(b) shows the resultant HR curves obtained using the proposed method (*, participant 1; +, participant 2; o, participant 3), and the black, red and pink lines represent the HR measurements extracted from the YX301, the near-infrared video and the RGB video, respectively. It can be seen that the HR curve acquired via the proposed method with the near-infrared images is closer to that from the YX301 throughout the experiments. It should be noted that the main reason to perform this group of tests is to identify the feasibility of the proposed HR measurement method, which would not be used to monitor the multi-driver at the same time.

5) HR MEASUREMENT CONSIDERING VEHICLE CHATTERING AND FACIAL CAMOUFLAGE

There would be slight vehicle vibration during the vehicle driving, and the facial expression of the driver would be changed slightly correspondingly which could bring certain impact on the measurement. In order to remove the impact and improve the measurement precision, the obtained signals can be segmented in pieces to calculate the standard deviation. Then the signals would be reconfigured at the locations with obvious deviation. The detailed calculation can be referred in [23].

Besides, in practical driving, the drivers would wear makeup, hat or glasses, especially for female drivers, which may bring certain disturbance to the HR calculation. A series of experiments have been conducted under these conditions. First, facial videos are recorded with three female participants after makeup, and 8 videos were recorded from each tester with 30seconds while the YX301 is used to acquire HR of the tester simultaneously. The recorded videos are processed for HR measurement and the obtained results with Bland-Altman graph are shown in Fig. 16. It can be seen that the measurement values are all fallen inside the 95% bound.

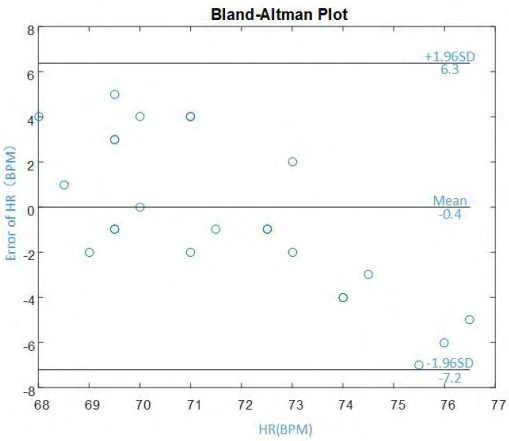


FIGURE 16. HR measurements when the participants are wearing makeup.

It should be noted from the dermatologist that the body lotion or facial cream would increase the degree of the light penetration of the faces. Only the makeup power would have negative effect to reduce the light penetration degree. Therefore, normal makeup on the face has little or positive effect on the HR measurement.

Next, a group of the experiments are performed when the participants are wearing hats while each tester is recorded with 8 videos lasting 30s. Here, it is assumed that the hats would not interfere the driving activities, so the brim of the hat would not hide the face, i.e., tennis hat. Similarly, the HRs of the participants are measured by the YX301 simultaneously. The recorded videos are processed for HR measurement and the obtained results with Bland-Altman graph are shown

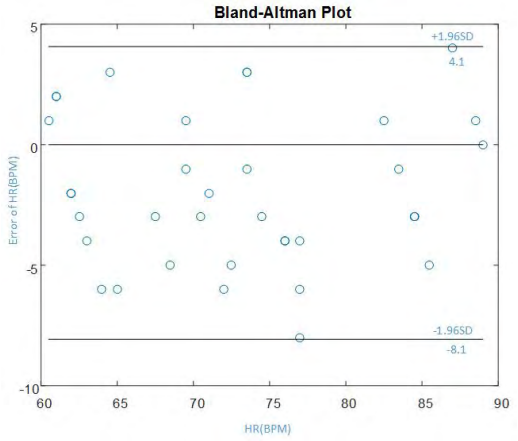


FIGURE 17. HR measurement when the participants are wearing hats.

in Fig. 17. It can be seen that the measurement values are all fallen inside the 95% bound.

