Remote Photoplethysmography Based on Implicit Living Skin Tissue Segmentation

Serge Bobbia, Yannick Benezeth, Julien Dubois
Univ. Bourgogne Franche-Comté
LE2I UMR6306, CNRS, ENSAM
F-21000 Dijon, France
Email: serge.bobbia@u-bourgogne.fr

Abstract—Region of interest selection is an essential part for remote photoplethysmography (rPPG) algorithms. Most of the time, face detection provided by a supervised learning of physical appearance features coupled with skin detection is used for region of interest selection. However, both methods have several limitations and we propose to implicitly select living skin tissue via their particular pulsatility feature. The input video stream is decomposed into several temporal superpixels from which pulse signals are extracted. Pulsatility measure for each temporal superpixel is then used to merge pulse traces and estimate the photoplethysmogram signal. This allows to select skin tissue and furthermore to favor areas where the pulse trace is more predominant. Experimental results showed that our method perform better than state of the art algorithms without any critical face or skin detection.

I. INTRODUCTION

Photoplethysmography (PPG) is a non-invasive technique for detecting microvascular blood volume changes in tissues. Nowadays, PPG is applied ubiquitously in many settings where a contact PPG sensor (also known as pulse oximeter) is typically attached to a finger or patched to the skin. Basically, contact PPG sensors are used to determine the heart rate and oxygen saturation in blood. The principle of this technology is actually very simple as it only requires a light source and a photodetector. The light source illuminates the tissue and the photodetector measures the small variations in transmitted or reflected light associated with changes in perfusion in the tissue [1].

However, conventional contact PPG sensor is not suitable in situations of skin damage or when unconstrained movement is required. Moreover, it has been showed that pressure of the conventional clip sensors tend to affect the waveform of PPG signal because of the contact force between the finger and the sensor [2].

With the emergence of camera-based health care monitoring, remote photoplethysmography (rPPG) has recently been developed as it allows remote physiological measurement without expensive and specialist hardware. Actually, it has been shown recently [3] that it is possible to recover the cardiovascular pulse wave measuring variations of back-scattered light remotely, using only ambient light and low-cost vision systems. Since this seminal work, there has been rapid growth in the literature pertaining to remote PPG techniques.
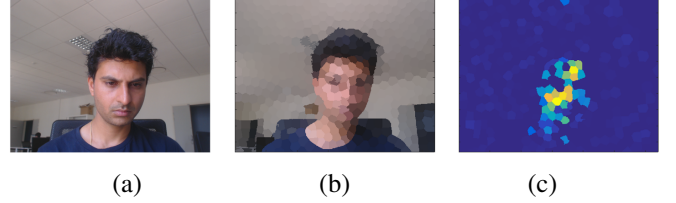

Fig. 1. Pulsatility measures estimated from various temporal superpixels. (a) input frame, (b) temporal superpixel segmentation and (c) pulsatility measures (blue means low pulsatility measures and yellow/orange is high).

Most methods share a common pipeline-based framework (*e.g.* [4]–[7]): regions of interest (ROI) are first detected and tracked over frames, RGB channels are then combined to estimate the pulse signal, which is filtered and analyzed to extract physiological parameters such as heart rate or respiration rate. An interesting and comprehensive state of the art paper on PPG and rPPG has been recently proposed by Sun and Thakor [1]. The selection of ROI is a critical first step to obtain reliable pulse signals. The ROI should contain as many skin pixels as possible. Several approaches have been proposed for ROI selection in the video stream. In earlier studies, manual selection of the ROI have been used [3], [8]. ROI can also be defined based on the result of classical face detection [9] and tracking [10] algorithms and possibly refined with skin pixel classification [11]. It is worth mentioning that this preliminary step can be computationally very expensive (*e.g.* [11]).

