Continuous Heart Rate Measurement from Face: A Robust rPPG Approach with Distribution Learning

Xuesong Niu[1,2], Hu Han[*,1], Shiguang Shan[1,2,3], and Xilin Chen[1,2]

[1]Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing 100190, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]CAS Center for Excellence in Brain Science and Intelligence Technology
xuesong.niu@vipl.ict.ac.cn {hanhu, sgshan, xlchen}@ict.ac.cn

Abstract

Non-contact heart rate (HR) measurement via remote photoplethysmography (rPPG) has drawn increasing attention. While a number of methods have been reported, most of them did not take into account the continuous HR measurement problem, which is more challenging due to limited observed video frames and the requirement of speed. In this paper, we present a real-time rPPG method for continuous HR measurement from face videos. We use a multi-patch ROI strategy to remove outlier signals. Chrominance feature is then generated from each ROI to reduce the color channel magnitude differences, which is followed by temporal filtering to suppress the artifacts. In addition, considering the temporal relationship of neighboring HR rhythms, we learn a HR distribution based on historical HR measurements, and apply it to the succeeding HR estimations. Experiment results on the public-domain MAHNOB-HCI database and user tests with commodity webcams show the effectiveness of the proposed approach.

1. Introduction

Heart rate (HR) is an important physiological feature which reflects the physical and emotional activities, *e.g*. exercise, emotion changes, illness, *etc*. Therefore, continuous HR measurement can be very helpful for many applications, such as training aid, health monitoring, nursing care, *etc*. Traditional HR measurement methods usually rely on contact monitors, such as electrocardiograph (ECG) and contact photoplethysmography(PPG). These approaches can be intrusive for the users in many application scenarios.

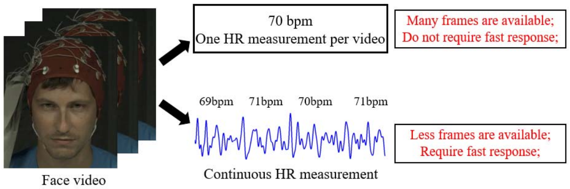

70 bpm
One HR measurement per video
Many frames are available;
Do not require fast response;
69bpm
71bpm
70bpm
71bpm
Less frames are available;
Require fast response;
Face video
Continuous HR measurement

Figure 1. While there are a number of approaches available for HR estimation from face videos, most of them were limited to offline scenarios, e.g., one HR measurement per video. This paper focuses on continuous HR measurement from face using a few observed video frames, which is more challenging and requires not only high measurement accuracy but also fast response. A demo of our approach can be seen at: `http://ddl.escience.cn/f/Ndme`

Recently, non-contact HR measurement methods based on remote photoplethysmography (rPPG) has drawn increasing attention [16, 17, 1, 3, 14, 8, 9]. The main reason is that HR measurement based on rPPG is more flexible, and can be applied without requiring users' cooperation.

HR measurement via rPPG is based on the principle of optical absorption by the skin varies periodically with the blood volume pulse (BVP) [20]. Human skin is usually treated as a three-layer model: subcutis, dermis, and epidermis from inner to surface. The hemoglobin in the blood of dermis and subcutis layers, and melanin in the epidermis layer are the major chromatophores of human skin. The changes of hemoglobin content during a cardiac cycle would cause tiny color variations in the skin. Although the color changes are invisible to human eyes, they can be captured by visible sensors, which makes it possible to measure HR remotely.

While rPPG based HR measurement under controlled environment can be accurate enough for offline analysis,

[*]H. Han is the corresponding author.

978-1-5386-1124-1/17/$31.00 ©2017 IEEE

rPPG based continuous HR measurement remains a challenging problem (see Fig 1). This is because that continuous HR measurement requires not only high accuracy but also fast response, and an estimate should be made given a small number of video frames. In addition, rPPG signals can be influenced by face movement and illumination lighting variations. At the same time, the HR of a subject is often stable and varies within a small range, e.g., 50 and 90 bpm, and the change of HR may appear because of strong emotions and strenuous exercise. Thus, in the situation of health and emotional monitor, estimating only one single HR measurement of the subject per video is often limited, and continuous HR measurement is required.

In order to achieve robust continuous HR measurement via rPPG, we argue that a system should consider the following factors: (1) reliable detection and tracking for region of interest (ROI) on the face, (2) efficient cardiac cycle signal extraction and enhancement, and (3) ability to handle temporal subtle changes. However, most of the previous methods focus on estimating a single HR value given a long video sequence; their effectiveness is not known given a small number of observed frames.

In this paper, we present a novel approach for continuous HR measurement aiming to address the above issues. Firstly, we use multi-patch ROIs calculated based on facial landmarks and skin segmentation, obtaining better local consistency and robustness against facial movement. Estimations of all ROIs are fused to get a final HR measurement. Secondly, we transform the cardiac cycle signal from RGB space into chrominance space to reduce the magnitude differences of individual color channels, and apply temporal filters to reduce the influence of white noise and noise from frequency domain we are not interested in. Finally, considering the contextual relationship of the temporal HR signals, we learn a HR distribution based on historical HR measurements, and apply it to the succeeding HR estimations. Experiments on the MAHNOB-HCI database [18], and user tests with a commodity webcam demonstrate the effectiveness of our proposed approach.

