

Block-based adaptive ROI for remote photoplethysmography

Lai-Man Po¹ · Litong Feng¹  · Yuming Li¹ ·
Xuyuan Xu¹ · Terence Chun-Ho Cheung² ·
Kwok-Wai Cheung³

Received: 9 April 2016 / Revised: 25 February 2017 / Accepted: 28 February 2017
© Springer Science+Business Media New York 2017

Abstract Remote photoplethysmography (rPPG) can achieve contactless human vital signs monitoring, but its signal quality is limited by the remote operation nature. In practical applications, improving the rPPG signal quality becomes an essential task. As a remote imaging technique, rPPG utilizes a camera to capture a video of a skin area, especially the facial area, then focuses on a particular sub-area as the region of interest (ROI). In this paper, we investigated a novel adaptive ROI (AROI) approach for improving the rPPG signal quality. In this approach, block-based spatial-temporal division is performed on a captured face video. Based on these segmented video pipelines, the spatial-temporal quality distribution of the rPPG signals is estimated using a signal-to-noise ratio (SNR) feature. Afterwards, AROIs are calculated through mean-shift clustering and adaptive thresholding in SNR maps. As the AROI can be dynamically adjusted according to the spatial-temporal quality distribution of rPPG signals on the face, the quality of the final recovered rPPG signal is improved. The performance of the proposed AROI approach was evaluated with both still and moving subjects. Compared to conventional ROI methods for rPPG, the proposed AROI obtained a higher accuracy in heart rate measurement. And the state-of-the-art motion-resistant rPPG techniques can be effectively enhanced through being integrated with the AROI.

Keywords Mean-shift · Photoplethysmography (PPG) · Region of interest (ROI) · Remote imaging · Thresholding

✉ Litong Feng
lightedfeng@gmail.com

¹ Department of Electronic Engineering, City University of Hong Kong, Hong Kong, SAR, China

² Department of Information Systems, City University of Hong Kong, Hong Kong, SAR, China

³ Department of Computer Science, Chu Hai College of Higher Education, Hong Kong, SAR, China

1 Introduction

Photoplethysmography (PPG) is a low-cost, non-invasive technique to measure the cardiovascular blood volume pulse (BVP). PPG has widespread clinical applications, such as pulse oximeters, vascular diagnostics, and cuff-less blood pressure measurement [1]. The conventional contact PPG (cPPG) needs dedicated light sources and contact probes to detect the reflection/transmission light intensity variation from the tissue due to BVP. Recently, the remote PPG (rPPG) technique has attracted much attention, because it can implement the PPG function using only a digital camera, without contact probes and dedicated light sources. Fig. 1 illustrates the working principles of cPPG and rPPG.

In rPPG, the cardiac pulsation information can be accessed remotely using a digital camera and ambient light [31]. Ambient light is used as the light source for rPPG, and a digital camera acts as a photoelectric converter. In actual use, the digital camera captures videos of skin areas, especially human faces as shown in Fig. 1(b). The principle of rPPG is based on the fact that hemoglobin in blood absorbs light, with the quantum absorbed being proportional to hemoglobin concentration in the blood vessel. Since hemoglobin concentration varies quasi-periodically with BVP, the amount of light absorbed by the skin is modulated by BVP underneath the skin surface with cardiac cycles, appearing as a slight color variation of the skin. Although this slight color variation of the skin due to BVP is invisible to human eyes, it can be detected using a digital camera [33]. To obtain the rPPG signal, a region of interest (ROI) is selected in the video frames to convert the three-dimensional video signal to a one-dimensional PPG signal through spatially averaging pixel values in the ROI frame by frame. An rPPG signal can be recovered through bandpass-filtering the temporal variation of the averaged pixel value in ROI.

Evolution from cPPG to rPPG has brought new challenges. Firstly, the signal-to-noise ratio (SNR) of the estimated rPPG signal from video sequence is much lower than that of the cPPG signal. cPPG utilizes contact probes and dedicated light sources to enhance the PPG signal quality. However, rPPG can only utilize ambient light as the light source, which is treated as a noise source for cPPG [14]. Moreover, the customized monochrome photodetector for cPPG is replaced by the complementary metal-oxide semiconductor (CMOS) sensor in a digital camera for rPPG. During the rPPG measurement, there is a certain distance between the camera and

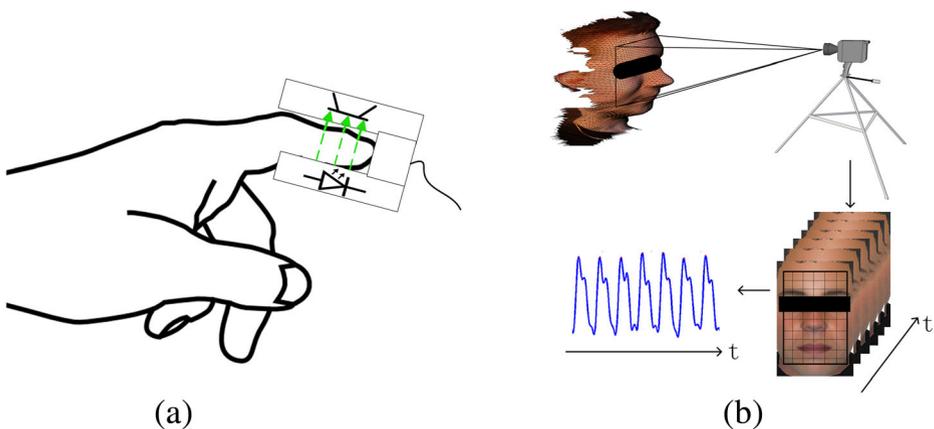


Fig. 1 Illustration of the working principle of PPG. **a** cPPG. **b** rPPG

the subject. All of these changes between cPPG and rPPG increase the difficulty of estimating a high-quality signal for rPPG. Secondly, the motion artifacts problem is much more serious for rPPG as compared to cPPG. Ambient light is not distributed with a spatially homogeneous intensity in daily environments. The CMOS sensor and optics of a digital camera also do not have a uniform response in the whole field of view. When measuring a moving subject, the size and position of the ROI for rPPG need to be adjusted in accordance with the subject's motions. Hence, the radiant flux on the ROI will change, as also the response of the digital camera on the ROI. These changes will distort the rPPG signal, because the skin color variation loading BVP information is very weak as compared with global shading effects. This kind of distortion is called motion artifacts [25]. Low SNR and motion artifacts severely impair the rPPG signal quality. To tackle these problems, several approaches have been proposed to enhance the rPPG signal quality, such as independent component analysis (ICA), wavelet filter, adaptive bandpass filter (ABF), and motion compensation through face tracking [5, 10, 25]. Different facial skin regions provide rPPG signals of different quality, and the quality distribution of rPPG signals on the face is also affected by motion and illumination. If the varying quality distribution of rPPG signals on the face can be estimated, the rPPG signal can be extracted only from facial regions containing good quality rPPG signals. Thus, the eventual rPPG signal quality can be improved. Based on this intuition, an adaptive ROI (AROI) approach is investigated to improve the rPPG signal quality, in which the captured video is divided into block-based pipelines in the spatial and temporal domains for generating spatial-temporal rPPG signal quality-maps on the whole facial area. The quality-maps can describe the quality distribution of rPPG signals on the face. Then AROIs are extracted through segmenting the spatial-temporal quality-maps. Since AROIs are able to adaptively select facial regions containing good-quality rPPG signals, signal quality of the final rPPG wave estimated using AROIs is improved.

The contributions of this paper are as follows: (1) An SNR map is utilized to describe the quality distribution of rPPG signals on the face, without a reference cPPG. (2) An AROI approach for the rPPG is implemented using mean-shift clustering and adaptive thresholding in SNR maps. A better performance in heart rate (HR) measurement can be achieved, compared with conventional ROI methods for the rPPG. (3) The state-of-the-art motion-resistant rPPG techniques can be effectively enhanced through integration with the proposed AROI.

The rest of the paper is organized as follows. Related work about rPPG estimation is briefly reviewed in Section 2. In Section 3, the SNR map for AROI calculation on the face is introduced. The proposed AROI approach for rPPG is explained in Section 4. In order to verify the proposed AROI approach, experimental assessment details are presented in Section 5 and our experimental results are given in Section 6. Finally, a conclusion is drawn in Section 7.

2 Related work

In previous works, there have been some worthwhile approaches toward improving the rPPG estimation process. These approaches can be mainly classified into three categories: chrominance-based, adaptive bandpass filter (ABF)-based, and ROI-based.