Other group of the experiments is performed when the participants are wearing glasses while each tester is recorded with 8 videos lasting 30s. Once again, the HRs of the participants are measured by the YX301 simultaneously. Fig. 18 shows the Bland-Altman graph of the obtained HR measurements. It can be seen that the measurement values are all fallen inside the 95% bound but with obvious deviation. So the obtained HR measurement when the driver is wearing glasses can still be used as the symptom indictor.

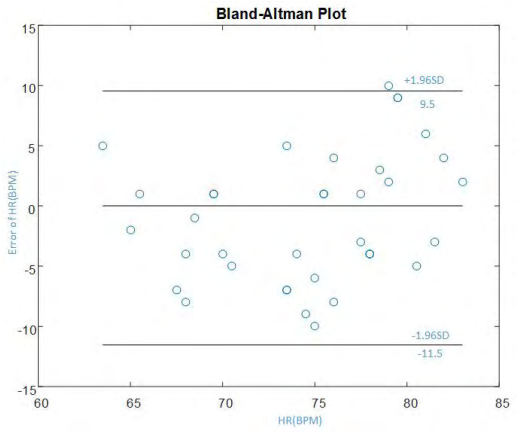


FIGURE 18. HR measurement when the participants are wearing glasses.

D. COMPARISON EXPERIMENTS

Bland-Altman plots and coefficient correlation analysis are used for the performance comparison. In these experiments, the dataset consists of 100 video fragments recorded from 100 participants, including 90 males and 10 females with ages ranging from 20 to 55 years old. The experiments are performed at the optimal measuring distance and video duration. The Bland-Altman plots obtained from the proposed method is depicted in Fig. 19. It is quite clear that almost all the

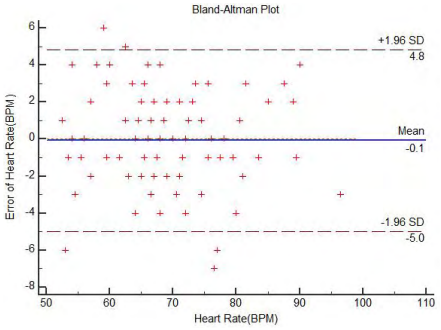


FIGURE 19. Bland-Altman Plot for the HR.

measured results fall into 95% tolerance interval (±1.96 SD).

Furthermore, the HR measurement obtained from the proposed method and the commercial finger pressure oximeter (YUYUE, YX301) are used for comparison. 100 experiments are performed and the test results are shown in Fig. 20. The RMSE error is 2.48 while the correlation coefficient is 0.96.

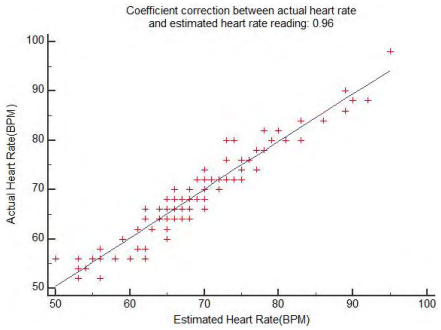


FIGURE 20. The obtained HR with the proposed method and the oximeter.

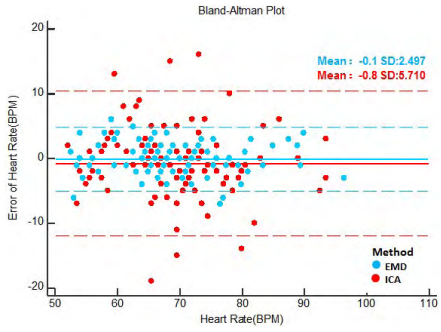


FIGURE 21. The Bland-Altman plots obtained by ICA (red) and our proposed approach (blue).

The ICA method is used for comparison to the proposed method with the infrared video [14]. The Bland-Alterman plots of the HR tests are depicted in Fig. 21. The solid lines are the mean bias of the HR measured by the YX301 and evaluated from the test dataset, and the dash lines are the tolerance interval. The Bland-Altman plot with the red line is the method with ICA and the blue line is with the proposed method. It is seen that nearly all the measured results with

the proposed method fall into the 95% tolerance interval (SD=2.497) and the measured results using ICA fall into 90% tolerance interval (SD=5.71).