2. Related Work

A wide variety of salient information can be obtained from human face, including a person's identity, demographic attributes [4, 6, 22], and even physiological features such as heart rate [16, 14, 8].

Blind signal separation (BSS) was introduced in [16] for remote HR estimation, in which ICA was used to seek the source signals that are maximally independent in an information-theoretic sense. The separated color signals were found to have high SNR, and were used for frequency analysis. In a latter work of [17], temporal filters, such as the moving average filter, and bandpass filter, were applied to reduce the noise in the temporal signal sequence. Other BSS methods such as PCA were also used to seek the source signals that are minimally correlated in a probabilistic sense [13].

Except for the methods using BSS, Haan and Jeanne proposed a HR measurement method using chrominance difference [3]. They used skin segmentation to separate skin and non-skin pixels, and then computed the chrominance feature using the combination of two orthogonal projections of RGB space to reduce the influence of face motion. In the work of [23], a pixel-wise chrominance feature calculation method is used for HR estimation.

Many of the previous methods reported their performance on private databases, leading to difficulties in performance comparison by the succeeding approaches. Li *et al.* [14] proposed a framework that achieved the state-of-the-art HR estimation accuracy on the public-domain MAHNOB-HCI database [18]. They used facial landmarks to locate the area of face. The influence of illumination was removed by comparing with the background, and the influence of non-grid face motion was suppressed by statistical analysis and a few temporal filters. However, they did not consider the scenarios of continuous HR measurement.

Recent studies on HR measurement focus on how to select ROIs from the face. In [10], Kumar *et al.* proposed a method to combine the green channel signal of different ROIs using the frequency characteristics as weights. Lam *et al.* selected a number of random patches from face, and used a majority vote scheme to find the optimal HR estimation [12]. Tulyakov *et al.* divided the face into multiple ROI regions, and used a matrix completion approach to purify rPPG signals [19].

Besides the color-based HR measurement methods, a motion-based method was proposed in[1]. Inspired by the Eulerian magnification method [24], they tracked subtle head motions caused by cardiovascular circulation, and used PCA to get the pulse signal from the trajectories of multiple tracked feature points. Since the method is based on subtle motion, no subjects' voluntary movements are allowed, leading to very limited use in real applications.

In summary, the published methods for HR measurement have the following limitations. First, most of these approaches use the average of color values in the whole ROI as the original rPPG signal, which ignores the local information within each ROI. Although average operation is helpful in reducing Gaussian noises under the assumption that all pixels in a ROI have similar baseline value and variations, the average operation becomes less effective when the assumption does not hold. At the same time, most of the published methods focused on measuring the average HR for a input video, and fail to reflect the continuous changes of HRs.

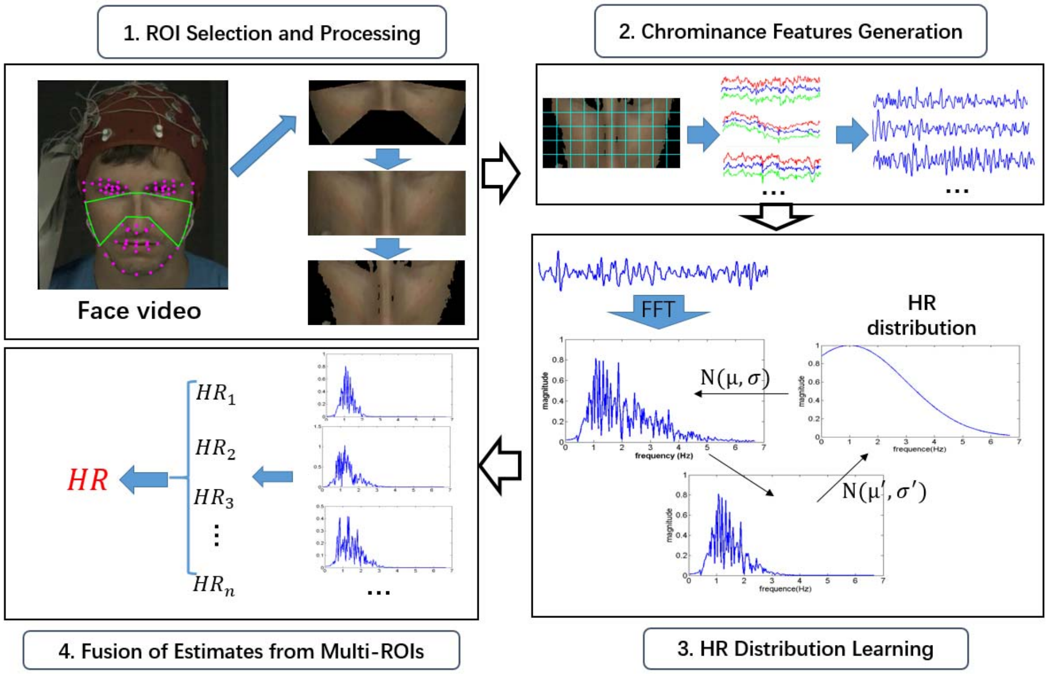


Figure 2. An overview of proposed approach for continuous HR measurement from face videos.