Pixels in the ROI are then usually spatially averaged and the process is repeated in each video frame. The result of this process is a time series, that is later used to obtain rPPG signal. It has been shown in several studies that the quality of the ROI has a direct impact on the quality of the rPPG signal (*e.g.* in [12]). First, because a smaller number of skin pixels leads to larger quantized RGB errors, it can be observed that the quality of rPPG signal deteriorates while down-sampling the ROI. This may be understood as the reduction of the sensor noise amplitude by a factor equal to the square root of the number of pixels used in the averaging process [13]. Second, the quality is also affected by the percentage of non-skin pixels in the ROI [7]. All rPPG algorithms suffer from performance degradation when the ROI is not properly selected. These two remarks are fairly intuitive but it is actually quite difficult in
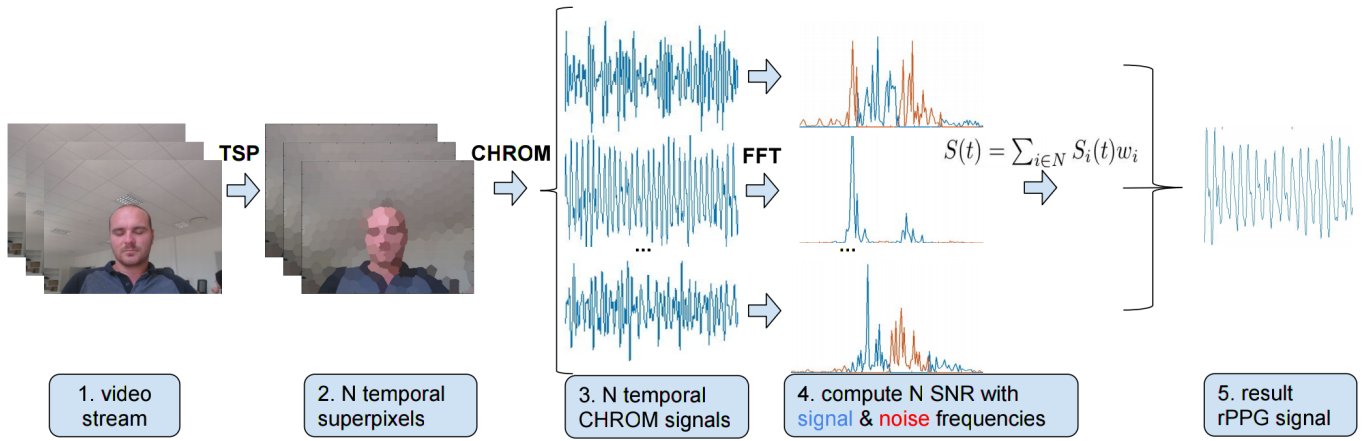
Fig. 2. Overview of the proposed method. (1) Input video stream is (2) decomposed into temporally consistent superpixels. (3) *Tentative* rPPG signal is extracted from each TSP. (4) A pulsatility measure is estimated for each ROI. Blue signal is the convolution of the periodogram by h_{signal} and red signal is the convolution by h_{noise}. (5) A weighted average of all the *tentative* rPPG signals is finally computed.

practice to get a well-defined ROI, that is stable over time, without performing complex calculations. Moreover, as shown by [14], the rPPG signal is not distributed homogeneously on skin. Some skin regions contain more PPG signal than others. For example, we experimentally observed that signal-to-noise ratio (SNR) of photoplethysmogram signals extracted from forehead or cheekbones are clearly higher than those obtained from the chin. Figure 1 presents the SNR of rPPG signals calculated from several skin regions.

From these observations, we propose a new method that implicitly selects ROI that represents living skin tissue and that favor regions of interest where the pulse trace is more predominant. We use the term *implicit* to differentiate our method with those that require critical pre-processing steps for ROI selection and tracking. As opposed to conventional approaches based on face detection, tracking and skin segmentation, ROI selection is based on the fact that only the skin tissue of an alive subject exhibits pulsatility. The input video stream is decomposed into several temporal superpixels from which pulse signals are extracted. Pulsatility measure for each temporal superpixel is then used to merge pulse traces and estimate the photoplethysmogram signal. This approach can be associated with any rPPG algorithm. In this paper, we experimentally validated the proposed automatic living skin tissue segmentation for ROI selection using the chrominance-based method (also known as CHROM) [6] because this method is definitely one of the most reliable rPPG methods. To the best of our knowledge, work by [15], called Voxel-Pulse-Spectral (VPS) is the closest contribution to ours. They propose an automatic and unsupervised subject detection via rPPG. However computational load of VPS is significantly heavier. Moreover, our method uses temporal superpixels tailored to video data rather than supervoxels such as in VPS that are designed for 3D volumetric data. In contrast to supervoxel methods, with temporal superpixels object parts in different frames are tracked by the same temporal superpixel. Guazzi *et*

al. [13] also use the fusion of several pulse traces but the video is simply divided into contiguous square blocks. The temporal superpixel segmentation is more suited to rPPG algorithms to handle motion scenarios.

The rest of the paper is organised as follows. The method is described in section II with the temporal superpixel segmentation, the pulsatility measure and the fusion procedure. The video dataset, metrics and results are described in section III while the conclusion is presented in section IV.

II. METHOD

The overview of the proposed method is shown in figure 2. The algorithm can be decomposed in three main steps: (1) video stream is first decomposed into temporally consistent superpixels (later called TSP for Temporal SuperPixels). Then, tentative rPPG signal is extracted from each TSP. (2) A pulsatility measure is estimated for each TSP and (3) a weighted average of all the rPPG signals is computed where weights are given by the pulsatility measure.

A. Temporal superpixel-based pulse extraction

The first step of our method is the segmentation of the video stream into temporally consistent superpixels. If a superpixel is a set of pixels that are local, coherent, and which preserve most of the structure necessary for segmentation [16], temporal superpixels can be defined as a set of video pixels that are local in space and track the same part of an object across time [17]. In this work, we use the TSP method proposed by Chang *et al.* [17] which we found to be a good compromise between precision and speed. This method is based on the Simple Linear Iterative Clustering (called SLIC) [18] decomposition. It has been shown that SLIC is among the fastest superpixel methods and is very efficient [19].