The chrominance-based approach utilizes the wavelength dependent characteristic of the rPPG signal magnitude due to the light absorption of hemoglobin [7]. A regular digital camera sensor consists of red, green, and blue (RGB) color channels. rPPG obtains the strongest pulse

signal in the green light, and weaker pulse signals in red and blue light. Though with different amplitudes, all the three rPPG signals in different color channels contain similar artifacts. An appropriate linear combination of rPPG signals from different color channels can enhance the final rPPG signal while reducing artifacts, simultaneously. The representative chrominance-based methods are ICA [25], principal component analysis (PCA) [16], the skin tone-based method [12], and the adaptive green/red difference [10].

The ABF-based methods focus on refining rPPG signals based on frequency characteristics in transform domains. The wavelet transform has been successfully applied in the field of biometrics due to its effective spatial-temporal representation of images and signals [17]. Wavelets especially can be constructed with special abilities, such as rotation-invariance and singular detection in the high-frequency sub-bands [35]. Bousefsaf et al. [5] utilized the continuous wavelet transform to denoise and refine the peaks of rPPG signals. After a skin mask-based pre-processing, the rPPG signal in the wavelet domain has a maximum corresponding to the HR on the frequency axis, occupying the dominant energy. Then, the wavelet transform coefficients are filtered using a wavelet energy curve to reduce noises. Tarassenko et al. [28] employed an autoregressive model to perform the spectral analysis. A pole cancellation algorithm is proposed to find the pole corresponding to the HR and cancel out aliased frequency components caused by artificial light flicker, simultaneously.

The ROI-based approach aims to select a good ROI on the skin area in the captured video frames for the rPPG signal estimation. Hence, the selected ROI can provide a higher quality rPPG signal as compared with the whole skin region such as the whole face. In the conventional ROI-based methods, all pixels within an ROI are averaged to generate the rPPG signal. However, different skin regions carry rPPG signals with different quality, due to the difference in hair density, epidermis thickness, capillary density, facial features, etc. To achieve high-quality rPPG signal estimation, only the regions that can provide relatively clear rPPG signals should be selected as the ROI, while regions containing low-quality rPPG signals should be discarded. Poh et al. [25] chose the whole face detected by the cascade classifier as the ROI. As the forehead and cheeks regions can provide stronger rPPG signals as compared with other regions on the face, they are chosen as the ROI instead of the whole face [21, 31]. Skin detection is utilized to mask skin regions and discard non-skin regions on the face frame by frame, since non-skin regions rarely contain rPPG information [5, 20]. Thus, the dynamic skin mask can be treated as a dynamic ROI for rPPG.

In most of the ROI-based methods for rPPG, the ROI selection focuses on pre-defined regions that have good potential to provide high-quality signals, such as the forehead, the cheeks, and the skin region masked by skin detection. However, an effective method to quantitatively assess the quality of rPPG signals on different facial regions for the ROI selection is still lacking. If the rPPG signal quality distribution on the face can be estimated, a good ROI is possible to be adaptively selected for high-quality rPPG signal estimation.

3 SNR map

The critical step for determining an AROI in the proposed approach is to estimate the spatial-temporal quality distribution of rPPG signals on the face. The SNR map is utilized to implement such a task.

3.1 SNR of a PPG signal

A good PPG signal is quasi-periodical, which should have a clear dominant peak and harmonics in its spectrum as shown in Fig. 2(a)-(b). When the PPG signal is distorted, the dominant peak and harmonics will be weakened with a flattened spectrum as shown in Fig. 2(c)-(d).

Based on the frequency characteristics of the PPG signal as shown in Fig. 2, a custom SNR for assessing quality of rPPG signals is defined as

$$SNR = \frac{S(f_{HR})}{\sum_{f=0}^{f_{sampling}/2} S(f) - S(f_{HR})} \quad (1)$$

where, $S(f)$ is the spectral density function of the rPPG signal, $S(f_{HR})$ is the spectral density at the HR frequency f_{HR} , and $f_{sampling}$ is the sampling rate of the rPPG signal. Similar SNR definitions have been proposed for measuring rPPG signal quality in [12, 20]. In actual use, there is no reference cPPG signal to provide reference f_{HR} for assessing the rPPG signal quality. Therefore, f_{HR} is estimated from the whole facial region first. In order to obtain a reliable HR estimation for reference, the adaptive green/red difference technique is utilized to enhance the rPPG estimation on the face [10]. Thus, a stable estimation of f_{HR} can be provided in the frequency domain.

3.2 SNR map on the face

The whole facial region is divided into non-overlapped blocks. The rPPG signal can be accessed from each block region. SNR for each rPPG signal measured from each block-based video pipeline is calculated using the estimated f_{HR} from the whole face. SNR values

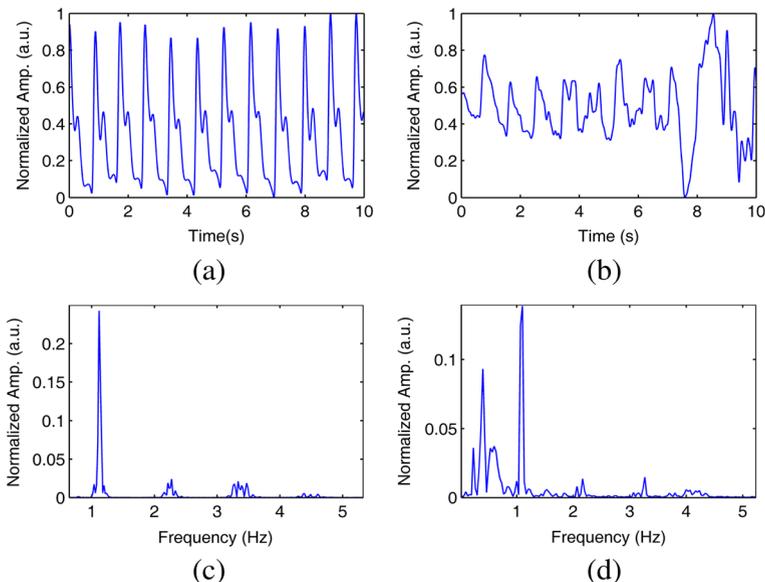


Fig. 2 Normal and distorted cPPG waveforms (detrended and normalized) and their normalized spectra. **a** A normal cPPG waveform. **b** The normalized spectrum of the normal cPPG waveform. **c** A distorted cPPG waveform. **d** The normalized spectrum of the distorted cPPG waveform

corresponding to all blocks form an SNR map, which can reflect the rPPG signal quality distribution on the face. High SNR represents good quality block-ROIs, and low SNR represents noisy block-ROIs.

Fig. 3 demonstrates SNR maps of three subjects with different skin colors, of skin color type I-II, type III-IV, and type V-VI respectively, according to Fitzpatrick's chromatic scale [11].

4 Methodology

Based on the estimated quality-maps of rPPG signals using SNR, the flow chart of the proposed AROI method for rPPG signal estimation is illustrated in Fig. 4. Firstly, a face video is captured using a digital camera. During the video recording, timestamps of every frame in the video are recorded for interpolation of the estimated rPPG signals, because low-end digital cameras do not have stable frame-rates [27]. In order to reduce motion artifacts in rPPG signals, face tracking is performed using the Kanade – Lucas – Tomasi (KLT) algorithm [22, 29]. After the motion matrix describing the face motion is obtained, the face image can be stabilized and resized into a uniform resolution, as shown in Fig. 4(a) and.