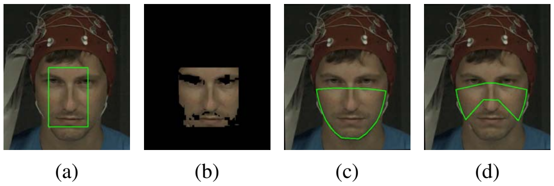


Figure 3. Different approaches for ROI selection: (a) a rectangle ROI at the center of the face [16], (b) the skin area ROI of the whole face [3], (c) the bottom face ROI [14], and (d) proposed ROI at the cheek area.

3. Proposed approach

Figure 2 gives an overview of the proposed approach for continuous HR measurement. Generally speaking, we use a frequency domain analysis with robust time domain features generated from local facial ROIs to capture the periodic color changes due to heart pulses. Additionally, considering the temporal relationship between neighboring HRs, we learn a HR distribution to determine the maximum expectation of HR estimation in the succeeding signals. We provide the details below.

3.1. ROI Selection and Processing

As we summarized in the introduction, the quality (*e.g*. SNR) of the raw color signal plays an important role for rPPG based HR estimation. Although color signal quality can be improved by using expensive RGB sensors, such a hardware based method will limit the application scenarios. A more valuable way is to improve the SNR of color signals captured by commodity webcams, which can be achieved by carefully choosing the ROIs from a face. Traditional ROI selection methods include rectangle ROI at the center of the face (Figure 3a), ROI of face skin segmentation (Figure 3b), and the bottom area of face (Figure 3c). All of these ROI choices contain some irrelevant areas, and may introduce non-grid motions. At the same time, all of ROI choices average over a large face region, which may contain different patterns of local variations and lose the local consistency.

As reported in [11], the most informative facial part containing color changes due to heart rhythms is the cheek area. The cheek area contains much less non-rigid motions due to smiling and talking than the other areas. Therefore, we choose to use the cheek area as the ROI. Specifically, we use an open source face detector [25] to localize 81 facial landmarks (see Figure 2), and calculate a polygon (based on the landmarks) ROI on the cheeks. Since the facial landmark detection is able to run at more than 30 fps, we can perform landmark detection on every frame in order to get stable ROI across the frames.

After obtaining a ROI, we use a piece-wise linear wrapping method to wrap the cheek area into a $M \times N$ rectangle for the convenience of computing. For the reason that each facial landmark has a particular semantic meaning, we can assume that each pixel in the wrapped ROI rectangle is aligned. Furthermore, in order purify the raw color signals, irrelevant pixels are removed from the rectangle ROI by using skin segmentation.

3.2. Local Chrominance Features Generation and Temporal Filtering

After the rectangle ROI is computed, we divide the whole rectangle ROI ($M \times N$) into K smaller ones considering that the face is not a perfect lambertian surface, and smaller ROIs should have better consistency than a larger one. As stated in [21], average pooling is helpful to reduce the sensor noises and improve the SNR of rPPG signals. Let $R(x, y, t)$ denote the red channel value at location (x, y) of the t^{th} frame, the average pooling of the i^{th} ROI for the red channel at time t is

$$\overline{R}_i(t) = \frac{\sum_{x,y \in ROI_i} R(x, y, t)}{|ROI_i|} \tag{1}$$

where $|ROI_i|$ denotes the ROI area (the number of pixels). So, for each ROI we obtain a temporal sequence for each of the R, G, and B channel, *e.g.*, $\mathbf{R}_i = \{\overline{R}_i(1), \overline{R}_i(2), \cdots, \overline{R}_i(n)\}$. Therefore, we totally have $3 \times K$ signals.

Given these raw rPPG signals extracted from multiple ROIs, we then transform the signals from RGB to chrominance space. The chrominance features are found to be more robust to motion and illumination variations [3]. Let X denote the pulse signal under the influence of face motions as

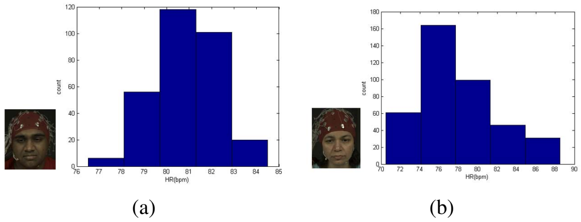
$$\mathbf{X} = I_c(\rho_{Cdc} + \rho_{Cac})M, \quad \mathbf{X} \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\} \tag{2}$$

where I_c is the intensity of light source, ρ_{Cdc} indicates the direct-current part of the reflection coefficients of the skin, ρ_{Cac} indicates the alternating part, and M is the influence factor of the motion. However, directly using the ratio of different color signals could not handle the problem of nonwhite illumination [7]. A skin-tone standardization approach was proposed to eliminate these influences. The final features could be expressed as

$$\mathbf{S} = X_f - \alpha Y_f \tag{3}$$

where X_f and Y_f are signals after bandpass filtering for X and Y, with $X = 3R_n - 2G_n$ and $Y = 1.5R_n + G_n - 1.5B_n$ (R_n, G_n, B_n are the normalized R, G, B signals) and $\alpha = \frac{\delta(X_f)}{\delta(Y_f)}$ ($\delta(X_f)$ and $\delta(Y_f)$ are the standard deviations of X_f and Y_f).

