

Detecting Pulse Wave From Unstable Facial Videos Recorded From Consumer-Level Cameras: A Disturbance-Adaptive Orthogonal Matching Pursuit

Xuenan Liu, Xuezhi Yang[✉], Jing Jin, and Alexander Wong[✉]

Abstract—Objective: Modern consumer-level cameras can detect subtle changes in human facial skin color due to varying blood flow; they are beginning to be used as noncontact devices to detect pulse waves. Little, however, do we know about their capacity to perform pulse wave detection when the recorded faces are unstable. **Methods:** Here, we propose a novel method that can extract pulse waves from videos with drastic facial unsteadiness such as head twists and alternating expressions. The method first uses chrominance characteristics in multiple facial sub-regions to construct a raw pulse matrix. Subsequently, it employs a disturbance-adaptive orthogonal matching pursuit (DAOMP) algorithm to recover the underlying pulse matrix corrupted by facial unsteadiness. **Results:** To evaluate the efficacy of the method, we perform analyses on two datasets including 268 samples from 67 testing subjects. The results demonstrate that the proposed method outperforms state-of-the-art algorithms, especially in the terrain where drastic facial unsteadiness is present. **Conclusion:** The proposed framework shows promise to achieve video-based noncontact pulse wave detection from both steady and unsteady faces recorded by consumer-level cameras. **Significance:** By employing the proposed method, disturbance robustness in noncontact pulse wave detection can be significantly improved.

Index Terms—Imaging photoplethysmography, pulse wave detection, orthogonal matching pursuit.

Manuscript received August 16, 2019; revised February 23, 2020; accepted March 24, 2020. Date of publication April 3, 2020; date of current version November 20, 2020. This work was supported by the Anhui Major Science and Technology Projects (201903c08020010). (*Corresponding author: Xuezhi Yang.*)

Xuenan Liu is with the School of Computer and Information, Hefei University of Technology, and also with the Anhui Key Laboratory of Industry Safety and Emergency Technology.

Xuezhi Yang is with the School of Software, Hefei University of Technology, Hefei 230601, China, and also with the Anhui Key Laboratory of Industry Safety and Emergency Technology, Hefei 230601, China (e-mail: xzyang@hfut.edu.cn).

Jing Jin is with the School of Computer and Information, Hefei University of Technology.

Alexander Wong is with the Department of Systems Design Engineering, University of Waterloo.

Digital Object Identifier 10.1109/TBME.2020.2984881

I. INTRODUCTION

THE pulse wave (PW) is a vital sign which can be used to estimate pulse rate [1], [2], pulse rate variability [3], [4], oxyhemoglobin [5], [6], blood pressure [7], [8], as well as cardiovascular fitness [9], [10]. PWs are traditionally detected via specialized optical sensors or pressure sensors [11]–[13], but the contact between devices and skin may make users uncomfortable and inconvenient. Researches have shown that color of human skin changes subtly in response to the pulses of blood vessels [14]. These variations, especially in facial skin with the densely distributed capillaries, can be captured on videos using consumer-level cameras, offering a new approach to detect PWs in a noncontact manner. Based on this finding, a video-based biomonitoring technique called imaging photoplethysmography (IPPG) was proposed. Its ease of use, low cost and convenience have made it become an attractive research direction in computer vision and biomedical engineering communities [14]–[16]. However, PWs detected by the IPPG technique are sensitive to complex disturbances in real life, which has recently impeded its further development.

Pho *et al.* extracted PWs from facial videos by employing the independent component analysis (ICA) algorithm [17], and later improved their method by adding a band-pass filter in post-processing stage [3]. Nevertheless, their testing subjects had to stay motionless and expressionless in video recording. To alleviate constraints on testing subjects, Li *et al.* proposed a signal segmentation algorithm to extract clean PWs from facial videos containing face movements [18]. In dealing with expression changes, PW models built on multiple facial sub-regions were devised to locate the region of interest (ROI) automatically [19]–[21]. In addition to head movements and expression changes, illumination variations also form an obstacle to PW extraction. The fusion of multi-wavelength skin chrominance features demonstrated great performance in accommodating illumination variations [22]–[24]. More recently, Tulyakov *et al.* [16] and Nowara *et al.* [25] proposed a matrix completion algorithm in order to remove sparse spike noise in raw PWs. Previous approaches could deal with moderate facial unsteadiness (e.g., local expression changes, speaking and blinking), but failed to effectively address drastic facial unsteadiness (e.g., head twists and drastic expression changes).

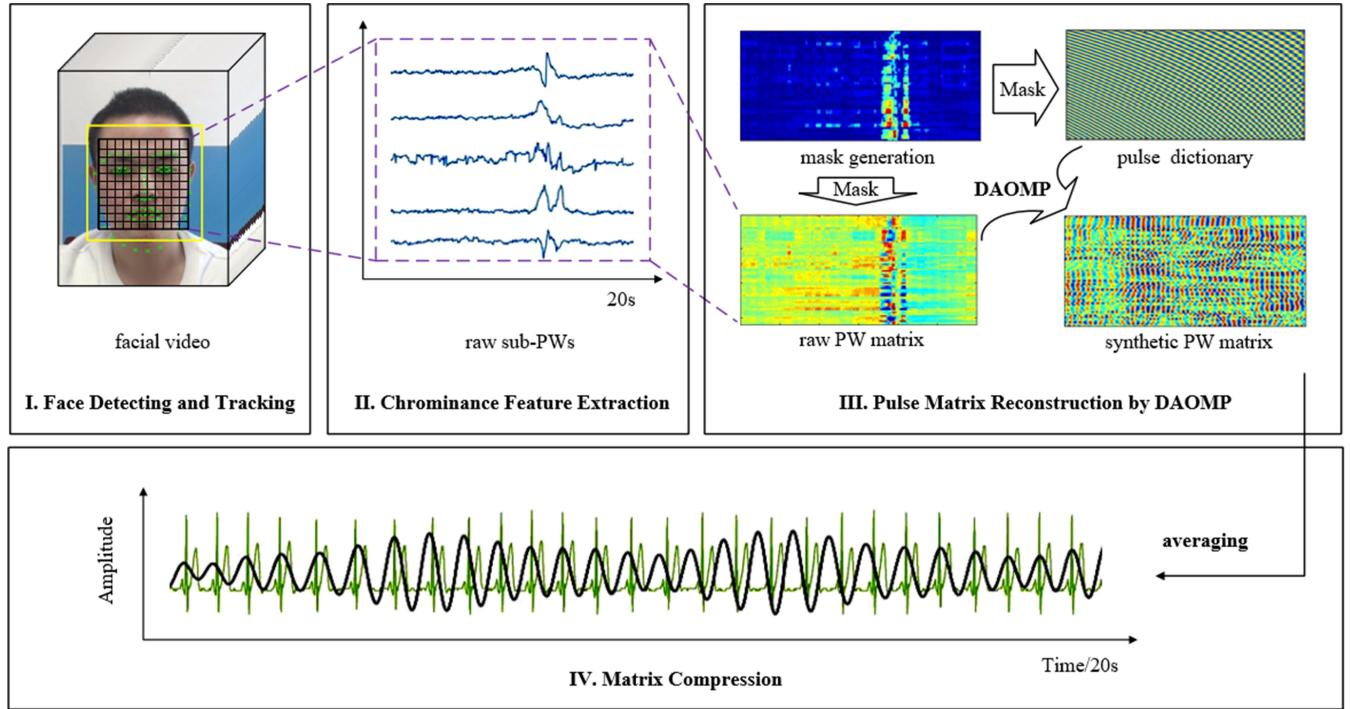


Fig. 1. Flowchart of the proposed method. (I) The face region is detected and tracked through video frames. (II) The face region is divided into several sub-regions, from which skin chrominance features are extracted to construct a set of raw pulse sub-PWs, i.e., raw PW matrix. (III) DAOMP is implemented on the raw PW matrix in order to construct a synthetic PW matrix which approximates underlying pulse components in the raw PW matrix. (IV) The synthetic PW matrix is ultimately compressed into a time series, i.e., synthetic PW.