4(b). Then each frame in the face video is divided into non-overlapped square blocks of $N \times N$ pixels. A typical block size is 8×8 pixels. After the spatial block-division, the face video is divided with overlap on the temporal axis with a time-interval of T seconds and an overlap of $T/2$ s. A typical time-interval of T is 8 s. Thus, the face video is segmented into block-based video pipelines, and each pipeline has a volume size of $N \times N$ -pixel \times T -second, as shown in Fig. 4(c). Then SNR of the rPPG wave is calculated in each video pipeline. For a long face video, a sequence of SNR maps is thus obtained. And each SNR map can describe the quality distribution of rPPG signals on the face within a time-interval, as shown in

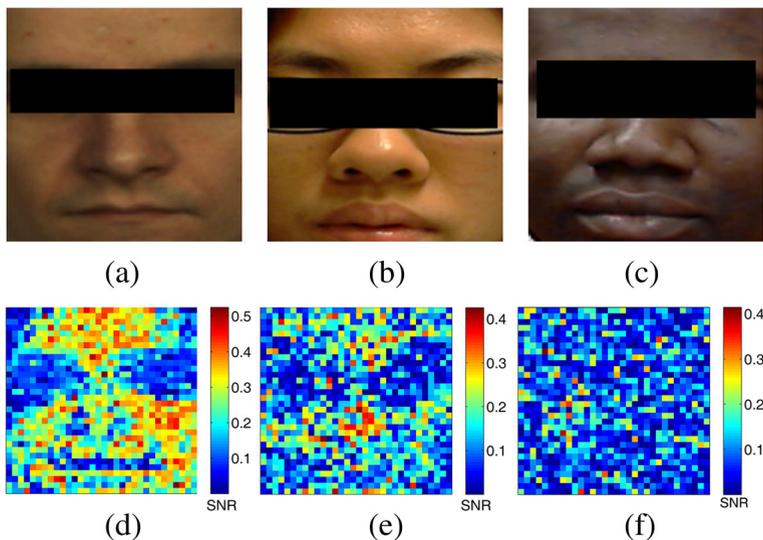


Fig. 3 Face images and the corresponding SNR maps of subjects with different skin colors. **a** Face image of type I-II skin color. **b** Face image of type III-IV skin color. **c** Face image of type V-VI skin color. **d** SNR map of type I-II skin color. **e** SNR map of type III-IV skin color. **f** SNR map of type V-VI skin color. The skin color is classified according to Fitzpatrick's chromatic scale

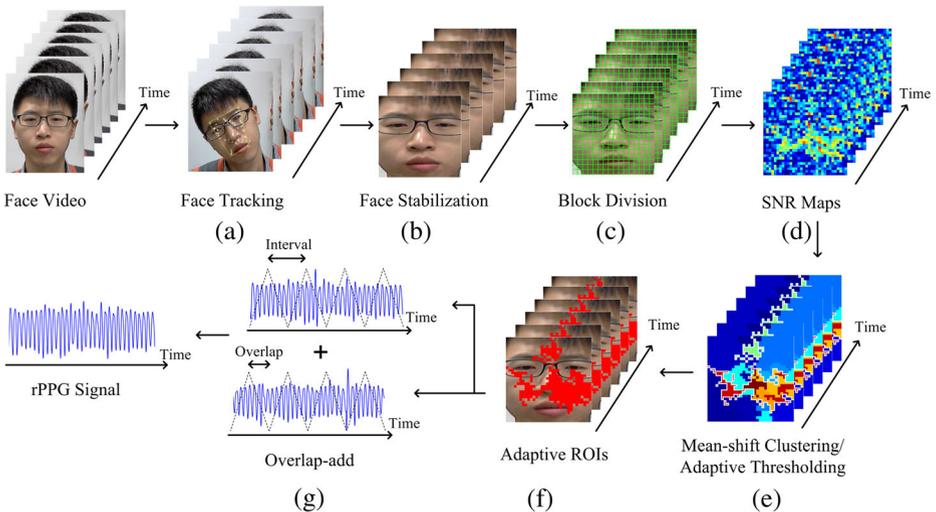


Fig. 4 Flow chart of the proposed AROI approach for rPPG. **a** Face is tracked using the KLT algorithm. **b** Face is stabilized through motion compensation and resized into a uniform resolution. **c** Each face image is divided into non-overlapped blocks. **d** SNR maps are calculated using video pipelines with a volume of block-size \times time-interval. **e** Mean-shift clustering is performed in SNR maps, and adaptive thresholds are determined. **f** Block-wise AROIs are obtained through thresholding in SNR maps. **g** The final rPPG signal is obtained through averaging pixel values within AROIs, and discontinuities are removed using the overlap-add operation

Fig. 4(d). Mean-shift clustering is performed in each SNR map to cluster facial regions based on SNR values. After the clustering, adaptive thresholding is utilized to select clusters with relatively higher mean SNR values to generate an AROI, because they have a higher chance of carrying high-quality rPPG information, as shown in Fig. 4(e). In this way, a sequence of AROIs is obtained through mean-shift clustering and adaptive thresholding, as shown in Fig. 4(f). Finally, the estimated rPPG signal is obtained by averaging pixel values of the face frames masked by AROIs, and discontinuities in an rPPG signal caused by AROI transitions are removed using an overlap-add operation, as shown in Fig. 4(g).

Two video clips illustrating the proposed AROI approach are provided as supplementary materials: (1) a still subject [8]; (2) a moving subject. Detailed descriptions for each module in the flow chart are given in the following subsections [9].

4.1 Face stabilization

Currently, object tracking can be handled well using robust sparse features and online training [15, 23]. However, this work focuses on locating bounding boxes of objects other than motion compensation. In order to stabilize the face for effectively estimating AROI, the local feature matching is selected, which has been utilized for the motion artifacts suppression for rPPG [10]. This face tracking is revised to perform face stabilization in this paper. The Viola – Jones (VJ) face detector [32] is used to detect a rectangular facial region (width \times height) in the initial frame of a video (green rectangle in Fig. 5(a)), and afterwards a trapezoidal region (yellow trapezoid in Fig. 5(a)) is automatically selected by ratios ($0.5 \times$ width, $0.4 \times$ width, $0.58 \times$ height). Speeded up robust features (SURF) points are detected in the trapezoid region [2]. Then, the KLT algorithm is used to track SURF points, as shown in Fig. 5(b). The motion

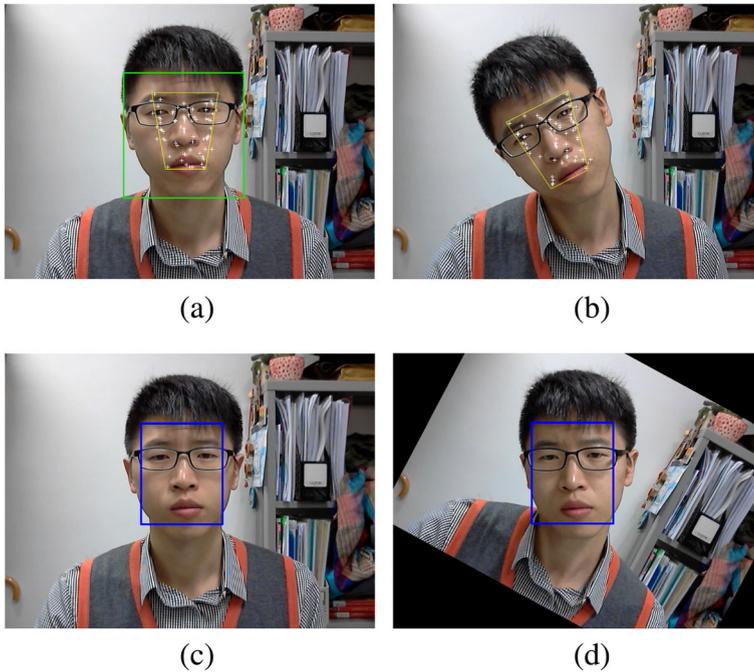


Fig. 5 Face tracking and stabilization. **a** Face detection and SURF points detection. **b** Face tracking using the KLT algorithm. **c** The reference face position in the first frame (Rectangle). **d** Following face images are stabilized referring to the reference face position

matrix describing the face motion between consecutive frames can be calculated during tracking. The face position in the first frame is treated as the reference position ($0.6 \times$ width, $0.9 \times$ height). All face images in the following frames are warped into the reference face position using motion matrices. Thus, face images can be stabilized, even if the subject is moving.

4.2 Mean-shift clustering and thresholding in SNR maps

After a face image is stabilized in the reference face position, the face image is resized into a uniform resolution. Thus, a uniform block-division can be performed for face videos with different resolutions, in which each face image will be divided into non-overlapped $N \times N$ blocks, as shown in Fig. 6(a). After the spatial block-division, the face video is divided temporally with a time-interval of T seconds and an overlap of $T/2$ s. Thus, the face video is divided into video pipelines with a volume size of $N \times N$ -pixel $\times T$ -second. Within each T seconds, $N \times N$ rPPG signals are measured from $N \times N$ face video pipelines using a bandpass filter. Then, a block-wise SNR map can be calculated for a T -second-long face video, in which each point records the SNR value of the rPPG signal measured from the corresponding video pipeline, as shown in Fig. 6(b).