After converting the rPPG signals into chrominance space, we use several additional filters to remove various artifacts. We first use a Gaussian smoothing filter with a window size of 5 frames to reduce noises introduced by ROI average pooling. Then, a 4th order butterworth bandpass filter with the transmission band of $[0.7, 4]$ Hz (corresponding to $[42, 240]$ bpm) is used to eliminate the frequencies that are less likely to be HR distributions.



(a)

(b)

Figure 4. HR distributions calculated from the ground-truth HRs of two subjects in the MAHNOB-HCI database.

3.3. HR Distribution

After filtering the chrominance signals of each ROI, we use Fast Fourier Transformation (FFT) to transform the unit sequence signal (i.e., l frames) from time domain to frequency domain, and get the power spectral density distribution $\hat{f}(\omega)$. Then, HR can be measured by searching the peak power spectral (*e.g.* $\overline{\omega} = \text{argmax}\hat{f}(\omega)$) in frequency domain. The HR estimate per minute can be calculated as $\overline{n} = 60\overline{\omega}$. After each estimation, we move the sequence signal by one second for the next HR estimation. Therefore, after the first l frames, the proposed approach can continuously output a HR estimation per second.

The above HR estimation can be accurate enough for still scenarios, but in a continuous HR measurement scenario, where rigid and non-rigid facial movement may appear likely to appear, the above measurement becomes not stable. Although the chrominance domain filtering could reduce such influence to some extent, the temporal relationship of neighboring rPPG signals are not explicitly utilized. As a result, the HR estimates may flicker during continuous HR measurement, leading to false alarms in some applications like a health monitor. So, to handle these issues, we propose to use the HR distribution to model the temporal relationship, and use it to modulate the succeeding HR measurement.

Specifically, as shown in Figure 4, it is reasonable to assume that the pulse frequency distribution of individual subjects follows a Gaussian distribution

$$HR_t \backsim N(\mu_{HR}, \sigma_{HR}) \tag{4}$$

where μ and σ are the mean and standard deviation of HR distribution, respectively (see Figure 4). For continuous HR measurement of a subject, we first learn μ_{HR} and σ_{HR}. We need a period of time T to estimate HR without the help of HR distribution. Then, the parameters (μ_{HR} and σ_{HR}) can be easily estimated by using the mean and standard deviation of prior estimations. The HR distribution models the subject's HR within a recent period, and it can be used as a constraint to remove outlier HR estimations. Specifically, we use prior estimations $HR_1, HR_2, \cdots, HR_T$ to compute HR distribution, and get the parameters μ_{HR}, σ_{HR}.

Given a new sequence signal, we firstly compute its frequency $\hat{F}(\omega)$, and then we modulate the frequency magnitudes using weights $P(\omega)$ as follow

$$\hat{\mathbb{F}}(\omega) = P(\omega) \circ \hat{F}(\omega) \tag{5}$$

where the entries of $P(\omega)$ are computed based on the HR distribution

$$p(\omega_i) = \frac{1}{(\sigma_{HR} + \sigma_0)\sqrt{2\pi}} e^{-\frac{(\omega_i - \mu_{HR})^2}{2(\sigma_{HR}+\sigma_0)^2}} \tag{6}$$

and σ_0 is the parameter to balance the influence of history and current estimation. During the continuous HR measurement, the parameter σ_{HR} is updated using all the historical estimates in the current video sequence. After generating the modified frequency domain $\hat{\mathbb{F}}(\omega)$, it is reasonable to find the frequency $\tilde{\omega}$ that has the maximum $\hat{\mathbb{F}}(\omega)$, and calculate the corresponding HR per minute as $\tilde{n} = 60\tilde{\omega}$. The introduced temporal context modeling method looks simple, but it is found to be useful to reduce the influence of head motions, and stabilize the HR estimations.

3.4. Fusion of Estimates from Multi-ROIs

As we described in Section 3.1, small areas of face would have better consistency than a larger one, and we have divided the ROI into k small regions. By using the method mentioned in Sections 3.2 and Sections 3.3, we would get k HR estimations for a unit signal sequence. A simple average of the k estimations is not enough to get an accurate HR estimation, because we notice that some of the estimations are extreme high or low. In order to reduce the influence of these extreme errors, we use a median HR estimation of the K estimates. Specifically, we firstly sort all the k estimations, and we get $\{hr_1\ hr_2, \cdots\ hr_k\}$. Then we choose the median $2l+1$ estimations $\{hr_{[k/2]-l}\ hr_{[k/2]-l+1}, \cdots\ hr_{[k/2]+l}\}$ as the stable estimations. Finally, we compute the HR estimation as

$$hr = \frac{\Sigma_{i=[\frac{k}{2}]-l}^{[\frac{k}{2}]+l} hr_i}{2l+1} \tag{7}$$

4. Experimental Results

In this section, we provide evaluations of the proposed approach from several perspectives: the effectiveness of key components in our approach, continuous HR measurement on public database and by user tests, single HR measurement, and computational cost on commodity desktop.