The orthogonal matching pursuit (OMP) is a classical algorithm in signal de-noising [26], [27]. Taking PW features into consideration, a disturbance-adaptive orthogonal matching pursuit (DAOMP) algorithm is proposed in this paper to cope with drastic facial unsteadiness in noncontact PW detection. A pulse dictionary containing sinusoidal bases whose frequencies are constrained within the range of human pulse rates is created firstly. Afterwards, a sparse approximation of the raw PW is constructed by linearly combining sinusoidal bases in the pulse dictionary. In this process, a mask is imposed on the raw PW in order to mitigate the impact of the PW distortion due to drastic facial unsteadiness. The rest of this paper is structured as follows. The proposed method is elaborated in Section II, followed by the introduction about the experimental setup and results in Section III. In Section IV, the significance of the proposed method is discussed. This paper ends with conclusions in Section V.

II. THE PROPOSED METHOD

As shown in Fig. 1, the proposed method principally comprises four steps: facial detecting and tracking, chrominance feature extraction, pulse matrix reconstruction using DAOMP, and matrix compression.

A. Phase I: Face Detecting and Tracking

The subject's face in a video is detected and tracked in order to eliminate rigid face movements [18]. Firstly, a facial video is

recorded by a consumer-level camera in natural ambient lighting conditions. A rectangular facial region, as well as 66 facial landmarks, is then located by the discriminative response map fitting (DRMF) algorithm [28]. On this basis, the facial region is tracked along the moving trial of these landmarks following the Kanede-Lucas-Tomasi (KLT) algorithm [29].

B. Phase II: Chrominance Feature Extraction

Chrominance features are extracted from the facial video to construct a raw PW matrix. A variety of chrominance models for noncontact PW detection have been proposed in previous works [22]–[24]. In this paper, the model defined in [23] is adopted due to its strong robustness to facial movements. For each pixel in the video, a chrominance signal (S) is computed as the combination of two signals \mathbf{X}_f and \mathbf{Y}_f , i.e., $S = \mathbf{X}_f - \alpha\mathbf{Y}_f$, where $\alpha = \sigma(\mathbf{X}_f)/\sigma(\mathbf{Y}_f)$ and σ refers to the standard deviation. \mathbf{X}_f and \mathbf{Y}_f are the band-passed filtered versions of signals \mathbf{X}_s and \mathbf{Y}_s respectively, and \mathbf{X}_s and \mathbf{Y}_s are constructed by fusing three normalized color signals, \mathbf{R}_n , \mathbf{G}_n and \mathbf{B}_n , from RGB channels respectively, i.e.,

$$\begin{aligned} \mathbf{X}_s &= 3\mathbf{R}_n - 2\mathbf{G}_n \\ \mathbf{Y}_s &= 1.5\mathbf{R}_n + \mathbf{G}_n - 1.5\mathbf{B}_n, \end{aligned} \quad (1)$$

The combination coefficients are acquired by the skin-tone standardization approach (see [23] for details).

Subsequently, the facial region is divided into N square sub-regions (L pixels \times L pixels), as several sub-regions failing to

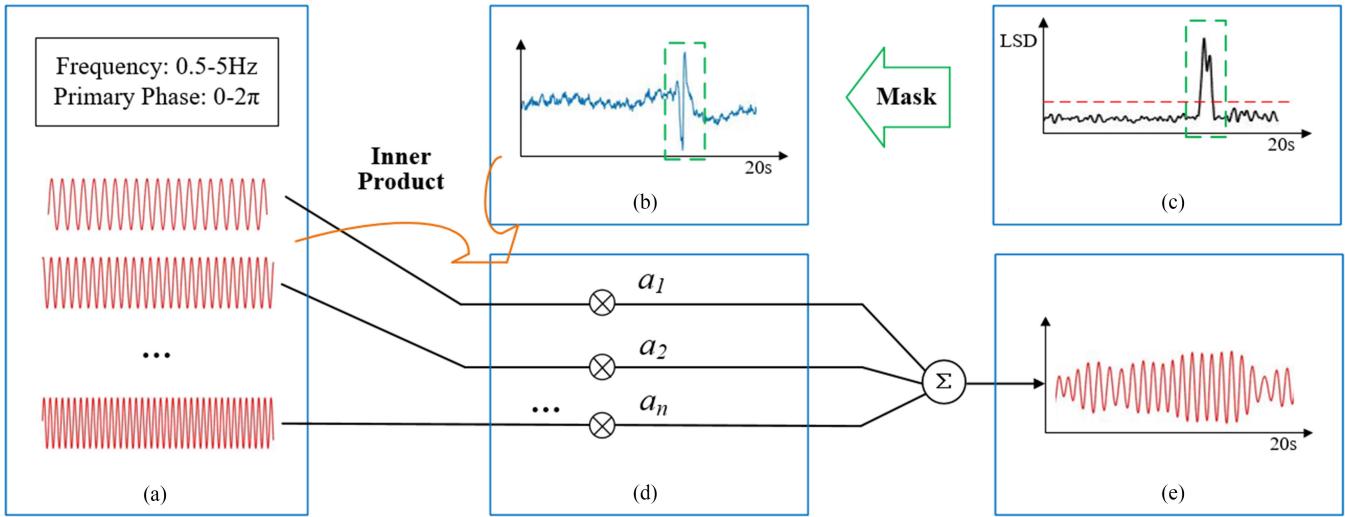


Fig. 2. Framework of the DAOMP algorithm. Sinusoidal bases that approximate the masked raw sub-PW are selected from the pulse dictionary to compose the synthetic sub-PW.

provide precise pulse information (e.g., boundary, skin fold, and hair) should be discarded in later steps. The sub-PW from the n^{th} sub-region (\mathbf{u}^n) can be obtained by implementing the spatial average on all the chrominance signals (\mathcal{S}) in this sub-region.

C. Phase III: Pulse Matrix Reconstruction by DAOMP

The DAOMP algorithm is performed on the raw sub-PWs to acquire an approximation of the underlying pulse components. A raw PW matrix (\mathbf{U}) is firstly constructed by arranging the raw sub-PWs in column, i.e., $\mathbf{U} = [\mathbf{u}^1, \mathbf{u}^2, \dots, \mathbf{u}^N]^T$, where the superscript T refers to matrix transposition. Normally, the raw PW matrix consists of an ideal pulse matrix ($\mathbf{U}_{\text{pulse}}$) and a noise matrix ($\mathbf{U}_{\text{noise}}$), i.e., $\mathbf{U} = \mathbf{U}_{\text{pulse}} + \mathbf{U}_{\text{noise}}$. The task of the DAOMP algorithm is to construct a synthetic PW matrix $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N]^T$ as an approximation of the ideal pulse matrix ($\mathbf{U}_{\text{pulse}}$). Based on the framework in Fig. 2, DAOMP will be elaborated in the rest of this sub-section.

It is known that a discrete signal can be represented as the linear combination of a number of sinusoidal bases with various amplitudes, frequencies and phases. As an approximation of the ideal PW, the n^{th} row of the synthetic PW matrix (\mathbf{x}^n) can therefore be composed of sinusoidal bases whose frequencies are constrained within the range of the human pulse rates (0.7–4 Hz). In view of this, a pulse dictionary (\mathbf{D}) is created as follows in order to store above-mentioned sinusoidal bases,

$$\begin{aligned} \mathbf{D}_{ij} &= \sin(2\pi k_i f / F_b + \phi_j) \\ \mathbf{D} &= [\mathbf{D}_{11}, \mathbf{D}_{12}, \dots, \mathbf{D}_{ij}, \dots, \mathbf{D}_{IJ}]^T, \end{aligned} \quad (2)$$

where \mathbf{D}_{ij} is a sinusoidal basis whose frequency and initial phase equal to k_i and ϕ_j respectively; k_i and ϕ_j denote the i^{th} frequency scale and j^{th} phase scale respectively; The gap between two frequency scales, k_i and k_{i+1} , is 1/60 Hz; the gap between two phase scales, ϕ_j and ϕ_{j+1} , is $\pi/100$; f refers to the frame series of a video; F_b stands for the frame rate of the video. Based on the pulse dictionary, the synthetic PW matrix (\mathbf{X}) can

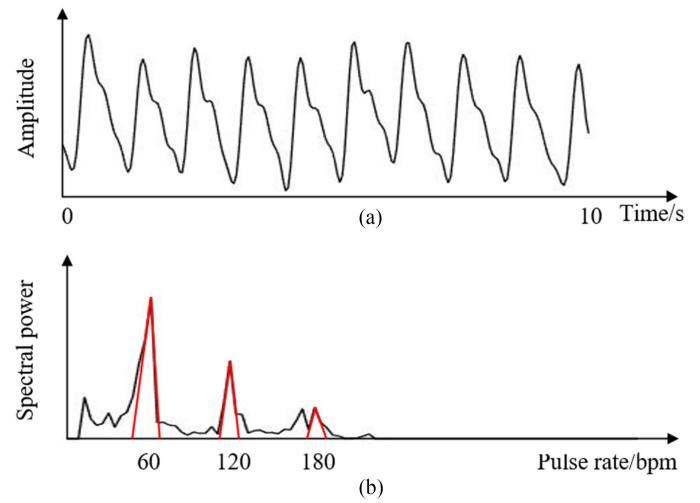


Fig. 3. Quasi-periodic nature of an ideal PW. In chart (a), intervals between adjacent crests of the PW roughly equal to each other in the time domain spectrum. In chart (b), the periodicity and dicrotic characteristics of the pulse waveform lead to predominant harmonics in the frequency domain spectrum.

be represented as

$$\mathbf{X} = \mathbf{a} \cdot \mathbf{D}, \quad (3)$$

where \mathbf{a} is the matching coefficient matrix.