Our assumption is that the rPPG signal quality can be improved, if only the block-based facial regions with mean SNR values above a threshold are selected as ROI for estimating the rPPG signal.

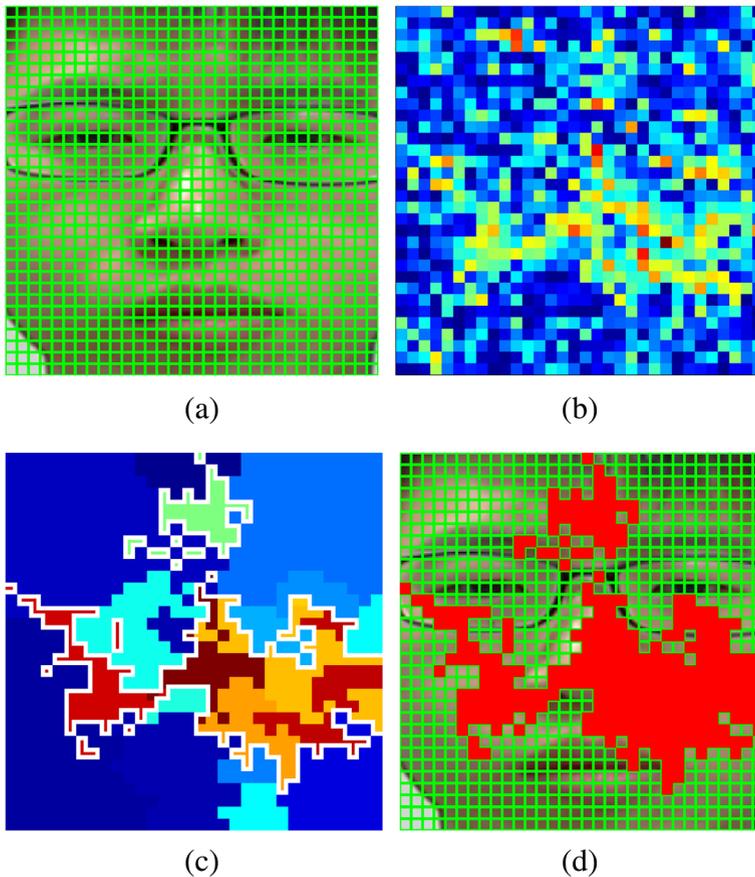


Fig. 6 Generating an AROI through mean-shift clustering and adaptive thresholding in a SNR map. **a** Block-division in a face image. **b** The SNR map for a T-second-long face video. **c** Mean-shift clustering and adaptive thresholding in the SNR map. **d** The AROI generated (Red blocks – AROI)

Mean-shift clustering Before thresholding, mean-shift clustering is performed in the SNR map, due to the fact that facial regions providing good-quality rPPG signals are distributed as clusters, not sparsely. The selection of mean-shift as the clustering algorithm is because mean-shift clustering does not require prior knowledge of the number of clusters, and does not constrain shapes of clusters [6]. This nonparametric clustering algorithm is suitable for the SNR map. Based on the observation that a cluster on the face should share a spatial neighborhood and similar rPPG signal amplitudes, the kernel density estimator utilized for mean-shift is expressed as

$$K_{h_{SNR}, h_S, h_A}(x) = \frac{C}{h_{SNR} h_S^2 h_A} k\left(\left\|\frac{x^{SNR}}{h_{SNR}}\right\|^2\right) k\left(\left\|\frac{x^S}{h_S}\right\|^2\right) k\left(\left\|\frac{x^A}{h_A}\right\|^2\right) \quad (2)$$

where x is the feature vector, x^{SNR} is the SNR part, x^S is the spatial part, x^A is the rPPG signal amplitude part, $k(x)$ is the common profile used in the three domains, C is the normalization constant, and h_{SNR} , h_S , and h_A are the bandwidth parameters for the SNR part, spatial part, and

signal amplitude part, respectively. Gaussian kernel is utilized here. Fig. 6(c) illustrates the clustering result of Fig. 6(b).

Adaptive thresholding After the clustering, an adaptive thresholding based on Otsu's method is utilized to automatically select the clusters with relatively higher mean SNR values on the face as the AROI [24]. It is assumed that all the clusters on the face can be classified into two classes based on their mean SNR values (foreground and background). The two classes are separated by minimizing the intra-class variance. The SNR threshold is determined as

$$t = \text{Arg Min} \{ \omega_1(t) \sigma_1^2(t) + \omega_2(t) \sigma_2^2(t) \} \quad (3)$$

subject to $0.2 < |AROI|/|WROI| < 0.5$

where t is the SNR threshold, $\omega_1(t)$ and $\omega_2(t)$ are the class probabilities of two separated classes respectively, $\sigma_1^2(t)$ and $\sigma_2^2(t)$ are variances of these two classes, $|AROI|$ is the size of AROI, $|WROI|$ is the size of the whole ROI available (whole face). The optimization of threshold t should be subject to AROI taking 20%–50% of the whole ROI size. Otherwise, the optimized threshold t is unreasonable. 20% and 50% are loose experiential limits. The white borderlines in Fig. 6(c) surround the clusters, which are determined as the AROI by the adaptive threshold t . Fig. 6(d) illustrates the AROI in red. The process of mean-shift clustering and adaptive thresholding is repeated in each overlapped T-second-long face video. Thus, a sequence of AROIs is obtained.

4.3 Averaging in the AROI and overlap-add

After an AROI is obtained for a T-second-long face video, the final rPPG signal is estimated by averaging the rPPG signals loaded in block-based video pipelines located within the AROI given as

$$\text{rPPG}_{\text{AROI}}(t) = \frac{\sum_{x,y \in \text{AROI}} \text{rPPG}(x,y,t)}{|\text{AROI}|} \quad (4)$$

where $\text{rPPG}_{\text{AROI}}(t)$ is an rPPG signal measured using an AROI for a T-second-long face video, $\text{rPPG}(x,y,t)$ is the rPPG wave located at (x,y) in the AROI at time t , and $|\text{AROI}|$ is the size of the AROI.

Due to the continuous updating of AROIs, the change of AROI will cause discontinuities in the rPPG signal at the transitions of consecutive AROIs, as shown in Fig. 7 (a). These discontinuities will affect peak detection in rPPG signals, and the accuracy of HR measurement will decrease. In order to remove these discontinuities due to AROI transitions, an overlap-add operation is performed as shown in Fig. 7. As mentioned in Section 4.2, the face video is divided overlappedly on the temporal axis with a time-interval of T seconds and an overlap of T/2 s. The overlap of T/2 s is designed for compensating discontinuities which occur at the start and end of a time-interval of T seconds. Each rPPG signal accessed from a T-second-long face video masked by an AROI is multiplied with a triangle window with a length of T seconds, as shown in Fig. 7(b). The triangle window zeros off the discontinuities caused by AROI transitions, and the overlap-add operation can complement the final rPPG signal, as shown in Fig. 7(c).

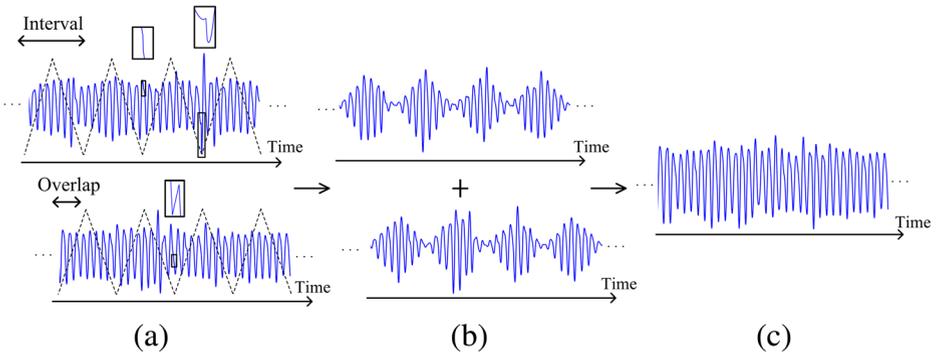


Fig. 7 Overlap-add operation for removing discontinuities in rPPG signals due to AROI transitions. **a** rPPG signals estimated using consecutive AROIs have an overlap of $T/2$ s. A T -second-long triangle window is designed to zero off discontinuities at the start and end of each time-interval through multiplication. **b** Consecutive rPPG signals estimated using consecutive AROIs are overlap-added. **c** Discontinuities due to AROI transitions are removed in the final rPPG signal

5 Assessment details

Instantaneous HR measurement is utilized to test the rPPG estimation performance. A custom HR algorithm is adopted to estimate instantaneous HR from the rPPG signal [10].