4.1. Experimental Settings

Different kinds of statistics have been used in the literature for evaluating the accuracies of different HR measurement methods, such as the HR error between estimated HR and ground-truth HR (HR$_e$), the mean and standard deviation of the HR error (HR$_{me}$, and HR$_{sd}$), the root mean squared HR error (HR$_{rmse}$), and the mean of error rate percentage (HR$_{mer}$) [14]. For both the continuous HR measurement, and one measurement per video scenarios, we use HR$_{me}$, HR$_{sd}$, HR$_{rmse}$ and HR$_{mer}$ to report our results.

The public domain MAHNOB-HCI database [18] is used in both the continuous HR measurement and one HR measurement per video experiments. The MAHNOB-HCI database is a multimodal database with 20 high resolution videos per subject, and 27 subjects (12 males and 15 females) in total. Each subject participated in the experiment of emotion elicitation and implicit tagging, during which the HR may float because of the change of subject's emotions. For continuous HR measurement, besides MAHNOB-HCI, we also perform user tests using a commodity Logitech C270 webcam (640 × 480 at 30 fps) on a Windows 10 desktop with Intel Core I7 3.6GHz CPU. The ground-truth HRs on MAHNO-HCI for both continuous HR measurement and one HR measurement per video are calculated based on the EEG signal provided in the database. For the user tests, we use a FDA approved Contec CMS50D finger pulse oximeter as the reference.[1] For continuous HR measurement with the MAHNOB-HCI database and user tests, the proposed approach generates one HR estimation per second.

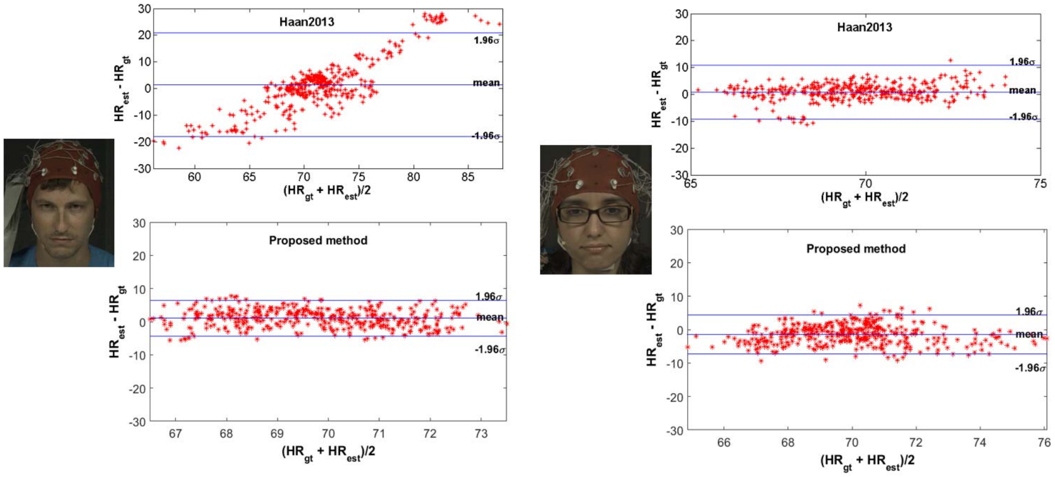
For the proposed approach, we use a ROI rectangle of 100×200, and divided it into 32 regions (4×8 grid). We use $l = 5$ for the median estimation in Section 3.4.

4.2. Continuous HR Measurement

We first provide evaluations under this scenario of continuous HR monitoring using the MAHNOB-HCI database and user tests.

- **Test on MAHNOB-HCI.** In this experiment, we use videos with the length of 90 seconds (frame 306 to 5490) from the MAHNOB-HCI database for continuous HR measurement, and we have 416 video chips in total for our experiments. We use the sliding window of 30s, and compute 30 prior HR estimations for HR distribution learning. Since there is not known result reported under the continuous HR measurement scenario, we implemented a few baseline methods based on the published methods. Specifically, we implemented the method in [3] and [16] with sliding windows, and report their accuracies with and without using our HR distribution learning. From Table 1, we can see that the HR distribution we proposed provides robustness against fluctuations in the raw color signals, e.g., due to motion, and significant improvement could be seen after using HR distribution in all the methods.

[1] The main reason why we used a portable finger pulse oximeter is mainly because of the users' concerns in privacy.



(a)

(b)

Figure 5. The Bland-Altman plots of two subjects from MAHNOB-HCI database. For each subject there are two figures to show, the top one is the result using Haan and Jeanne's method (denoted as Haan2013) [3], while the bottom is using our method.

Table 1. Comparisons on the MAHNOB-HCI database under the scenario of continuous HR measurement. We present results of the same methods with and without our HR distribution learning. (DL: HR distribution learning described in Section 3.3.)