As shown in Fig. 3, an ideal PW is of quasi-periodic nature and is close to a sinusoidal signal in shape. Thus, it can be represented as the weighted sum of a limited number of sinusoidal bases. Accordingly, the synthetic PW matrix (\mathbf{X}) can be constructed by combining a small amount of bases (\mathbf{D}_{ij}). In other words, the matching coefficient matrix (\mathbf{a}) is a sparse matrix containing only a few non-zero elements. The sparsity of the matching coefficient matrix (\mathbf{a}) can be modeled as the minimization of its 0-norm, i.e., $\min \|\mathbf{a}\|_0$. In addition, aiming to guarantee the fidelity of

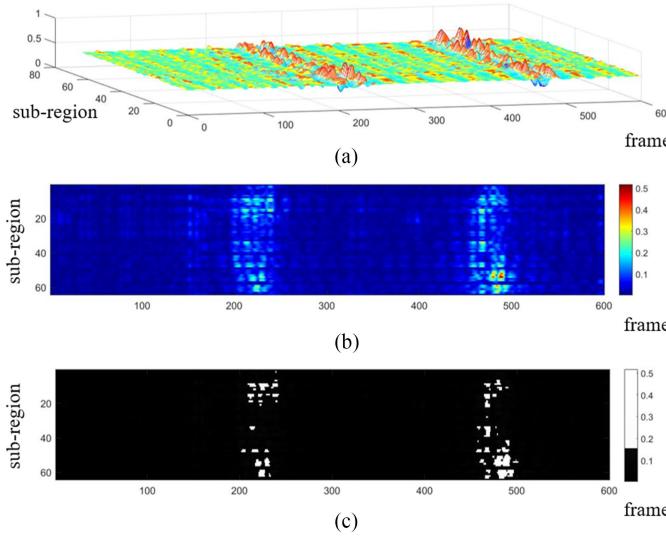


Fig. 4. Schematic diagram for the generation of the mask in one test. Chart (a) shows a raw PW matrix, whose LSD matrix is displayed in chart (b). Chart (c) presents the mask generated according to LSD.

synthetic PW (\mathbf{X}), the error between the synthetic PW matrix (\mathbf{X}) and the raw PW matrix (\mathbf{U}) should be constrained at a low level. The fidelity of synthetic PW (\mathbf{X}) can be modeled as the minimization of the error 2-norm, i.e., $\min \|\mathbf{X} - \mathbf{U}\|_2$. Based on above analysis, the construction of the synthetic PW matrix (\mathbf{X}) can be formulated as the following minimization problem:

$$\min_{\mathbf{X}} \|\mathbf{X} - \mathbf{U}\|_2 + \mu \|\alpha\|_0, \quad (4)$$

where μ is the regularization parameter used to balance the fidelity term $\|\mathbf{X} - \mathbf{U}\|_2$ and the regular term $\|\alpha\|_0$.

The video-based PW is sensitive to drastic facial unsteadiness in real life. Inspired by [16], a mask matrix denoted by $\mathbf{M} = [\mathbf{m}^1, \mathbf{m}^2 \dots \mathbf{m}^N]^T$ is adopted to discard the disturbed elements of the raw PW matrix (\mathbf{U}), as presented in Fig. 4. Specifically, local standard deviation (LSD) of each element of the raw PW matrix (\mathbf{U}) over eight neighbors is calculated firstly. Then the elements whose LSDs exceed a threshold (Q_T) are covered by the mask matrix (\mathbf{M}). Thus, (4) is changed to

$$\begin{cases} \min_{\mathbf{X}} \|\mathbf{M} \circ (\mathbf{X} - \mathbf{U})\|_2 + \mu \|\alpha\|_0 \\ \mathbf{M} = \frac{1}{2} \{\text{sgn}[Q_T - \text{LSD}(\mathbf{U}^\tau)] + 1\} \end{cases}, \quad (5)$$

where \circ is the Hadamard product, and the superscript τ refers to the normalization. Elements of the mask matrix (\mathbf{M}) are set to 0 if corresponding LSDs top the threshold (Q_T), and set to 1 otherwise.

Equation (5) is a NP-hard problem, but it can be solved by a greedy algorithm, DAOMP. Taking the first masked sub-PW ($\mathbf{m}^1 \circ \mathbf{u}^1$) into consideration, sinusoidal bases close to this sub-PW are selected from the pulse dictionary to compose a synthetic sub-PW. It is known that the inner product of two signals reflects their similarity [26]. Thus, the inner product (p_{ij}) of the sinusoidal basis and the masked sub-PW is calculated as

follows to estimate their similarity,

$$p_{ij} = \langle \mathbf{m}^1 \circ \mathbf{u}^1, \mathbf{D}_{ij} \rangle. \quad (6)$$

The sinusoidal basis with the maximal $p_{ij}(\mathbf{D}_1^c)$ is the basis closest to the masked sub-PW. The matching coefficient (a_1) between this basis and the masked sub-PW can be computed as follows,

$$a_1 = \langle \mathbf{m}^1 \circ \mathbf{u}^1, (\mathbf{D}_1^c)^g \rangle, \quad (7)$$

where the superscript g stands for pseudo-inverse operation. Afterwards, the residual vector (\mathbf{r}_1) between this basis and the sub-PW can be obtained by following equation

$$\mathbf{r}_1 = \mathbf{m}^1 \circ (\mathbf{u}^1 - a_1 \cdot \mathbf{D}_1^c), \quad (8)$$

and the matching error (e_1) can be formulated as

$$e_1 = \|\mathbf{r}_1^\tau\|_2 / \sqrt{\|\mathbf{m}^1\|_0}. \quad (9)$$

Similar to the sub-PW (\mathbf{u}^1), the residual vector (\mathbf{r}_1) can further be linearly decomposed into a basis (\mathbf{D}_2^c) and a residual vector (\mathbf{r}_2),

$$\mathbf{r}_1 = a_2 \cdot \mathbf{D}_2^c + \mathbf{r}_2, \quad (10)$$

where \mathbf{D}_2^c is obtained in the same way with \mathbf{D}_1^c , and the matching coefficient a_2 is computed with the update of a_1 , i.e.,

$$[a_1, a_2] = [\mathbf{m}^1 \circ \mathbf{u}^1] \cdot \begin{bmatrix} \mathbf{D}_1^c \\ \mathbf{D}_2^c \end{bmatrix}^g. \quad (11)$$

This process is iteratively implemented until the matching error of the k^{th} iteration (e_k) drops below the threshold (E_T), or the number of iterations exceeds the threshold (K). When the iteration satisfies the stop criteria, the synthetic sub-PW (\mathbf{x}^1) can be represented as the linear combination of several sinusoidal bases, i.e.,

$$\mathbf{x}^1 = \sum_{k=1}^K a_k \cdot \mathbf{D}_k^c. \quad (12)$$

Other sub-PWs $\{\mathbf{x}^2, \mathbf{x}^3 \dots \mathbf{x}^N\}$ are constructed in the same manner as the first sub-PW (\mathbf{x}^1). Filling matching coefficients into the corresponding places of the coefficient matrix (\mathbf{a}), the synthetic PW matrix (\mathbf{X}) can be constructed by (3).