Three parameters are selected for the performance comparison, the standard deviation error (SDE), the correlation coefficient (Pcc) and time consumption (Tc), listed in Table 4. The SDE of the proposed method is about 2.49, which is smaller than that with ICA. The correlation coefficient and completion time of the proposed method is 0.96 and 0.2575 second, which is larger and faster than that with ICA method. Therefore, the HR measurement with the proposed method is more robust than that with ICA method.

TABLE 4. The performance comparison of the proposed method and ICA.

Method	SDE	Pcc	TC (second)
ICA	5.71	0.81	0.3397
Ours	2.49	0.96	0.2575

E. FIELD EXPERIMENTS

The tachograph (see Fig. 22(a)) is installed inside the vehicle for facial images collection, and the distance is kept roughly 60cm from the driver, which is used to record the facial video of the driver during driving activity. In the mean time, the HR bracelet is worn by the driver to record his/her heart rate, shown in Fig. 22(b). The sampling of the video is 30 fps, and the pixels of the images are 640 × 480.

(a) (b)

FIGURE 22. The field test setting. (a) The installed tachograph. (b) The field test scene.

These field experiments are performed by the participants (students and colleagues) when they are in normal healthy medical conditions under the same environmental settings, i.e., location, period of time, weather etc. The tests are performed during the morning 10am and on the road behind our institute, quite similar to the standard test environment. The reason for the field test is to testify the effectiveness of the proposed method of the online HR measurement. During the tests, the captured images are transferred to the laptop via wireless communication network for real-time processing. If the proposed method is applied in practice, an onboard micro-chip could be used for the data processing, which is easily to be engineered.

In this group of the field experiments, the video of 20 drivers during driving in different periods of time are

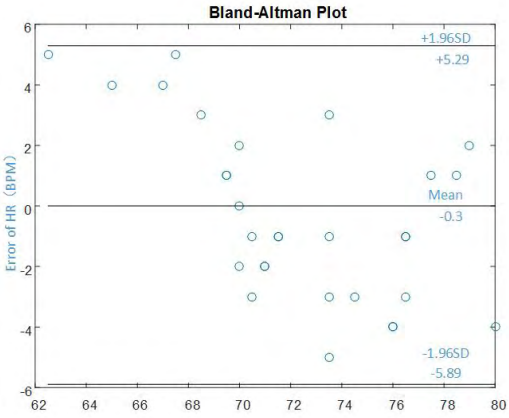


FIGURE 23. The HR measurements results in the field test.

TABLE 5. The HR results of the field test.

Tester	Daylight					
	Proposed	Detector	Proposed	Detector	Proposed	Detector
1	72	75	71	78	73	74
2	71	74	70	72	72	69
3	70	74	70	65	72	75
4	74	78	76	77	76	77
5	69	71	70	69	69	72
Tester	Night					
	Proposed	Detector	Proposed	Detector	Proposed	Detector
1	73	76	71	72	74	78
2	69	65	65	60	67	62
3	75	79	80	78	79	78
4	78	77	76	72	78	82
5	70	70	70	67	70	69

selected for the HR measurement. The HR Bland-Alterman plots from the proposed method and the YX301 for the selected 5 participants is illustrated in Fig. 23 and summarized in Table 5. Since the heads of the drivers could not be kept fully still during driving, part of the video with heads kept relatively still are extracted to verify the effectiveness of the proposed algorithm. Although there is rather difference of the HR measurement from Tester1, the results from other four participants are quite similar. It could be used to demonstrate the feasibility of monitoring drivers' physiological parameters via the near-infrared camera in real-time.

It should be noted that the HR measurement tests are performed when the participants are in normal physical conditions. Other conditions, i.e., sleepiness, fatigue are not considered here. Further research could be carried out considering more complex physical scenarios.