Method	HR_{me} (bpm)	HR_{sd} (bpm)	HR_{rmse} (bpm)	HR_{mer}
Poh2010[16] w/o DL	-0.38	16.08	11.11	15.4%
Poh2010[16] with DL	-1.52	13.94	9.93	13.7%
Haan2013[3] w/o DL	0.68	14.03	9.06	12.7%
Haan2013[3] with DL	**-0.17**	12.16	8.70	11.4%
Proposed method w/o DL	-1.87	10.98	9.21	12.4%
Proposed method with DL	-0.98	**10.42**	**7.82**	**11.1%**

Table 2. Comparisons on the MAHNOB-HCI database under the scenario of continuous HR measurement using different ROI strategies. We use chrominance signals generated from different ROIs for comparison. 'Proposed ROI (avg.)' denotes the method using the average of the whole ROI described in 3.1, and 'Proposed ROI (local)' denotes the proposed multi-patch ROIs approach.

ROI selection	HR_{me} (bpm)	HR_{sd} (bpm)	HR_{rmse} (bpm)	HR_{mer}
Rectangle [16]	-1.62	13.97	9.76	13.4%
Skin segmentation [3]	**-0.17**	12.16	8.70	11.4%
Bottom face [14]	-0.96	13.36	9.18	12.7 %
Proposed ROI (avg.)	1.60	11.29	8.98	13.2%
Proposed ROI (local)	-0.98	**10.42**	**7.82**	**11.1%**

Furthermore, to demonstrate the effect of our ROI selection and processing methods, we report the HR measurement results using different ROI determination strategies, including the rectangle ROI at the center of the face [16], skin area ROI [3], bottom face area[14], and the proposed ROIs. From the results in Table 2, we can see that using local regions to estimate HR is helpful to eliminate noise introduced by grid and non-grid facial movement as well as the inconformity of different areas, and outperforms all the methods using global areas. The proposed approach benefits from the accurate landmark detection [25], which is helpful to obtain stable ROI localizations, and therefore stable signals of local-ROIs.

We also evaluate the consistency between the ground truth HR and the estimated HR, by showing the Bland-Altman plot [2] for some subjects in Figure 5. The Bland-Altman plot for a state-of-art method [3] is also given for comparation. It can be seen that our method has smaller standard deviation and the $HR_{est} - HR_{gt}$ are closer to 0. We further check the error distributions of the proposed method and [3]. As shown in Figure 6, 67.2% of the cases are estimated with an error less than 5bpm using the proposed method, while the percentage for [3] is only 58.7%.

Finally, Figure 7 shows the performance of proposed approach for different σ_0, which is used to balance the effect of HR distribution and current estimation. As we can see from Figure 7, too small or too large value of

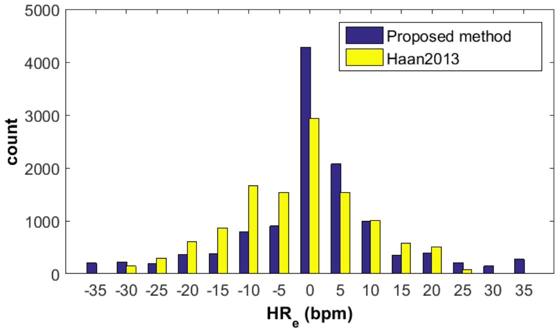


Figure 6. Comparison of the HR estimation error distributions of the proposed approach and Haan and Jeanne's method [3].

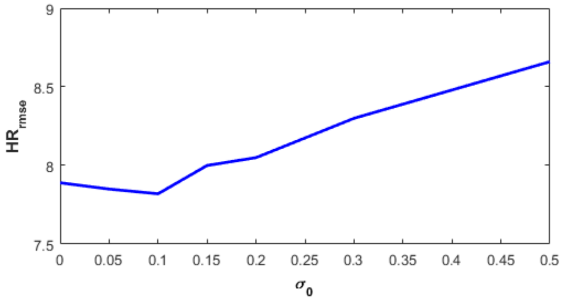


Figure 7. The influence of parameter σ_0 in our HR distribution learning.

σ_0 may decrease the performance; so we use $\sigma_0 = 0.1$ in all experiments.

- **User tests.** As described in the experimental settings, we also have a number of user tests using a Logitech 310 camera and desktop. We have collected 27 videos of 10 subjects, containing 9 males and 1 female. All of the videos are recorded indoor, but with natural lighting variations. The ground-truth HR for reference is given by a FDA approved Contec CMS50D finger pulse oximeter, where the display screen of the pulse oximeter is recorded together with the HR estimation values by the proposed approach (see Figure 8). We then calculate the HR$_{me}$, HR$_{sd}$, HR$_{rmse}$, and HR$_{mer}$. The results by our approach are HR$_{me}$ $= 0.1bpm$, HR$_{sd}$ $= 0.3bpm$, HR$_{rmse}$ $= 0.1bpm$, and HR$_{mer}$ $= 1\%$, which is very promising. The user test results again show that our method is effective for continuous HR measurement.