D. Phase IV: Matrix Compressing

The synthetic PW is finally constructed by averaging the sub-PWs in the synthetic PW matrix (\mathbf{X}). However, some non-homogenous facial sub-regions cannot provide precise pulse information. In order to remove the irregular sub-PWs extracted from the non-homogenous sub-regions, the sub-PW (\mathbf{u}^n) whose matching error (e_k) fails to decrease to the error threshold (E_T) when the number of iterations meets the upper limit (K) is discarded. Thus, the synthetic PW (\mathbf{x}_{ave}) can be represented as the average of the remaining sub-PWs, i.e.,

$$\begin{aligned} \mathbf{x}_{ave} &= 1/N \times \sum_{n=1}^N \mathbf{x}^n, \\ \text{s.t. } e_K^n &< E_T \end{aligned} \quad (13)$$

TABLE I
INFORMATION ON SUBJECTS IN TWO DATASETS

| Dataset | Subjects | | |
|------------|----------|-------|--------|
| | Gender | Age | Number |
| MAHNOB-HCI | Male | 19-40 | 12 |
| | Female | 19-40 | 15 |
| IPPG-PW | Male | 18-60 | 28 |
| | Female | 24-36 | 12 |

where e_K^n refers to the matching error of the n^{th} sub-PW in the K^{th} iteration.

III. TESTING AND RESULTS

This section firstly presents information about the experimental datasets, evaluation metrics and parameter settings, followed by the detailed introduction about experimental results.

A. Datasets

As shown in **Table I**, two datasets were used in our experiments. The one was the publicly available MAHNOB-HCI dataset provided by Soleymani *et al.* [30]. The other was the IPPG-PW dataset established by the authors' research team, which is available upon request.

The MAHNOB-HCI is a multi-modal dataset including videos in two categories: emotion elicitation data and implicit tagging data. It contains 27 subjects (15 females and 12 males) whose ages range from 19 to 40. Twenty frontal face videos were recorded for each subject at 60 frame per second (fps) with resolution of 780×580 . Electrocardiograph (ECG) of the subjects was recorded synchronously serving as the ground truth data. 108 facial videos (2 static videos and 2 dynamic videos for each subject) were randomly selected from the MAHNOB-HCI dataset, and all the videos lasted for 20 seconds.

The IPPG-PW dataset comprises 160 videos from 40 subjects (12 females and 28 males). Four videos were recorded for each subject, including two static videos where the subject kept motionless and expressionless, and two dynamic videos where the subject turned head and changed expressions freely. The dynamic videos can be further divided into the moderately dynamic videos (involving blink, smile, and frown) and the drastically dynamic videos (involving head twists and alternating expressions). Detailed information about the videos in the IPPG-PW dataset is given in **Table II**.

Facial videos in the IPPG-PW dataset were recorded at 30 fps with resolution of 640×480 in natural ambient lighting conditions. All videos lasted for 20 seconds. **Fig. 5** illustrates the video recording scenario, where the subject was sitting in front of a fixed camera (Logitech C920), with her face around 50 centimeters far from the camera lens. Three-lead ECG, regarded as the ground truth, was recorded synchronously with the facial videos using a portable device called Heal Force Electrocardiograph (see <http://www.healforce.com/cn> for details).

TABLE II
THE NUMBER OF VIDEOS IN THE IPPG-PW DATASET

| D | M | 0 | 1 | 2 | 3 | 4 |
|-----|-----|----|---|---|---|---|
| V | D | 80 | 3 | 3 | 2 | 0 |
| 0 | 0 | 80 | 3 | 3 | 2 | 0 |
| 1 | 1 | 9 | 9 | 3 | 4 | 1 |
| 2 | 2 | 15 | 8 | 4 | 0 | 0 |
| 3 | 3 | 6 | 4 | 1 | 0 | 0 |
| 4 | 4 | 3 | 3 | 1 | 0 | 1 |

M : the amount of moderate disturbances in one video; D : the amount of drastic disturbances in one video; V : the amount of videos containing M moderate disturbances and D drastic disturbances. (e.g., $V(8)-M(1)-D(2)$ means there are eight videos including one moderate disturbance and two dynamic disturbances.)

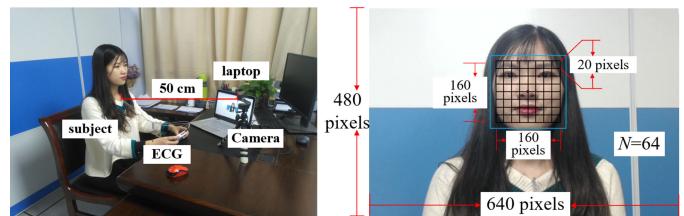


Fig. 5. The testing scenario of the IPPG-PW dataset.

All recorded data were then transmitted to a laptop (Hewlett-Packard) and processed using MATLAB of version 2014b on Window 10 system.

B. Evaluation Metrics

Average pulse rate has been frequently used in previous papers [16]–[18] as an indicator to evaluate noncontact PW detection algorithms. In consistence with [16], we acquired average pulse rate by detecting the peak of PW frequency spectrum. However, the average pulse rate failed to reflect the consistency between the estimated PW and the ground truth in detail. In order for a more comprehensive evaluation, inter-beat-interval (IBI) was used as another indicator. **Fig. 6** presents IBIs of a video-based PW and the corresponding ECG. Inflection points of each polyline segment are aligned with the peaks of the physiological signals, and the height of each polyline segment equals to the value of the current IBI. Intuitively, fluctuation of polylines reflects variations of IBIs. Four statistics based on the average pulse rate and IBI respectively formed the evaluation metrics.

- S1. Pulse rate error of the v^{th} sample.

$$PE_v = PR_{\text{PW}}^v - PR_{\text{ECG}}^v, \quad (14)$$

where PR_{PW}^v and PR_{ECG}^v refer to the average pulse rate extracted from the video-based PW and the ECG respectively, and v means the v^{th} testing sample.

- S2. Absolute mean of pulse rate error.

$$|M_{PE}| = \sum_{v=1}^V |PE_v| / V, \quad (15)$$

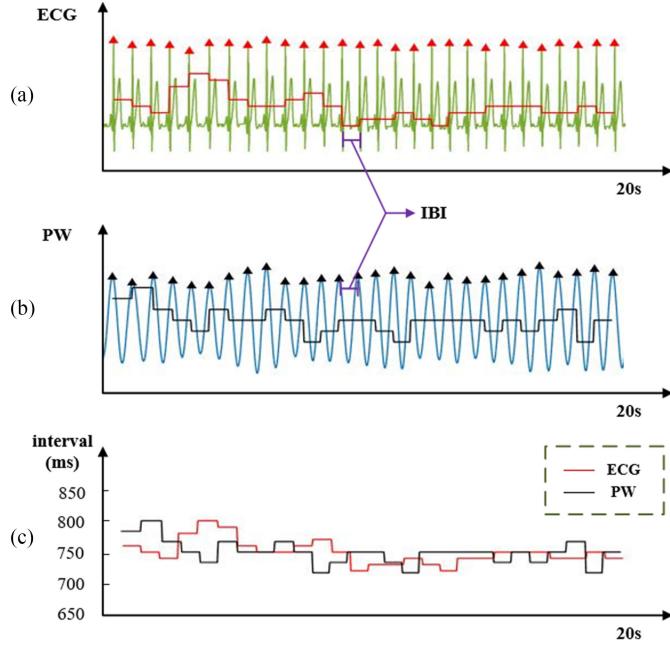


Fig. 6. Chart (a): ECG and its IBI polyline. Chart (b): video-based PW and its IBI polyline. Chart (c): agreement in IBIs between the ECG and the video-based PW.

where V is the amount of testing samples.

- S3. Standard deviation of pulse rate error.

$$SD_{PE} = \sqrt{\sum_{v=1}^V (PE_v - M_{PE})^2 / V}. \quad (16)$$

- S4. Accuracy of pulse rates.

$$PP = 1 - \sum_{v=1}^V |PE_v| / PR_{ECG}^v / V. \quad (17)$$

- S1. IBI error for the v^{th} sample.

$$AE_v = E(|R_{PW}^v - R_{ECG}^v|),$$

where R_{PW}^v and R_{ECG}^v stand for the IBI series of the v^{th} video-based PW and the corresponding ECG, respectively, and E refers to expectation.

- S2. Mean of IBI error.

$$M_{AE} = \sum_{v=1}^V AE_v / V. \quad (18)$$

- S3. Standard deviation of IBI error.

$$SD_{AE} = \sqrt{\sum_{v=1}^V (AE_v - M_{AE})^2 / V}. \quad (19)$$

- S4. Accuracy of IBIs.

$$AD = 1 - \sum_{v=1}^V AE_v / (T / (B - 1)) / V, \quad (20)$$

where B stands for the amount of pulsations in a PW sample.