5.1 Material

There are two datasets for testing the proposed AROI approach as listed in Table 1. The test dataset I is for testing the performance of the proposed AROI approach in measuring rPPG signals for still subjects. The test dataset II is for testing the performance of the AROI approach in measuring rPPG signals for moving subjects.

In the test dataset I, 21 healthy volunteers (14 males, 7 females) with different skin colors distributed uniformly from type I to type VI (light – dark) were enrolled. Fitzpatrick’s chromatic scale was utilized to classify different skin colors [11]. The volunteers’ age ranged from 20 to 33. None of them had any known cardiovascular diseases. Informed consent was obtained from each subject, and the study was approved by the Research Committee of City University of Hong Kong. A low-cost webcam (Logitech C270) was used as the imaging device. The webcam worked at a frame-rate of 30 fps. All videos were recorded in color space (24-bit RGB) with a resolution of 640×480 , and saved in uncompressed AVI format. The auto white balance function of the webcam was turned off. During the video-recording, there were no dedicated light sources except normal indoor fluorescent light (no windows in the room). One 1-min-long video was recorded for each subject, who sat steadily in front of the webcam at a distance of 80 cm. A cPPG device (HRS-06UF Heart Sensor) was attached to the subject’s index finger to record the reference cPPG signal with a sampling rate of 300 Hz.

Table 1 Datasets for experiments

Dataset	Number of subjects	Videos per subject	Skin color	Lighting condition	Motion
Test dataset I	21	1	I – VI	Indoor fluorescent light	Still
Test dataset II	6	5	I – II	Indoor daylight	Moving

The test dataset II is taken from Pulse Rate Detection Dataset (PURE) [20]. PURE consists of 10 persons (skin color type I – II) performing different, controlled head motions in front of a camera. The face videos and reference cPPG signals were recorded simultaneously. There are six different motion setups in PURE, which are steady, talking, slow translation, fast translation, small rotation, and medium rotation. Since the test dataset II is utilized for testing the motion-resistance of the AROI approach, just the last five motion setups in PURE are taken as the test dataset II, as listed in Table 2. In addition, the videos of subjects with IDs of 2, 5, 7, and 9 were not adopted, because their reference cPPG signals were not stable enough to provide accurate references.

5.2 Benchmark methods

To benchmark the proposed AROI method, we compared our experimental results with the state-of-the-art ROI methods for rPPG [5, 21, 25]. In [25], the ROI is the whole facial region. Since the ROI is fixed on the whole skin region available, we call this method the whole fixed ROI (WFROI) method. In [21], the cheeks region is regarded as the optimal fixed ROI on the face, which can provide an rPPG signal with the strongest amplitude on the facial region. Since the ROI on the cheeks is fixed on a partial region of the skin available, it is named as the partial fixed ROI (PFROI) method. In [5], a simple and fast skin-pixel detection based on a piecewise YCbCr linear classifier is utilized to produce a skin mask for each face image. The skin mask is updated every frame. The rPPG signal is measured using this dynamic skin mask as a dynamic ROI. We call this method the skin mask-based ROI method. For a stable performance of skin detection, Gaussian mixture model-based human skin detection was utilized to replace the piecewise YCbCr linear classifier [19]. Screenshots of different ROI methods for rPPG are shown in Fig. 8.

In order to test the motion-robustness of the proposed AROI approach, three representative motion-resistant rPPG methods were selected for a comparison. The first is the ICA-based method, which has been widely used to reduce motion artifacts for rPPG [26, 30]. We utilized the robust ICA algorithm with maximization of the kurtosis contrast for the implementation of the ICA-based method [34]. The second is the wavelet filtering-based method [5]. The third is the ABF method proposed in [10], which can automatically build a narrow custom bandpass filter according to the estimated HR frequency using adaptive green/red difference. The proposed AROI approach was also integrated with the three motion-resistant rPPG methods respectively to test whether the AROI approach can improve the performances of different rPPG methods. The face tracking and stabilization in Section 3.1 were applied in the implementation of all the benchmark methods. Thus, the three motion-resistant rPPG methods were tested with the same pre-processing.

Table 2 The types of motions in test dataset II

Motion type	Setup
I	Talking
II	Slow translation
III	Fast translation
IV	Small rotation
V	Medium rotation

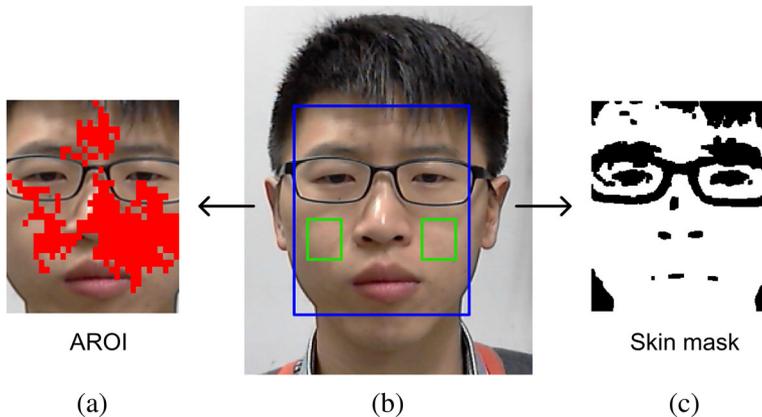


Fig. 8 ROIs on the face. **a** AROI. **b** Fixed ROIs (The outer blue rectangle is WFROI and the inner green rectangles are PFROIs). **c** Skin mask-based ROI

5.3 Statistics

We employed Bland–Altman analysis of HR measurements as the main method to verify agreement between rPPG and cPPG [3]. We examined the agreement between methods of measurement with multiple observations per subject [4]. In addition, Pearson’s correlation coefficients (PCC) and the corresponding p -values were calculated to measure the linear dependency between the rPPG and the cPPG [18]. When plotting Bland–Altman plots, the HR measurement differences between the rPPG and the cPPG were plotted against the mean results of the rPPG and the cPPG. The measurement bias and 95% limits of agreement between the rPPG and the cPPG are calculated to measure the rPPG accuracy. The 95% limits of agreement (LoA) are defined as the bias ± 1.96 times the standard deviation (SD) of the differences, consisting of an upper limit (UL) and a lower limit (LL). LoA shows how far apart the HR measurements made by rPPG are likely to be for most individuals.

5.4 Setting of experimental parameters

In the experiments, the uniform resolution for resizing face images during face stabilization is 256×256 pixels. The pass band of the bandpass filter utilized for accessing rPPG signals is 0.75–3.5 Hz, covering heart rates ranging from 45 to 210 beats per minute (bpm). The green channel in videos is utilized for calculating AROIs, because pulse information is strongest in the green channel among RGB channels.

6 Experimental results

In the experimental study, the performance of the proposed AROI approach was assessed for still subjects first. Then the effect of the AROI approach in the integration with motion-resistant rPPG methods was investigated with moving subjects. Last, time complexities of different ROI methods were analyzed.

Table 3 PCC values of the AROI method with different block-sizes and time-intervals in HR measurement for still subjects

Block-size Duration	4 × 4 pixels	8 × 8 pixels	16 × 16 pixels
4 s	0.7842	0.8543	0.8394
8 s	0.8372	0.9341	0.8976
16 s	0.8323	0.9161	0.8830

The highest PCC value is shown in bold

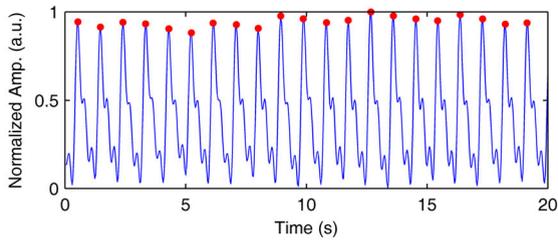
6.1 The performance of AROI in measuring still subjects

The performance of the proposed ROI was first evaluated using the test dataset I, in which only the videos of still subjects were collected. There are two tunable parameters for the AROI estimation, which are the block-size and the time-interval of every block-based video pipeline. In order to find a set of empirical values for these two parameters, the PCC values obtained by the AROI methods with different block-sizes and time-intervals were measured on the test dataset I, as shown in Table 3. When the block-size is small (such as 4 × 4 pixels), the rPPG signal extracted from every block-based video pipeline is too weak for measuring SNR effectively. Then AROI cannot be well estimated. With increase in the block-size, the resolution of the AROI becomes low. A low-resolution AROI cannot precisely describe the quality distribution of rPPG signals. When the time-interval is small (such as 4 s), the frequency resolution of the rPPG signal is too coarse for a bandpass filter. If the time-interval is long, the bandpass filter can affect the rPPG signal waveform, because the Fourier transform cannot describe the temporal variation of the HR frequency. Hence both the block-size and the time-interval should take a moderate value. Based on experimental results in Table 3, the block-size for the spatial division on the face image is set as 8 × 8 pixels. The time-interval for the temporal division of the face video is 8 s, and the overlap accordingly is 4 s. This set of parameters was utilized in the following sub-sections.