In summary, with the proposed method for signal processing, it is promising to use facial video information to detect physiological parameters of the human body. Thus the EMD is applied to remove the noise of the signal for HR extraction. The experimental results presented in this paper can verify the efficacy of the proposed method. However, the measurement results could be affected by the motion artifacts, hence, the facial videos of the drivers with relatively small movements, i.e., looking up/down, nodding the head should be avoided for the HR measurement. Field experiments are also performed to measure the HR of the drivers during driving activities with relatively acceptable satisfaction.

IV. CONCLUSIONS

Based on the near-infrared video image processing and empirical mode decomposition technology, a non-invasive HR detection method is proposed in this paper. The captured facial images are processed and useful signals are extracted for HR calculation. The effects of the luminosity, video recording duration, and distance between the face of the participant and the camera on the non-invasive HR measurement are discussed in detail. Various experiments have been performed in different sceneries for the HR measurement, including single tester and multi-participants participation. Furthermore, compared with the YX301 instruments and other existing non-invasive measurement methods based on the video analysis, the proposed method is more accurate, convenient and robust, especially in the environments with complex illumination variation. Since real-time HR acquirement of the drivers for continuous physiological status monitoring is expected to have a significant impact on the safety improvement in the transportation, field experiments have also been performed for the HR obtainment of the drivers to provide precaution during driving.

Further research will be carried out to investigate the relationship between the HR status of the drivers and driving for the comprehensive evaluation index of the safety transportation.

ACKNOWLEDGMENT

(Qi Zhang and Yimin Zhou contributed equally to this work.)

REFERENCES

[1] J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G.-Z. Yang, "Big data for health," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1193–1208, Jul. 2015.

[2] T. Wartzek, B. Eilebrecht, J. Lem, H.-J. Lindner, S. Leonhardt, and M. Walter, "ECG on the road: Robust and unobtrusive estimation of heart rate," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 11, pp. 3112–3120, Nov. 2011.

[3] J. Gomez-Clapers and R. Casanella, "A fast and easy-to-use ecg acquisition and heart rate monitoring system using a wireless steering wheel," *IEEE Sensors J.*, vol. 12, no. 3, pp. 610–616, Mar. 2012.

[4] A. D. Hellicar *et al.*, "An algorithm for the automatic analysis of signals from an oyster heart rate sensor," *IEEE Sensors J.*, vol. 15, no. 8, pp. 4480–4487, Aug. 2015.

[5] H. Wang, C. Zhang, S. Ma, T. Shi, and F. Wang, "Real-time EEG-based detection of fatigue driving danger for accident prediction," *Int. J. Neural Syst.*, vol. 25, no. 2, p. 1550002, 2015.

[6] B.-S. Lin, W. Chou, H.-Y. Wang, Y.-J. Huang, and J.-S. Pan, "Development of novel non-contact electrodes for mobile electrocardiogram monitoring system," *IEEE J. Transl. Eng. Health Med.*, vol. 1, Jun. 2013, Art. no. 2700108.

[7] K. Lee, C. Park, and B. Lee, "Tracking driver's heart rate by continuous-wave Doppler radar," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2016, pp. 5417–5420.

[8] S. Bounyong, M. Yoshioka, and J. Ozawa, "Monitoring of a driver's heart rate using a microwave sensor and template-matching algorithm," in *Proc. IEEE Int. Conf. Consum. Electron.*, Jan. 2017, pp. 43–44.

[9] C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image," *Med. Eng. Phys.*, vol. 29, no. 8, pp. 853–857, 2007.

[10] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a Webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.

[11] G. R. Tsouri, S. Kyal, S. Dianat, and L. K. Mestha, "Constrained independent component analysis approach to nonobtrusive pulse rate measurements," *J. Biomed. Opt*, vol. 17, no. 7, p. 077011, 2012.