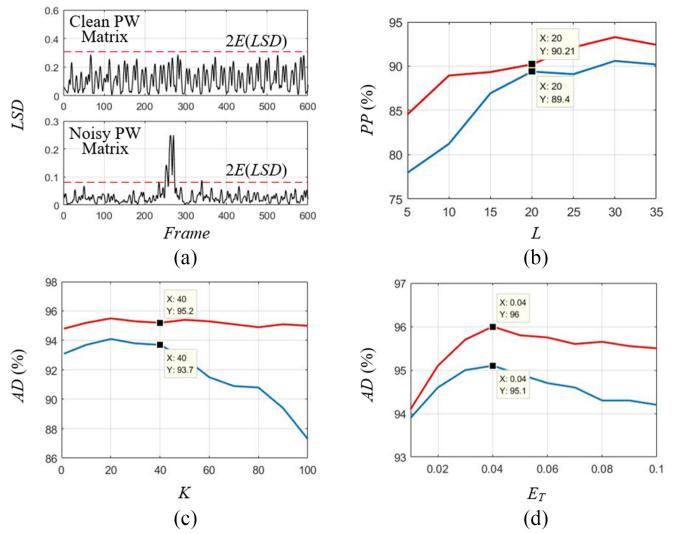


Fig. 7. Parameterization of the proposed method. Chart (a): Cross-sections of LSD of two PW matrix. Chart (b), (c), (d): Performance of the proposed method at varying L , K and E_T , respectively, where red curves mean the static scenario, and blue curves stand for the dynamic scenario.

C. Parameter Settings

The proposed method involves four adjustable parameters: the size of sub-regions ($L \times L$), the threshold of LSD (Q_T), the maximum number of iterations (K) and the threshold of matching error (E_T). Fig. 7(a) illustrates a LSD cross-section of a clean PW matrix and a LSD cross-section of a distorted PW matrix. In general, the maximal LSD of a clean PW matrix was less than twice as large as the averaged LSD . In contrast, elements of a PW matrix corrupted by drastic facial unsteadiness had $LSDs$ significantly more than twice the averaged LSD . Therefore, the double mean of the LSD was leveraged as the threshold to judge whether the pulse matrix involved drastic distortion, i.e., $Q_T = 2E(LSD)$. Values of L , K and E_T were investigated based on all testing samples in both datasets and were determined according to the best experimental results. L is a parameter in face area segmentation, whereas K and E_T are associated with PW post-processing. Since the face area segmentation is independent with the PW post-processing, we studied the optimal value of L without performing the DAOMP in order to avoid the confusion from K and E_T (we extracted pulse rates from all sub-regions and computed their mean as the final testing results). In Fig. 7(b), PP increased with the growth of L , as the quantization noise of the camera sensor was reduced with the expansion of sub-regions. The uptrend of PP was remarkable at the beginning and began to level off after L exceeded 20 pixels. Considering that the overlarge facial sub-region reduced the flexibility of sub-PW selection, we finally set L to 20 pixels. Fig. 7(c) presents the performance of the proposed method at varying K in two scenarios without regard to E_T (i.e., $E_T \rightarrow \infty$). It can be seen that the acceptable values of K ranged from 20 to 40, as the fidelity of the synthetic PW failed to meet requirements given an excessively small K , whereas synthetic PWs might be affected by noise if the value of K was higher than 40. The

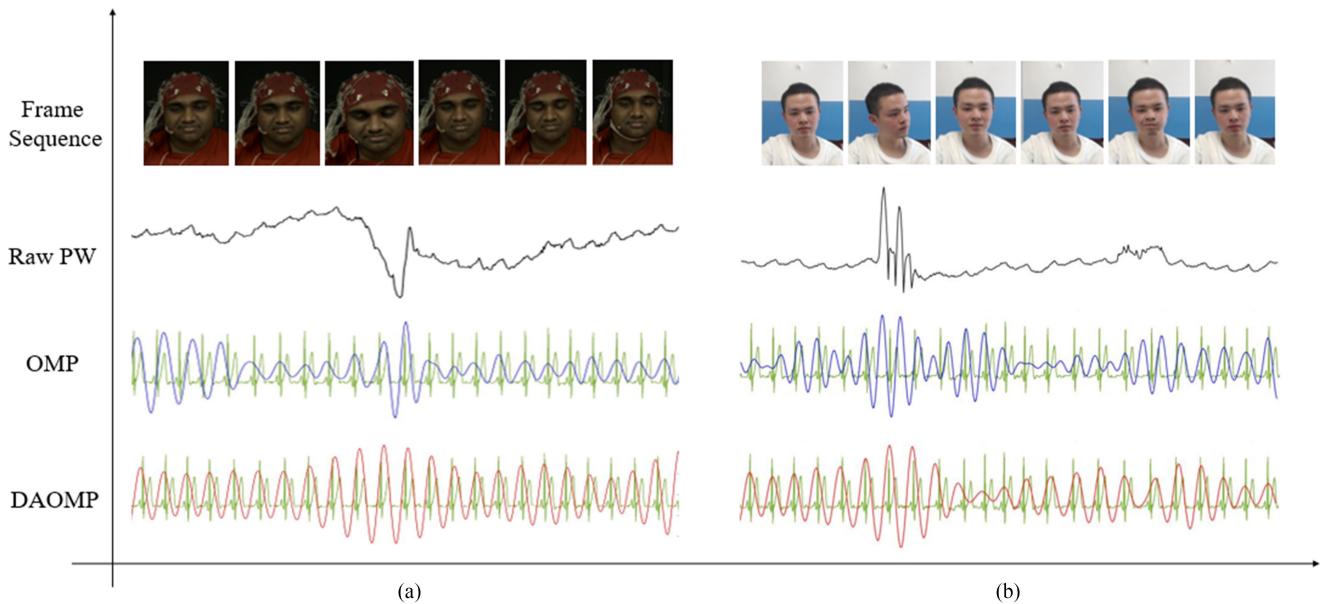


Fig. 8. Performance comparison between the OMP and DAOMP.

TABLE III

PERFORMANCE OF THE PROPOSED METHOD AT VARYING FRAME RATES AND VARYING VIDEO LENGTHS

| Frame rate (fps) | $ M_{PE} $ (bpm) | PP (%) | M_{AE} (ms) | AD (%) |
|------------------|------------------|--------|---------------|--------|
| 5 | 2.12 | 97.06 | 46.07 | 94.24 |
| 10 | 1.76 | 97.56 | 44.69 | 94.41 |
| 15 | 1.39 | 98.01 | 43.68 | 94.73 |
| 30 | 1.34 | 98.14 | 41.00 | 95.06 |

| Video length (s) | $ M_{PE} $ (bpm) | PP (%) | M_{AE} (ms) | AD (%) |
|------------------|------------------|--------|---------------|--------|
| 5 | 4.56 | 93.92 | 94.46 | 88.62 |
| 10 | 2.07 | 97.41 | 45.98 | 94.48 |
| 15 | 1.82 | 97.73 | 44.56 | 94.54 |
| 20 | 1.34 | 98.14 | 41.00 | 95.06 |

optimal value of K was finally set to 20 in order to reduce the computational burden. Fig. 7(d) shows the relationship between AD and E_T supposing that $K = 20$. AD in both testing scenarios peaked when E_T equaled to 0.04. It is because a small E_T might remove the useful sub-PW by mistake, whereas a large E_T was likely to accept poor-quality sub-PWs.

D. Performance Testing

The proposed method was tested in diverse conditions, including varying frame rates, video lengths and mask coverage rates.

Effect of frame rates and video lengths: To study the performance of the proposed method at varying frame rates, the frame rate of all dynamic videos in both datasets was decreased to 20 fps, 10 fps and 5 fps, with video length fixed at 20 s. As shown in Table III, four statistics almost kept unchanged with the

decrease of the frame rate, indicating that the proposed method applied to cameras with low frame rates. In addition, in order to study the impact of the video lengths, all videos in both datasets were cut into 15 s, 10 s and 5 s, with their frame rate fixed at 30 fps. Table III presents that the performance of the proposed method declined dramatically when the video length dropped to 5 s. Therefore, the proposed method failed to work effectively supposing that the video length was less than 5 s. It is because an excessively short video could not provide sufficient information for DAOMP to reconstruct a clean PW matrix.