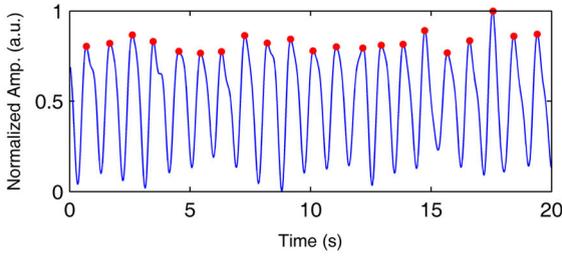
A set of rPPG waveforms and the reference cPPG waveform measured from a sample still subject (skin color III-IV) are given for an intuitive illustration in Fig. 9. Compared with rPPG waveforms estimated using the WFROI, PFROI, and skin mask, the rPPG waveform estimated using the AROI obtains the best waveform quality. Artifacts or distortions are removed by the AROI approach.

rPPG signals were measured from all videos in the test dataset I using different ROI methods. Then HR measurement was performed on these rPPG signals and reference cPPG signals. The Bland–Altman plot with multiple instantaneous HR measurements per subject for the rPPG using the AROI method is given in Fig. 10. The HR measurement bias between the cPPG and the rPPG using the AROI method is 0.1 bpm. The 95% degree of agreement is between −5.5 bpm and 5.7 bpm. The PCC between instantaneous HRs measured by the cPPG and the rPPG using the AROI method is 0.9341 (p -value < 0.001). The above analyses of the HR measurement accuracy were also performed for different benchmark ROI methods for the rPPG. The statistical results are compared in Table 4.

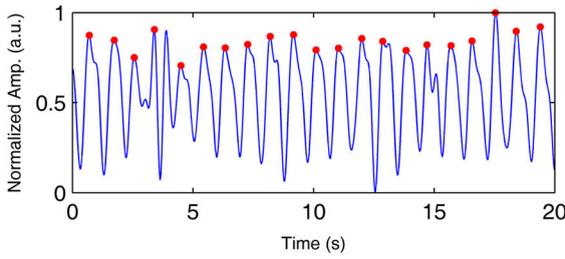
Fig. 9 rPPG waveforms measured from the sample still subject using different ROI approaches and the corresponding reference cPPG waveform. **a** The reference cPPG waveform. **b** The rPPG waveform measured using the AROI. **c** The rPPG waveform measured using the WFROI. **d** The rPPG waveform measured using the PFROI. **e** The rPPG waveform measured using the skin mask



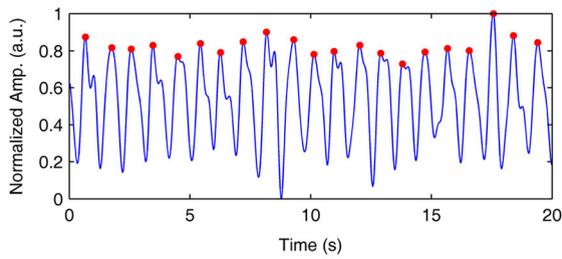
(a)



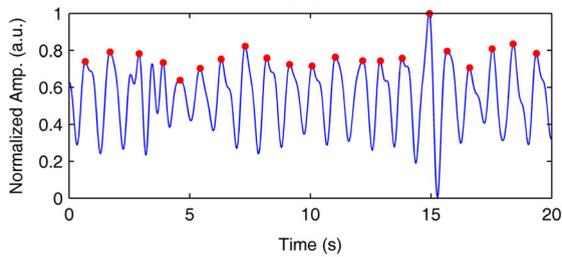
(b)



(c)

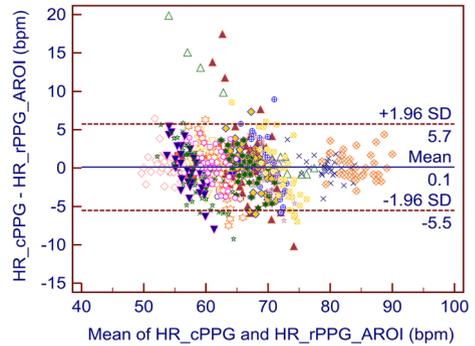


(d)



(e)

Fig. 10 Bland – Altman plot with multiple instantaneous HR measurements per subject for the comparison of cPPG and rPPG using the AROI method. (Different markers correspond to different subjects)



Compared with the benchmark ROI methods for rPPG, the proposed AROI method achieved the best HR measurement accuracy for still subjects, as shown in Table 4. The WFROI and PFROI obtained comparable HR measurement accuracies. The performance of the skin mask-based ROI method is relatively worse, because the ROI transition every face frame brought artifacts. The reason is similar to that of discontinuities as explained in Section 4.3.

6.2 The performance of AROI in measuring moving subjects

Subjects under measurement do not always keep still in practice. The effect of the proposed AROI approach on motion-robustness of the rPPG was tested using the test dataset II. First, HR measurements using different motion-resistant rPPG methods were performed for different types of motion in test dataset II. For a performance comparison among different ROI methods, three different types of ROI methods, including WFROI, skin mask, and AROI were integrated with the three motion-resistant rPPG methods respectively. Last, the HR measurement performance of the AROI approach alone was investigated for different types of motion. The HR measurement accuracies of the proposed AROI approach and three benchmark motion-resistant rPPG methods integrated with three different ROI methods respectively are listed in Tables 5, 6, 7, 8 and 9. For motion I, motion II, and motion III, the AROI-aided wavelet filtering-based method obtained the best performances. For motion IV and motion V, the AROI-aided wavelet filtering-based method and the AROI-aided ABF-based method achieved comparable best performance. A normal HR varies slowly on the temporal axis, which means the HR frequency has a slow temporal variation. The Fourier transform can only analyze frequency information without the temporal decomposition.

Table 4 Performances of different ROI methods for rPPG in HR measurement for still subjects

Statistic		WFROI	PFROI	Skin mask	AROI
Bias (bpm)		0.8	0.8	1.3	0.1
LoA	UL (bpm)	10.8	10.8	12.6	5.7
	LL (bpm)	-9.2	-9.3	-10.0	-5.5
PCC		0.8237	0.8123	0.7851	0.9341
<i>p</i> -value		<0.001	<0.001	<0.001	<0.001

Best results are shown in bold

Table 5 Performances of different rPPG methods in HR measurement for moving subjects with motion I

Statistic	ICA+ WFROI	ICA+ Skin	ICA+ AROI	Wavelet + WFROI	Wavelet + Skin	Wavelet + AROI	ABF+ WFROI	ABF+ Skin	ABF+ AROI	AROI
Bias (bpm)	0.4	1.1	0.5	0.9	1.6	0.6	0.9	0.9	0.7	0.4
LoA UL(bpm)	7.9	10.8	5.1	10.8	12.1	4.8	6.8	8.9	5.4	6.6
LL(bpm)	-7.0	-8.5	-4.1	-9.0	-8.9	-3.6	-5.0	-7.1	-4.2	-5.9
PCC	0.9060	0.8552	0.9580	0.8268	0.8033	0.9640	0.9363	0.8941	0.9517	0.9224
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Best results are shown in bold

Table 6 Performances of different rPPG methods in HR measurement for moving subjects with motion II

Statistic	ICA+ WFROI	ICA+ Skin	ICA+ AROI	Wavelet + WFROI	Wavelet + Skin	Wavelet + AROI	ABF+ WFROI	ABF+ Skin	ABF+ AROI
Bias (bpm)	0.3	0.3	0.5	0.5	0.5	0.5	0.5	0.5	0.5
LoA UL(bpm)	4.4	4.6	3.3	2.7	2.6	2.5	4.1	3.7	2.9
LoA LL(bpm)	-3.9	-3.9	-2.2	-1.7	-1.5	-1.5	-3.2	-2.7	-1.9
PCC	0.9468	0.9449	0.9753	0.9765	0.9859	0.9868	0.9542	0.9619	0.9813
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Best results are shown in bold