4.3. One HR Measurement per Video

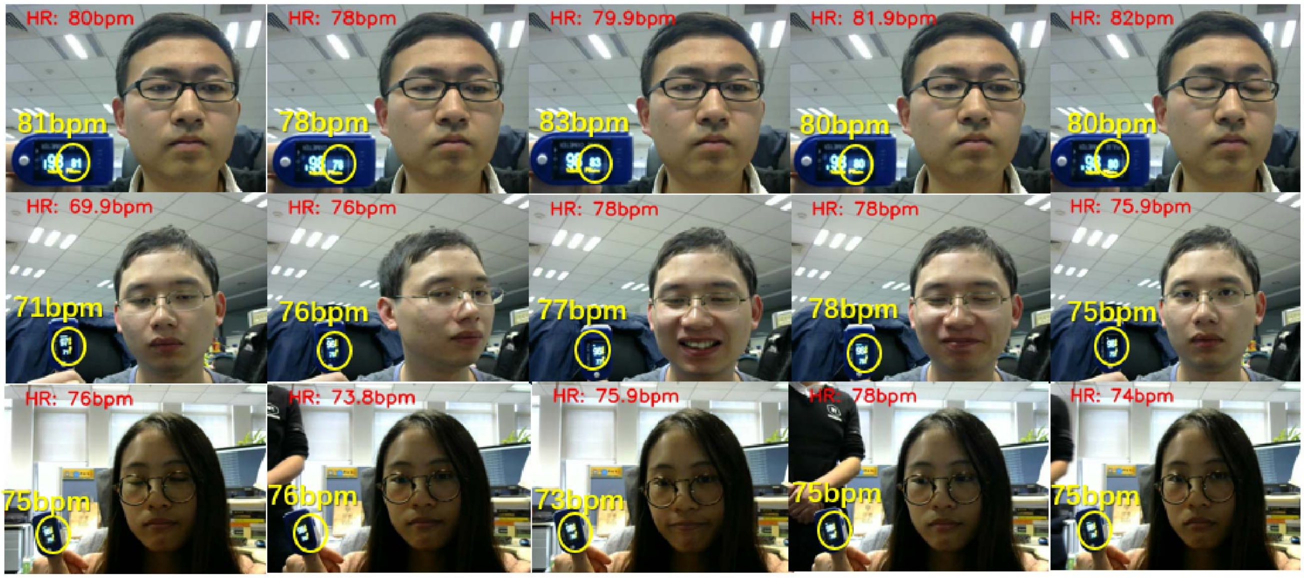
Following the experiments in [14], we also report the results on MAHNOB-HCI database under the scenario of one HR measurement per video. We no longer use the HR distribution in the one HR measurement per video tests. We also compare the proposed approach with a number of the state of the art methods, such as [1] [3] [14] [16] [17] [19], in which [3] and [16] are implemented by ourselves, and the results of the other methods are directly from the published papers. From Table 3, we can see that the proposed approach outperforms the methods [16] and [17], and is comparable to [14]. Another most recent work in [12] also reported higher performance than ours, but as the authors stated in their paper that their method is very slow (see running time comparisons below). These experiments show that although the proposed approach is designed for continuous HR measurement scenarios, its performance under single HR measurement per video is also comparable to the state of the art methods.

Table 3. Comparisons on the MAHNOB-HCI database under the scenario of one HR measurement per video.

Method	HR$_{me}$ (bpm)	HR$_{sd}$ (bpm)	HR$_{rmse}$ (bpm)	HR$_{mer}$
Poh2010 [16]	-8.95	24.3	25.9	25.0 %
Poh2011 [17]	2.04	13.5	13.6	13.2 %
Balakrishnan2013 [1]	-14.4	15.2	21.0	20.7%
Haan2013 [3]	-2.89	13.67	10.7	12.9%
Li2014 [14]	-3.30	6.88	7.62	6.87%
Tulyakov2016 [19]	3.19	5.81	6.23	5.93%
Proposed method	-0.38	10.81	8.72	11.5 %

4.4. Running Time

For rPPG based HR measurement problem, running time is an important factor that affects its application scope. We profiled the running time of each step of the proposed approach on a Windows 10 desktop with Intel Core I7 3.6GHz CPU and 32G RAM. For the key components of ROI wrapping and processing, chrominance feature and filtering, HR distribution learning modulation and FFT, each takes about 5ms, and overall the proposed approach takes less than 20ms for generating one HR estimate, and uses less than 30MB memory in total. Thus, the proposed approach can run as fast as 50 fps, and satisfy the requirement of various application scenarios. By contrast, most of the state of the art methods did report their running time on commodity desktop machines. The only running time we can find is reported in [12], which is about 4 fps. Feedback from the authors of [14] indicates that their method runs at about 10 fps on a desktop. Thus, the proposed approach is much faster than these two state of the art methods.



(a)

Figure 8. Examples of the recorded video frames in our user tests, where both the display screen of the pulse oximeter and our output are recorded for later evaluation in terms of HR$_{mae}$. The yellow number is the HR measured by the pulse oximeter, and the red number is our estimation.