Effect of the mask: In order to study the improvement made by the mask in DAOMP, we extracted PWs from all dynamic videos in both datasets using the OMP algorithm (OMP does not include the mask). Testing results of two samples illustrated in Fig. 8 demonstrated significance of the mask. In the sample from the MAHNOB-HCI dataset, the subject shook his head at the 7th second, leading to a sharp distortion of the raw PW at the corresponding moment. Likewise, in the sample from the IPPG-PW dataset, the subject turned his head at the 5th seconds and changed expression at about the 14th second, inducing a spike distortion and a modest distortion of the raw PW. Intuitively, in both samples, the synthetic PW constructed by DAOMP presented greater consistence with ECG compared to the counterpart obtained by OMP.

Overall comparison results between OMP and DAOMP are shown in Table IV. DAOMP increased PP and AD by 5.49% and 17.53% respectively compared to OMP, indicating that the mask in DAOMP improved the quality of synthetic PWs effectively. Table IV also presents the performance of the proposed method at varying mask coverage rates. The acceptable mask rates ranged from 10 to 70 percentages. There was a trade-off in choosing the mask coverage rate, as an oversized mask might cover the useful part of the raw PW matrix, whereas an excessively small mask might fail to cover all distorted elements of the raw PW matrix.

TABLE IV
EFFECT OF THE MASK ON THE PROPOSED METHOD

| Scheme | $ M_{PE} $ (bpm) | PP (%) | M_{AE} (ms) | AD (%) |
|----------------|------------------|--------|---------------|--------|
| OMP | 5.58 | 92.65 | 182.34 | 77.53 |
| DAOMP | 1.34 | 98.14 | 41.00 | 95.06 |
| Mask rates (%) | 0 | 10 | 20 | 30 |
| AD (ms) | 72.14 | 88.14 | 94.74 | 91.53 |
| Mask rates (%) | 50 | 60 | 70 | 80 |
| AD (ms) | 85.23 | 78.46 | 72.15 | 65.35 |
| | | | | 57.65 |

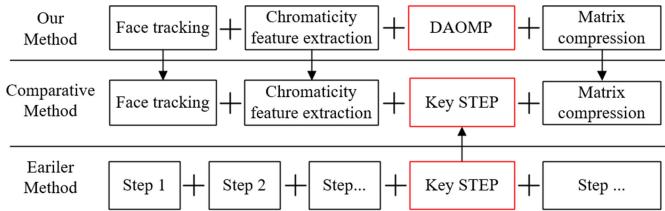


Fig. 9. Composition of the comparative method.

E. Comparative Experiments

To highlight the performance of the proposed method, we conducted data analysis on the same datasets using competing, state-of-the-art methods from Richard *et al.* [31], Chwyl *et al.* [21], Wang *et al.* [19] and Tulyakov *et al.* [16]. For consistency purpose, we extracted the key steps from each competing method and combined it with the proposed method, excluding the step DAOMP (See Fig. 9 for details, where half of the sub-PWs with the largest SNR (signal noise ratio) were chosen in comparative methods, see [21] for the computing process of the SNR).

Testing in the static scenario: Table V shows the performance of five methods in the static testing scenario where subjects kept motionless and expressionless. For the MAHNOB-HCI dataset, $|M_{PE}|$ and SD_{PE} of all the methods were below 0.6 bpm and 1.0 bpm respectively, indicating that all the methods could measure average pulse rates accurately in the static scenario. Besides, IBIs extracted by all the methods were highly consistent with the ground truth, with M_{AE} and SD_{AE} lower than 35 ms and 15 ms respectively. Similar conclusions could be drawn from the experimental results based on the IPPG-PW dataset. To summarize, all methods performed well in estimating pulse rates and IBIs in the static testing scenario.

Testing in the dynamic scenario: Video-based PWs are sensitive to facial unsteadiness such as head twists and alternating expressions in the dynamic testing scenario. As a result, the performance of the state-of-the-art methods declined dramatically as shown in Table VI. SAMC [16] had the best experimental results among the state-of-the-art methods on the MAHNOB-HCI dataset. However, with a low AD at a mere 81.36%, its performance in IBI estimation was not satisfactory. The performance of the state-of-the-art methods saw more significant declines on the IPPG-PW dataset, as the IPPG-PW dataset contained more drastically dynamic videos than the MAHNOB-HCI dataset. With M_{AE} topping 160 ms and AD

less than 80%, these methods presented insufficient robustness to the drastic facial unsteadiness.

On the other hand, there was a modest gap in the performance of the proposed method between the static scenario and the dynamic scenario. As shown in Table VI, PP and AD of the proposed method fell by 1.25% and 0.78% respectively on the MAHNOB-HCI dataset, and decreased by 1.29% and 1.65% respectively on the IPPG-PW dataset. On the IPPG-PW dataset, with $|M_{PE}|$ at 1.55 bpm and SD_{PE} at 2.30 bpm, the proposed method satisfied the precision requirement of daily pulse rate detection (i.e., the pulse rate error is less than 5 bpm [18]). In addition, with AD at 94.27%, the proposed method was capable of constructing PWs that had IBIs approximate to the ground truth. The proposed method also demonstrated validity on the MAHNOB-HCI dataset, as all the evaluation indicators were better than the counterparts on the IPPG-PW dataset. In conclusion, the proposed method outperformed the state-of-the-art methods in the dynamic testing scenario.

F. Stepwise Testing

The proposed method was tested step-by-step on both datasets in order to study the contribution made by each individual step. The testing results are presented in Table VII. DAOMP increased AD by up to 14.08%, in contrast to 6.29%, 3.70% and 4.49% due to other steps. Besides, it also saw the largest increase in PP (4.81%) among all the steps. The chromaticity model with multiple wavelengths (in CFE) could alleviate the illumination-induced noise; the face tracking (in FDT) could deal with planar facial movements; the facial area segmentation (in SC) was able to remove the noise attributable to local expression changes. However, they failed to address drastic facial unsteadiness which had a remarkable impact on video-based PWs. On the opposite, based on the sparse feature of PWs, DAOMP removed the sharp distortion arising from drastic facial unsteadiness and thus played a key role in improving the accuracy of PW detection.

IV. DISCUSSION

Disturbance robustness in noncontact detection of PWs has been shown in earlier works by combining ROI-trackers [16], [18], [21], multiple wavelength techniques [22]–[24], multiple sub-regions of skin [19]–[21], and post-processing [2], [16], [25]. However, drastic facial unsteadiness, such as head twists and global expression changes, is still a huge challenge for PW detection. With the results of a comparative experiment depicted in Fig. 10, the abilities of the proposed method, as well as five state-of-the-art methods, to deal with drastic facial unsteadiness are to be analyzed. In Fig. 10(a), there were two distortions of the raw PW induced by subject's head twists. Li *et al.* [18] cut off the distorted parts directly and then spliced signal segments end-to-end. However, as shown in Fig. 9(b), it was difficult to joint these ends perfectly as they might not be on the same level. Besides, Richard *et al.* [3] extracted a clear PW from the raw PW using an improved ICA algorithm. The distortion was however not be filtered out completely as shown in Fig. 10(c), as only three input signals (RGB signals) failed to provide sufficient

TABLE V
PERFORMANCE OF FIVE METHODS IN THE STATIC TESTING SCENARIO

| Database | Method | Key Algorithm | Pulse Rate | | | IBI | |
|------------|-----------------------------|---------------|------------------|-----------------|--------|---------------|----------------|
| | | | $ M_{PE} $ (bpm) | SD_{PE} (bpm) | PP (%) | M_{AE} (ms) | SD_{AE} (ms) |
| MAHNOB-HCI | Richard <i>et al.</i> [31] | C-ICA | 0.50 | 0.84 | 99.26 | 32.25 | 10.68 |
| | Wang <i>et al.</i> [19] | SP | 0.41 | 0.66 | 99.40 | 31.34 | 12.20 |
| | Chwyl <i>et al.</i> [21] | BM | 0.50 | 0.93 | 99.36 | 32.67 | 11.19 |
| | Tulyakov <i>et al.</i> [16] | SAMC | 0.43 | 0.64 | 99.42 | 32.33 | 11.67 |
| | Our method | DAOMP | 0.39 | 0.57 | 99.54 | 31.07 | 12.69 |
| IPPG-PW | Richard <i>et al.</i> [31] | C-ICA | 0.59 | 0.61 | 99.21 | 35.46 | 10.63 |
| | Wang <i>et al.</i> [19] | SP | 0.60 | 0.63 | 99.19 | 35.65 | 12.42 |
| | Chwyl <i>et al.</i> [21] | BM | 0.61 | 0.60 | 99.14 | 35.79 | 12.02 |
| | Tulyakov <i>et al.</i> [16] | SAMC | 0.53 | 0.65 | 99.26 | 35.30 | 8.31 |
| | Our method | DAOMP | 0.50 | 0.64 | 99.35 | 33.58 | 10.12 |

C-ICA: Constrained Independent Component Analysis; SP: Spatial Pruning; BM: Bayesian Minimization; SAMC: Self-adaptive Matrix Completion.