[12] Q. Zhang, G.-Q. Xu, M. Wang, Y. Zhou, and W. Feng, "Webcam based non-contact real-time monitoring for the physiological parameters of drivers," in *Proc. 4th Annu. Int. Conf. Cyber Technol. Automat., Control, Intell. Syst.*, Jun. 2014, pp. 648–652.

[13] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas*, vol. 28, no. 3, pp. R1–R39, 2007.

[14] 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*, vol. 18, no. 10, pp. 10762–10774, 2010.

[15] H. Monkaresi, R. A. Calvo, and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a Webcam," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1153–1160, Jul. 2014.

[16] G. R. Tsouri and Z. Li, "On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras," *J. Biomed. Opt*, vol. 20, no. 4, p. 048002, 2015.

[17] M. C. Moreira, R. N. do Prado, and A. Campos, "Application of high-brightness LEDs in tissue human and their therapeutic interaction," in *Proc. Ind. Appl. Soc. Annu. Meeting*, Oct. 2009, pp. 1–6.

[18] G. R. Bradski, "Real time face and object tracking as a component of a perceptual user interface," in *Proc. 4th IEEE Workshop Appl. Comput. Vis.*, Oct. 1998, pp. 214–219.

[19] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE CVPR*, Dec. 2001, pp. I-511–I-518.

[20] *Detection and Tracking Using the KLT Algorithm*. Accessed: 2010. [Online]. Available: https://ww2.mathworks.cn/help/vision/examples/face-detection-and-tracking-using-the-klt-algorithm.html

[21] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[22] Q. Zhang, Q. Wu, Y. Zhou, X. Wu, Y. Ou, and H. Zhou, "Webcam-based, non-contact, real-time measurement for the physiological parameters of drivers," *Measurement*, vol. 100, pp. 311–321, Mar. 2017.

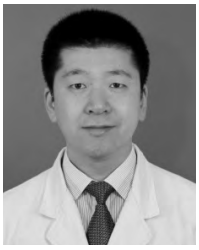
[23] X. Li, J. Chen, G. Zhao, and M. Pietikäinen, "Remote heart rate measurement from face videos under realistic situations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 4264–4271.

QI ZHANG received the B.S. and M.S. degrees from the School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2002 and 2004, respectively, and the Ph.D. degree from the School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai, China, in 2008. He is currently an Associate Professor with the School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, China. His research interests include computer vision and microrobot control.

YIMIN ZHOU received the Ph.D. degree from the University of Oxford, U.K., in 2008. She is currently an Associate Professor with the Shenzhen Institutes of Advanced Technology, Chinese Academy Sciences, China. Her research interests include nonlinear control, fault diagnosis, neuro-fuzzy modeling, machine learning, and energy management.

SHUANG SONG received the B.S. degree in computer science and technology from North Power Electric University, the M.S. degree in computer architecture from the Chinese Academy of Sciences, and the Ph.D. degree in computer application technology from the University of Chinese Academy of Sciences, China, in 2007, 2010, and 2013, respectively. He is currently an Assistant Professor with the Harbin Institute of Technology, Shenzhen Graduate School, China. His main research interests include magnetic tracking and actuation for bioengineering applications, such as surgical robots, micro manipulation, and so on.

GUOYUAN LIANG received the Ph.D. degree from the School of Information Technology, Peking University, Beijing, China, in 2005. He was a Post-Doctoral Fellow and Research Scientist with the Department of Electrical and Computer Engineering, University of Iowa, USA, from 2007 to 2011. He is currently an Associate Professor with the Center for Intelligent and Biomimetic systems, Shenzhen Institutes of Advanced Technology, Shenzhen, China. His current research interests include computer vision and pattern recognition.



HAIYANG NI received the M.D. degree from the Department of Clinical Medicine, Harbin Medical University, Harbin, China, in 1998, and the M.H.A. degree from Flinders University, Adelaide, Australia, in 2006.

He is currently an Associate Chief Physician of the Dermatology Department, Tianjin Academy of Traditional Chinese Medicine Affiliated Hospital, Tianjin, China. His research interests include dermatopathology and artificial intelligence.