5. CONCLUSIONS AND FUTURE WORKS

Non-contact continuous heart rate measurement via remote photoplethysmography is useful but a challenging problem due to the limited number of video frames in observation and the requirement of quick response and high accuracy. In this paper, we address these issues from the perspectives of ROI selection, chrominance feature generation, filtering, and heart rate distribution learning. We proposed a multi-patch ROI method to assure the local consistency of color signals. Chrominance feature generation from color space is applied to reduce the color channel magnitude differences, followed by temporal filtering to suppress the artifacts. In addition, the temporal relationship of neighboring heart rate rhythms is modeled via heart rate distribution, and applied to the succeeding heart rate estimations. Experimental results on the public domain MAHNOB-HCI database and user tests show the effectiveness of the proposed approach. Finally, our system is able to run in real-time (about 50 fps) on a commodity desktop machine.

In our future work, we would like to improve the robustness of rPPG based heart rate measurement under uncooperative scenarios by considering multiple facial component ROIs [5, 15]. The influence of different cameras will also be studied including both webcams and smartphone cameras.

Acknowledgement

This research was partially supported by 973 Program (grant 2015CB351802) and Natural Science Foundation of China (grant 61390511, 61672496, 61650202, and 61472398). H. Han gratefully acknowledges the support of NVIDIA Corporation with the donation of the Titan X GPU used for his research. The authors would like to thank Xiaoyan Li for her proofreading with the paper.

References

[1] G. Balakrishnan, F. Durand, and J. Guttag. Detecting pulse from head motions in video. In *Proc. IEEE CVPR*, pages 3430–3437, 2013.

[2] J. M. Bland and D. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476):307–310, 1986.

[3] G. de Haan and V. Jeanne. Robust pulse rate from chrominance-based rPPG. *IEEE Trans. Biomed. Eng.*, 60(10):2878–2886, 2013.

[4] H. Han, A. K. Jain, S. Shan, and X. Chen. Heterogeneous face attribute estimation: A deep multi-task learning approach. *arXiv 1706.00906*, Jun. 2017.

[5] H. Han, B. F. Klare, K. Bonnen, and A. K. Jain. Matching composite sketches to face photos: A component-based approach. *IEEE Trans. Inf. Forensics Security*, 8(1):191–204, Jan. 2013.

[6] H. Han, C. Otto, X. Liu, and A. K. Jain. Demographic estimation from face images: Human vs. machine performance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(6):1148–1161, Jun. 2015.

[7] H. Han, S. Shan, X. Chen, and W. Gao. A comparative study on illumination preprocessing in face recognition. *Pattern Recognition*, 46(6):1691 – 1699, Jun. 2013.

[8] M. A. Haque, R. Irani, K. Nasrollahi, and T. B. Moeslund. Heartbeat rate measurement from facial video. *IEEE Intelligent Systems*, 31(3):40–48, 2016.

[9] M. A. Haque, K. Nasrollahi, and T. B. Moeslund. Heartbeat signal from facial video for biometric recognition. In *Proc. SCIA*, pages 165–174, 2015.

[10] M. Kumar, A. Veeraraghavan, and A. Sabharwal. DistancePPG: Robust non-contact vital signs monitoring using a camera. *Biomed. Opt. Express*, 6(5):1565, May 2015.

[11] S. Kwon, J. Kim, D. Lee, and K. Park. Roi analysis for remote photoplethysmography on facial video. In *Proc. EMBS*, pages 851–862, 2015.

[12] A. Lam and Y. Kuno. Robust heart rate measurement from video using select random patches. In *Proc. IEEE ICCV*, pages 3640–3648, 2015.

[13] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak. Measuring pulse rate with a webcam - a non-contact method for evaluating cardiac activity. In *Proc. ComSIS*, pages 405–410, 2011.

[14] X. Li, J. Chen, G. Zhao, and M. Pietikainen. Remote heart rate measurement from face videos under realistic situations. In *Proc. IEEE CVPR*, pages 4264–4271, 2014.

[15] C. Otto, H. Han, and A. K. Jain. How does aging affect facial components? In *Proc. ECCV Workshop*, pages 189–198, 2012.

[16] M.-Z. Poh, D. J. McDuff, and R. W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express*, 18(10):10762–10774, 2010.

[17] M.-Z. Poh, D. J. McDuff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans. Biomed. Eng.*, 58(1):7–11, 2011.

[18] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.*, 3(1):42–55, 2012.

[19] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe. Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. In *Proc. IEEE CVPR*, 2016.

[20] W. Verkruysse, L. O. Svaasand, and J. S. Nelson. Remote plethysmographic imaging using ambient light. *Opt. Express*, 16(26):21434–21445, 2008.

[21] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE CVPR*, pages I–511–518, 2001.

[22] F. Wang, H. Han, S. Shan, and X. Chen. Deep multi-task learning for joint prediction of heterogeneous face attributes. In *Proc. IEEE FG*, pages 1–7, 2017.

[23] W. Wang, S. Stuijk, and G. De Haan. Exploiting spatial redundancy of image sensor for motion robust rppg. *IEEE Trans. Biomed. Eng.*, 62(2):415–425, 2015.

[24] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman. Eulerian video magnification for revealing subtle changes in the world. 2012.

[25] J. Zhang, S. Shan, M. Kan, and X. Chen. Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In *Proc. IEEE ECCV*, pages 1–16, 2014.