TABLE VI
PERFORMANCE OF FIVE METHODS IN THE DYNAMIC TESTING SCENARIO

| Database | Method | Key Algorithm | Pulse Rate | | | IBI | |
|------------|-----------------------------|---------------|------------------|-----------------|--------------|---------------|----------------|
| | | | $ M_{PE} $ (bpm) | SD_{PE} (bpm) | PP (%) | M_{AE} (ms) | SD_{AE} (ms) |
| MAHNOB-HCI | Richard <i>et al.</i> [31] | C-ICA | 5.25 | 5.04 | 92.60 | 150.59 | 48.59 |
| | Wang <i>et al.</i> [19] | SP | 4.94 | 5.99 | 93.62 | 158.73 | 57.51 |
| | Chwyl <i>et al.</i> [21] | BM | 4.67 | 5.06 | 93.84 | 149.61 | 52.24 |
| | Tulyakov <i>et al.</i> [16] | SAMC | 3.98 | 5.04 | 94.55 | 135.45 | 46.65 |
| | Our method | DAOMP | 1.26 | 1.82 | 98.29 | 40.25 | 13.19 |
| IPPG-PW | Richard <i>et al.</i> [31] | C-ICA | 5.61 | 6.41 | 92.52 | 176.74 | 51.19 |
| | Wang <i>et al.</i> [19] | SP | 5.26 | 6.84 | 93.14 | 167.65 | 55.98 |
| | Chwyl <i>et al.</i> [21] | BM | 4.78 | 6.29 | 93.68 | 161.87 | 49.14 |
| | Tulyakov <i>et al.</i> [16] | SAMC | 8.21 | 7.90 | 88.76 | 286.05 | 79.90 |
| | Our method | DAOMP | 1.55 | 2.30 | 98.06 | 42.49 | 14.36 |

TABLE VII
TESTING RESULTS OF EACH STEP OF THE PROPOSED METHOD

| Step | $ M_{PE} $ (bpm) | PP (%) | M_{AE} (ms) | AD (%) |
|------------------------------|------------------|--------------|---------------|--------------|
| None | 8.25 | 88.54 | 235.35 | 67.50 |
| FDT | 5.96 | 91.72 | 175.64 | 73.79 |
| FDT&CFE | 5.57 | 92.26 | 170.69 | 76.49 |
| FDT&CFE&DAOMP | 4.23 | 97.07 | 70.45 | 90.57 |
| FDT&CFE&DAOMP&MC | 1.34 | 98.14 | 41.00 | 95.06 |

FDT: Face Detecting and Tracking; CFE: Chromaticity Feature Extraction; SC: Matrix Compression.

information for the ICA algorithm. Sub-regions of interest could be selected automatically according to the stability of facial skin [19] or the SNR of sub-PWs [21]. Nevertheless, the drastic facial unsteadiness was likely to disturb the entire facial area,

leading to the invalidation of these methods as presented in Fig. 9(d) and (e). The method proposed by Tulyakov *et al.* [16] removed the abnormal elements of the raw PW matrix and then made up the incomplete PW matrix following the self-adaptive matrix completion algorithm. Unfortunately, under the goal of minimizing the rank of the completed PW matrix, the PW matrix might be filled in with zero elements supposing that the irregular elements occupied entire rows or columns of the raw PW matrix, thus aggravating the distortion of the detected PW as shown in Fig. 10(f). In sum, drastic facial unsteadiness had not been addressed effectively by state-of-the-art methods.

The DAOMP algorithm was proposed in this paper to cope with the drastic facial unsteadiness. The disturbed elements of the raw PW matrix were discarded by a mask firstly. Several sinusoidal bases close to the masked raw PW were then selected from the pulse dictionary to synthesize a sparse approximation of the underlying pulse components. In Fig. 10(g), the IBI polyline of the video-based PW (in red) and the IBI polyline of ECG (in green) twined with each other closely, reflecting that the

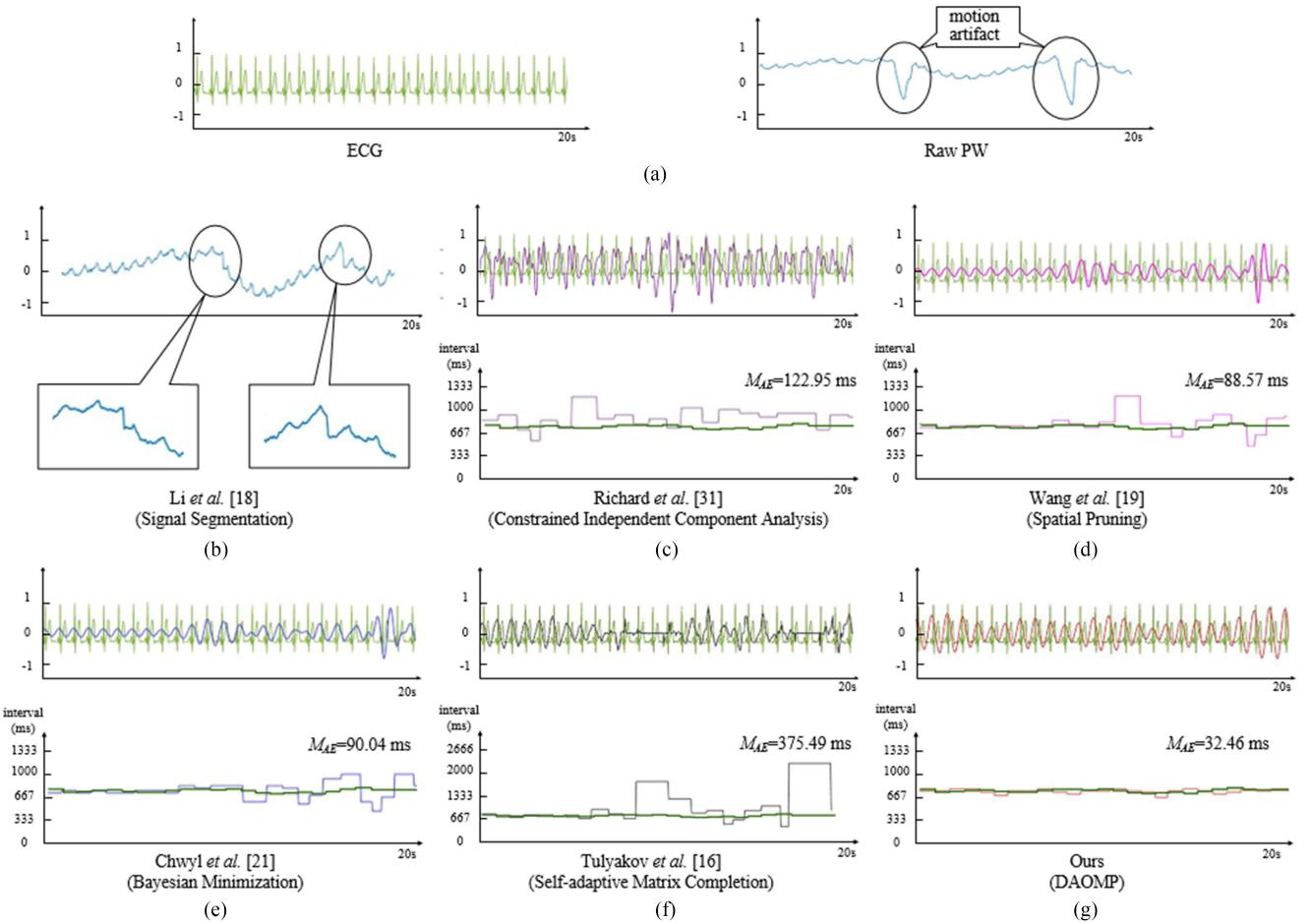


Fig. 10. Results of a comparative experiment in the dynamic scenario among the proposed method and five state-of-the-art methods.