Table 7 Performances of different rPPG methods in HR measurement for moving subjects with motion III

Statistic	ICA+ WFROI	ICA+ Skin	ICA+ AROI	Wavelet + WFROI	Wavelet + Skin	Wavelet + AROI	ABF+ WFROI	ABF+ Skin	ABF+ AROI
Bias (bpm)	0.4	0.6	0.7	0.5	3.5	0.7	0.6	0.7	0.8
LoA	5.8	17.6	5.0	4.9	16.5	3.7	4.1	14.3	4.0
UL(bpm)	-5.1	-11.2	-3.7	-4.0	-9.6	-2.3	-2.9	-9.0	-2.5
LL(bpm)	0.9510	0.8930	0.9673	0.9680	0.9192	0.9838	0.9778	0.9216	0.9807
PCC	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
p-value									

Best results are shown in bold

Table 8 Performances of different rPPG methods in HR measurement for moving subjects with motion IV

Statistic	ICA+ WFROI	ICA+ Skin	ICA+ AROI	Wavelet + WFROI	Wavelet + Skin	Wavelet + AROI	ABF+ WFROI	ABF+ Skin	ABF+ AROI
Bias (bpm)	0.6	0.4	0.6	0.5	0.6	0.6	0.5	0.5	0.6
LoA	4.8	6.8	4.3	4.3	6.8	4.0	4.3	5.1	4.1
LL(bpm)	-3.7	-6.0	-3.2	-3.3	-5.1	-2.9	-3.4	-4.1	-2.8
PCC	0.9640	0.9319	0.9701	0.9736	0.9249	0.9753	0.9681	0.9622	0.9748
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Best results are shown in bold

Table 9 Performances of different rPPG methods in HR measurement for moving subjects with motion V

Statistic	ICA+ WFROI	ICA+ Skin	ICA+ AROI	Wavelet + WFROI	Wavelet + Skin	Wavelet + AROI	ABF+ WFROI	ABF+ Skin	ABF+ AROI	AROI
Bias (bpm)	0.6	0.4	0.6	0.8	1.0	0.0	1.4	0.6	0.9	0.6
LoA	7.8	7.6	7.2	6.5	8.8	6.1	9.0	7.3	6.9	9.7
UL(bpm)	-6.6	-6.8	-5.9	-5.5	-6.9	-6.1	-6.3	-6.1	-5.1	-8.6
LL(bpm)	0.9322	0.9284	0.9446	0.9444	0.9228	0.9458	0.9287	0.9354	0.9476	0.8766
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Best results are shown in bold

Compared with the Fourier transform, the wavelet transform can decompose signals with both frequency and location information. Hence the wavelet filtering can reduce artifacts of the rPPG signal and maintain the temporal variation of the HR frequency within a temporal window. Thus, the wavelet filtering-based method performs better than the ABF-based method, and the AROI can improve the wavelet filtering-based method further. Generally, the motion-robustness of the AROI approach alone cannot overtake the three benchmark motion-resistant rPPG methods. However, HR measurement accuracies of all the three benchmark methods can be improved through being integrated with the AROI approach. And, all the best results were achieved by the AROI-aided motion-resistant rPPG methods. This implies that the AROI approach is an effective direction to enhance the rPPG signal quality, in addition to the chrominance-based approach and the ABF-based approach.

6.3 Time complexity analysis

A real-time implementation is important for rPPG in health monitoring applications. Hence, the time complexity is a critical issue to be addressed. The time consumption in processing a 1-min-long video using every ROI method is shown in Table 10. The video has a frame rate of 30 fps. This time complexity comparison was performed using a laptop computer with a 4-core 2.6 GHz CPU. Even though a multi-core CPU was utilized, the implementation is only single-threaded. The proposed AROI has a higher time complexity compared with the state-of-the-art ROI methods. However, the AROI still attains a real-time speed for processing a video with a typical frame rate of 30 fps, because only 23.16 s are needed for a 1-min-long video.

7 Conclusion

rPPG is a promising technique, which can realize remote human vital signs monitoring. On the one hand, the remote operation can be very helpful in special situations, such as monitoring neonates, telemedicine, fitness monitors, etc. On the other hand, the rPPG signal quality is affected by its remote imaging mode, especially when low-cost cameras are utilized. The proposed AROI approach can improve the rPPG signal quality through dynamically selecting skin regions loading good-quality rPPG information as ROI. Experimental results show that the proposed AROI method achieved a better HR measurement accuracy than the state-of-the-art ROI methods for rPPG. Besides good performance in measuring still subjects, the proposed AROI method can also be integrated with other popular motion-resistant rPPG methods to further improve their performance, because the AROI provides a new effective approach in addition to the chrominance-based approach and the ABF-based approach. Although only the facial region was analyzed in this paper, the proposed AROI can also be utilized in other skin

Table 10 Time consumption of processing a 1-min-long video (30 fps) using different ROI methods

ROI method	WFROI	PFROI	Skin	AROI
Time consumption (second)	16.2	16.2	21.6	23.16

areas, such as the palm. Since the proposed AROI approach for rPPG is able to adjust ROI according to the spatial-temporal rPPG signal quality distribution on a skin region, we believe that the AROI method can enhance the adaptability of rPPG in practice.

The pipeline of salient region detection can be transferred into the AROI application. A spatially weighted region contrast is used to describe saliency of a region, like SNR describing rPPG signal quality after mean-shifting. Then a GrabCut-like segmentation is performed to extract the saliency mask, just like adaptive thresholding is utilized to extract the AROI [13]. Han et al. developed stacked denoising autoencoders to model the background. Then salient objects can be separated from the background through measuring reconstruction residuals of deep autoencoders [36]. Similarly, if there are enough training data, the deep learning architecture can be utilized to learn the AROI extraction, with a prior that low-quality rPPG signals are concentrated around the mouth and eyes regions. With well-developed salient region detection, the AROI extraction may get refined. In future work, non-rigid face tracking and video segmentation with soft boundaries will be investigated to enhance rPPG.

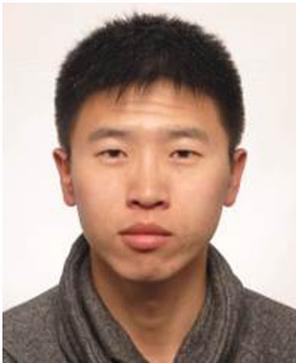
References

- Allen J (2007) Photoplethysmography and its application in clinical physiological measurement. *Physiol Meas* 28:R1–R39. doi:10.1088/0967-3334/28/3/R01
- Bay H, Ess A, Tuytelaars T, Gool LV (2008) SURF: speeded up robust features. *Comput Vis Image Und* 110:346–359. doi:10.1007/11744023_32
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310. doi:10.1016/S0140-6736(86)90837-8
- Bland JM, Altman DG (2007) Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 17:571–582. doi:10.1080/10543400701329422
- Bousefsaf F, Maaoui C, Pruski A (2014) Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomed Signal Process Control* 8:568–574. doi:10.1016/j.bspc.2013.05.010
- Comaniciu D, Meer P (2002) Mean shift: A robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell* 24:603–619. doi:10.1109/34.1000236
- Crowe JA, Damianou D (1992) The wavelength dependence of the photoplethysmogram and its application to pulse oximetry. In: *Proc IEEE Annu Int Conf Eng Med Biol Soc*. p 2423–2424. doi:10.1109/IEEMBS.1992.5761522
- Feng L (2015a) Block-based adaptive ROI for remote photoplethysmography (still subject) <https://youtu.be/7bR8pEM5bnc>. Accessed 24 July 2015
- Feng L (2015b) Block-based adaptive ROI for remote photoplethysmography (moving subject) https://youtu.be/TR_fkny2tt0. Accessed 24 July 2015
- Feng L, Po LM, Xu X, Li Y, Ma R (2014) Motion resistant remote imaging photoplethysmography based on optical properties of skin. *IEEE Trans Circuits Syst Video Technol* 25:879–891. doi:10.1109/TCSVT.2014.2364415
- Fitzpatrick TB (1975) Sun and skin (Soleil et peau). *J Med Esthétique* 2:33–34. doi:10.1159/000251345
- Haan G, Jeanne V (2013) Robust pulse rate from chrominance-based rPPG. *IEEE Trans Biomed Eng* 60:2878–2886. doi:10.1109/TBME.2013.2266196
- Han J, Zhang D, Hu X, Guo L, Ren J, Wu F (2015) Background prior-based salient object detection via deep reconstruction residual. *IEEE Trans Circuits Syst Video Technol* 25:1309–1321. doi:10.1109/TCSVT.2014.2381471
- Hayes MJ, Smith PR (1998) Artifact reduction in photoplethysmography. *Appl Opt* 37:7437–7446. doi:10.1364/AO.37.007437
- He Z, Yi S, Cheung YM, You X, Tang YY (2016) Robust object tracking via key patch sparse representation. *IEEE Trans Cybern* 47:354–364. doi:10.1109/TCYB.2016.2514714

16. Holton BD, Mannapperum K, Lesniewski PJ, Thomas JC (2013) Signal recovery in imaging photoplethysmography. *Physiol Meas* 34:1499–1511
17. Huang J, You X, Yuan Y, Yang F, Lin L (2010) Rotation invariant iris feature extraction using Gaussian Markov random fields with non-separable wavelet. *Neurocomputing* 73:883–894. doi:[10.1016/j.neucom.2009.09.016](https://doi.org/10.1016/j.neucom.2009.09.016)
18. Jacob B, Chen J, Huang Y, Cohen I (2009) Pearson correlation coefficient. In: Cohen I (ed) *Noise reduction in speech processing*. Springer, Berlin, pp 1–4
19. Jones MJ, Rehg JM (2002) Statistical color models with application to skin detection. *Int J Comput Vis* 46: 81–96. doi:[10.1023/A:1013200319198](https://doi.org/10.1023/A:1013200319198)
20. Kawasaki N (2011) Improving motion robustness of contact-less monitoring of heart rate using video analysis. MS thesis, Department of Mathematics and Computer Science, Eindhoven University of Technology
21. Lempe G, Zaunseder S, Wirthgen T, Zipser S, Malberg H (2013) ROI selection for remote photoplethysmography. *Bildverarbeitung für die Medizin* p:99–103. doi:[10.1007/978-3-642-36480-8_19](https://doi.org/10.1007/978-3-642-36480-8_19)
22. Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: *Proc IJCAI*. p 674–679 https://cecas.clemson.edu/~stb/kl/lucas_bruce_d_1981_1.pdf
23. Ma X, Liu Q, He Z, Zhang X, Chen WS (2016) Visual tracking via exemplar regression model. *Knowl-Based Syst* 106:26–37. doi:[10.1016/j.knosys.2016.05.028](https://doi.org/10.1016/j.knosys.2016.05.028)
24. Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst, Man, Cybern, Syst SMC-9(1):62–66*. doi:[10.1109/TSMC.1979.4310076](https://doi.org/10.1109/TSMC.1979.4310076)
25. Poh M, McDuff DJ, Picard RW (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt Express* 18:10762–10774. doi:[10.1364/OE.18.010762](https://doi.org/10.1364/OE.18.010762)
26. Sun Y, Hu S, Azorin-Peris V, Greenwald S, Chambers J (2011) Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *J Biomed Opt* 16:077010–1–077010-9. doi:[10.1117/1.3602852](https://doi.org/10.1117/1.3602852)
27. Sun Y, Hu S, Azorin-Peris V, Kalawsky R, Greenwald S (2013) Noncontact imaging photoplethysmography to effectively access pulse rate variability. *J Biomed Opt* 18:061205–061205. doi:[10.1117/1.JBO.18.6.061205](https://doi.org/10.1117/1.JBO.18.6.061205)
28. Tarassenko L, Villarroel M, Guazzi A, Jorge J, Clifton DA, Pugh C (2014) Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiol Meas* 35:807–831. doi:[10.1088/0967-3334/35/5/807](https://doi.org/10.1088/0967-3334/35/5/807)
29. Tomasi C, Kanade T (1991) Detection and tracking of point features. Carnegie Mellon University Technical Report CMU-CS-91-132 <https://cecas.clemson.edu/~stb/kl/tomasi-kanade-techreport-1991.pdf>
30. Tsouri GR, Kyal S, Dianat S, Mestha LK (2012) Constrained independent component analysis approach to nonobtrusive pulse rate measurements. *J Biomed Opt* 17:077011–1–077011-4. doi:[10.1117/1.JBO.17.7.077011](https://doi.org/10.1117/1.JBO.17.7.077011)
31. Verkruyse W, Svaasand LO, Nelson JS (2008) Remote plethysmographic imaging using ambient light. *Opt Express* 16:21434–21445. doi:[10.1364/OE.16.021434](https://doi.org/10.1364/OE.16.021434)
32. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: *Proc CVPR*. p 1-511–1-518. doi: [10.1109/CVPR.2001.990517](https://doi.org/10.1109/CVPR.2001.990517)
33. Wu HY, Rubinstein M, Shih E, Guttag JV, Durand F, Freeman WT (2012) Eulerian video magnification for revealing subtle changes in the world. *ACM Trans Graph* 31:65–1–65–8. doi:[10.1145/2185520.2185561](https://doi.org/10.1145/2185520.2185561)
34. Zarzoso V, Comon P (2010) Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size. *IEEE Trans Neural Netw* 21:248–261. doi:[10.1109/TNN.2009.2035920](https://doi.org/10.1109/TNN.2009.2035920)
35. Zhang D, You X, Wang P, Yanushkevich SN, Tang YY (2009) Facial biometrics using nontensor product wavelet and 2D discriminant techniques. *Int J Pattern Recogn* 23:521–543. doi:[10.1142/S0218001409007260](https://doi.org/10.1142/S0218001409007260)
36. Zhang D, Han J, Han J, Shao L (2016) Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining. *IEEE Trans Neural Netw Learn Syst* 27:1163–1176. doi:[10.1109/TNNLS.2015.2495161](https://doi.org/10.1109/TNNLS.2015.2495161)



Lai-Man Po received the B.S. and Ph.D. degrees in electronic engineering from the City University of Hong Kong, Kowloon, Hong Kong, in 1988 and 1991, respectively. Since 1991, he has been with the Department of Electronic Engineering, City University of Hong Kong, and is currently an Associate Professor and lab director of TI Educational Training Centre. He has published over 160 technical journal and conference papers. His current research interests are in the areas of image/video coding, content-based video copy detection, image/video quality assessment and biomedical signal processing. Dr. Po was the chairman of the IEEE Signal Processing Hong Kong Chapter in 2012 and 2013. He is a member of the Technical Committee on Multimedia Systems and Applications, IEEE Circuits and Systems Society and an Associate Editor of the HKIE Transactions.



Litong Feng received his B.E. degree in Electronic Science and Technology from Harbin Institute of Technology, Harbin, China, in 2008, his M.E. degree in Optical Engineering from Tianjin Jinhang Institute of Technical Physics, Tianjin, China, in 2011, and his PhD degree in Electronic Engineering from City University of Hong Kong, Hong Kong, in 2016. His research interests include video processing for vital signs, machine learning, and optical system design.



Yuming Li received his B.Eng. degree from Huazhong University of Science and Technology in 2011, his M.Phil. degree from Huazhong University of Science and Technology in 2013. He is currently pursuing his Ph.D. degree in City University of Hong Kong. His research interests include image & video processing, multiscale analysis and machine learning.



Xuyuan Xu received his B.E. in Information Engineering in 2010 in City University of Hong Kong. His final year project “Stereoscopic Video generation from Monoscopic Video” won the Best Tertiary Student Project of Asia Pacific International and Communication Awards (APICTA) in 2010. He is currently a PhD student in the Department of Electronic Engineering in City University of Hong Kong. His research interests include 3D video coding and 3D view synthesis.



Chun-Ho Cheung received his BEng (Hons) degree in computer engineering and PhD degree in electronic engineering from City University of Hong Kong, in 1996 and 2002, respectively. Currently, he is an Assistant Professor at the Department of Information Systems, City University of Hong Kong. His research interests include image coding, motion estimation and e-Learning.



Kwok-Wai Cheung received B.E., M.S., and Ph.D. degrees, all in electronic engineering, from City University of Hong Kong in 1990, 1994, and 2001, respectively. He joined the Chu Hai College of Higher Education, Hong Kong, in 2002. Currently, he is an Associate Professor with the Department of Computer Science, Chu Hai College of Higher Education. His current research interests include the areas of image/video coding and multimedia database.