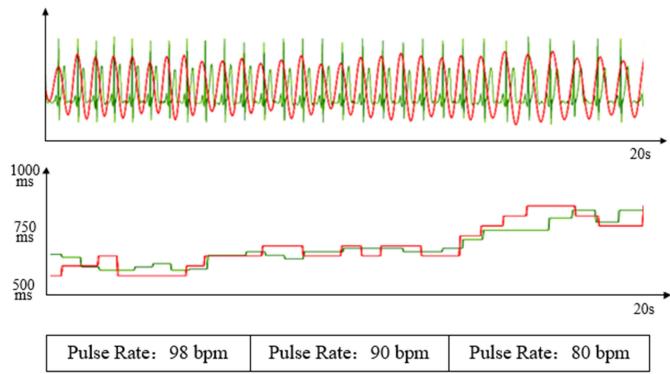


Fig. 11. Results of one test where the subject's IBIs were changing.

PW constructed by our method were highly consistent with the ground truth in IBIs.

Changes in IBIs can reflect a person's physiological status, providing important information for exercise guidance and daily healthcare. Taking the PW in Fig. 11 as an example, the validity of the proposed method in capturing the changes in IBIs is analyzed. It is known that a signal composed of several peridical bases is not necessarily a periodic signal. Therefore, the

proposed method could construct an approximation of a PW containing changing IBIs via combining sufficient sinusoids in the pulse dictionary. Fig. 11 illustrates results of a test in the static testing scenario, where the subject had just stopped exercising and hence his instantaneous pulse rate continued to slide down. In Fig. 11, the peaks of PW constructed by the proposed method were well aligned with the peaks of ECG, and their IBI polyline presented a simultaneously upward trend. Therefore, the proposed method was able to detect changing IBIs effectively, thus making more contribution compared to previous methods [16]–[18] focusing on the average pulse rate detection only.

However, as an exchange of the disturbance robustness, the synthetic PW inevitably lost some detailed components of the raw PW (e.g., ascending limb, descending limb, and dicrotic peak). Methods to protect the detailed components will be further studied in our future work.

V. CONCLUSION

Noncontact detection of the PWs has been rapidly developed thanks to previous researches, but it is still challenged by drastic facial unsteadiness such as head twists and drastic expression changes. A novel method is proposed in this paper

to address drastic facial unsteadiness. Preprocessing including face tracking, facial area separation, and chromaticity feature extraction is firstly performed on a facial video to extract a raw PW matrix. Afterwards, a pulse dictionary is constructed with sinusoidal bases whose frequencies are restricted within the range of human pulse rates. By employing the DAOMP algorithm, a few sinusoidal bases close to the raw PW are chosen from the pulse dictionary to compose a sparse approximation of the underlying pulse components. Experiments conducted on two datasets verified the validation of the proposed method in measuring both averaged pulse rates and IBIs. In the dynamic testing scenario, the proposed method increased *PP* and *AD* by 3.74% and 13.88% respectively on the MAHNOB-HCI dataset, and improved *PP* and *AD* by 4.38% and 14.92% respectively on the IPPG-PW dataset, compared to the best-performing comparative method.

ACKNOWLEDGEMENT

The authors would like to thank M. Soleymani, J. Lichtenauer, T Pun, and M. Pantic for the public video dataset. The authors would also like to thank the associate editor and the anonymous reviewers for their valuable comments and suggestions which substantially helped to improve this paper.

REFERENCES

- [1] W. Wang, S. Stuijk, and H. G. De, "Exploiting spatial redundancy of image sensor for motion robust rPPG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 415–425, Feb. 2015.
- [2] X. Liu *et al.*, "Self-adaptive signal separation for non-contact heart rate estimation from facial video in realistic environments," *Physiol. Meas.*, vol. 39, no. 6, Jun. 2018, Art. no. 06NT01.
- [3] 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.
- [4] Y. Sun *et al.*, "Noncontact imaging photoplethysmography to effectively access pulse rate variability," *J. Biomed. Opt.*, vol. 18, no. 6, Oct. 2013, Art. no. 061205.
- [5] A. R. Guazzi *et al.*, "Non-contact measurement of oxygen saturation with an RGB camera," *Biomed. Opt. Express.*, vol. 6, no. 9, pp. 3320–3338, Oct. 2015.
- [6] L. Kong *et al.*, "Non-contact detection of oxygen saturation based on visible light imaging device using ambient light," *Opt. Express.*, vol. 21, no. 15, pp. 17464–17471, Jul. 2013.
- [7] S. C. Huang *et al.*, "A new image blood pressure sensor based on PPG, RRT, BPTT, and harmonic balancing," *IEEE. Sens. J.*, vol. 14, no. 10, pp. 3685–3692, Oct. 2014.
- [8] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A Neural Network-based method for continuous blood pressure estimation from a PPG signal," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, Minneapolis, MN, USA, 2013, pp. 280–283.
- [9] T. Willum-Hansen *et al.*, "Prognostic value of aortic pulse wave velocity as index of arterial stiffness in the general population," *Circulation*, vol. 113, no. 5, pp. 664–670, Feb. 2006.
- [10] K. Sutton-Tyrrell *et al.*, "Elevated aortic pulse wave velocity, a marker of arterial stiffness, predicts cardiovascular events in well-functioning older adults," *Circulation*, vol. 111, no. 25, pp. 3384–3390, Jun. 2005.
- [11] T. Tamura *et al.*, "Wearable photoplethysmographic sensors—Past and present," *Electronics*, vol. 3, no. 2, pp. 282–302, Apr. 2014.
- [12] Y. Maeda, M. Sekine, and T. Tamura, "The advantages of wearable green reflected photoplethysmography," *J. Med. Syst.*, vol. 35, no. 5, pp. 829–34, Oct. 2011.
- [13] Z. Zhang, Z. Pi, and B. Liu, "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE. Trans. Bio-Med. Eng.*, vol. 62, no. 2, pp. 522–531, Feb. 2015.
- [14] A. T. Wu, V. Blazek, and H. J. Schmitt, "Photoplethysmography imaging: A new noninvasive and noncontact method for mapping of the dermal perfusion changes," *Proc. SPIE*, vol. 4163, pp. 62–70, Nov. 2000.
- [15] H. Y. Wu *et al.*, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–8, Jul. 2012.
- [16] S. Tulyakov *et al.*, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *Proc. IEEE Comput. Vision Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 2396–2404.
- [17] 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, May 2010.
- [18] X. Li *et al.*, "Remote heart rate measurement from face videos under realistic situations," in *Proc. IEEE Comput. Vision Pattern Recognit.*, Columbus, OH, USA, 2014, pp. 4264–4271.
- [19] W. Wang, S. Stuijk, and G. D. Haan, "Exploiting spatial redundancy of image sensor for motion robust RPPG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 415–425, Sep. 2014.
- [20] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomed. Opt. Express.*, vol. 6, no. 5, pp. 1565–1588, Apr. 2015.
- [21] B. Chwyl *et al.*, "SAPPHERE: Stochastically acquired photoplethysmogram for heart rate inference in realistic environments," in *Proc. IEEE Int. Conf. Image Process.*, Phoenix, AZ, USA, 2016, pp. 1230–1234.
- [22] T. Hassner *et al.*, "Effective face frontalization in unconstrained images," in *Proc. IEEE Comput. Vision Pattern Recognit.*, Boston, MA, USA, 2015, 4295–4304.
- [23] G. D. Haan and V. Jeanne, "Robust pulse rate from chrominance-based Rppg," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.
- [24] T. Yamamoto *et al.*, "Derivation and clinical application of special imaging by means of digital cameras and Image J freeware for quantification of erythema and pigmentation," *Skin. Res. Technol.*, vol. 14, no. 1, pp. 26–34, Feb. 2018.
- [25] E. M. Nowara *et al.*, "SparsePPG: Towards driver monitoring using camera-based vital signs estimation in near-infrared," in *Proc. IEEE Comput. Vision Pattern Recognit. Workshops*, Salt Lake City, Utah, USA, 2018, pp. 1271–1281.
- [26] M. Cui and S. Prasad, "Sparse representation-based classification: Orthogonal least squares or orthogonal matching pursuit," *Pattern. Recogn. Lett.*, vol. 84 pp. 120–126, Dec. 2016.
- [27] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, 2002, pp. 40–44.
- [28] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. Comput. Vision Pattern Recognit.*, Kauai, Hawaii, USA, 2001, p. 1.
- [29] C. Tomasi, "Detection and tracking of point features," Tech. Rep. CMU-CS-91-132, Carnegie Mellon University, 1991.
- [30] M. Soleymani *et al.*, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Aug. 2011.
- [31] M. Richard, B. Yannick, and M. Alamin, "Remote photoplethysmography with constrained ICA using periodicity and chrominance constraints," *BioMed. Eng. OnLine*, vol. 17, no. 1, p. 22, Dec. 2018.