In order to extract tentative rPPG signal from each TSP, for each video frame, pixels in each TSP are spatially averaged. The result of this process is a set of N RGB time series $x_i^c(t)$,

where $c \in \{R, G, B\}$ is the color channel, t is the frame index and $i = 1, 2, \ldots, N$ with N is the number of TSP:

$$x_i^c(t) = \frac{\sum_{k=1}^{M_i(t)} I_{k,i}^c(t)}{M_i(t)} \qquad (1)$$

where $M_i(t)$ is the number of pixels in the ith TSP at time t and $I_{k,i}^c(t)$ the kth pixel value at time t and color channel c.

The RGB temporal traces are then pre-processed by zero-mean and unit variance normalization, detrended using smoothness priors approach [20] and band-pass filtered with Butterworth filter. The rPPG signal is then extracted using the chrominance-based method (later called CHROM) [6]. This method applies simple linear combinations of RGB channels and obtains very interesting performance with low computational complexity. Let $y_i^c(t)$ be the RGB time series obtained after pre-processing, CHROM method projects RGB values onto two orthogonal chrominance vectors X_i and Y_i:

$$\begin{aligned} X_i(t) &= 3y_i^R(t) - 2y_i^G(t), \\ Y_i(t) &= 1.5y_i^R(t) + y_i^G(t) - 1.5y_i^B(t). \end{aligned} \qquad (2)$$

The pulse signal S_i of the ith TSP is finally calculated with $S_i(t) = X_i(t) - \alpha Y_i(t)$ where $\alpha_i = \sigma(X_i)/\sigma(Y_i)$. Because X_i and Y_i are two orthogonal chrominance signals, PPG-induced variations will likely be different in X_i and Y_i, while motion affects both chrominance signals identically. S_i is called a *tentative* rPPG signal because some pulse signals calculated from background superpixels do not contain relevant information.

B. Pulsatility measure

Only the skin tissue of an alive subject exhibits pulsatility, therefore pulse signals calculated from some superpixels only contain noise (*e.g.* on non-skin areas). Figure 3 (a) presents a periodogram of a pulse signal estimated from skin area while figure 3 (b) presents periodogram of a pulse signal estimated from the background. In the frequency domain, the pulsatile, cardiac-synchronous signal, exhibits an important peak centered on the fundamental frequency of heart rate, possibly its second harmonic and limited information at other frequencies. To measure the quality of rPPG signals, we estimate signal-to-noise ratio (SNR) defined as the ratio of the power of the main pulsatile component and the power of background noise, compute in dB due to the wide dynamic range of the signals.

The pulsatility measure of the ith TSP is estimated by:

$$SNR_i = 10 \log_{10} \left(\frac{\int_{f_1}^{f_2} h_{signal}^i(f) |\mathcal{F}\{S_i(t)\}|^2 \mathrm{d}f}{\int_{f_1}^{f_2} h_{noise}^i(f) |\mathcal{F}\{S_i(t)\}|^2 \mathrm{d}f} \right) \qquad (3)$$

where $\mathcal{F}\{S_i(t)\}$ is the Fourier transform of the rPPG signal of the ith TSP, f_1 and f_2 the lower and upper limit of the integral defined by the possible physiological range of the heart rate
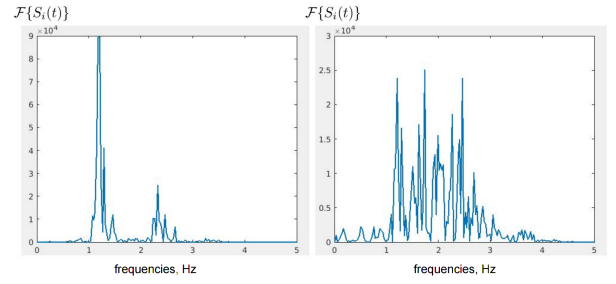

Fig. 3. Periodogram examples of 2 *tentative* rPPG signals estimated from (a) skin area and (b) background.

(40 to 240 bpm in our case), and a double-step function h, for the first and second harmonics, defined by the convolution:

$$\begin{aligned} h_{signal}^i(f) &= [\delta(f - f_0^i) + \delta(f - 2f_0^i)] * \prod(\pm f_r) \\ h_{noise}^i(f) &= 1 - h_{signal}^i(f) \end{aligned} \qquad (4)$$

with δ the Dirac delta function, f_0^i the fundamental frequency (*i.e.* peak of the periodogram), convoluted with the $rect$ function, noted as $\prod$ of half-width f_r. SNR_i will be high for skin TSP and low for background ones.

C. rPPG signal fusion

The final rPPG signal $S(t)$ is obtained by a weighted average of all *tentative* pulse signals $S_i(t)$, *i.e.* $S(t) = \sum_{i \in N} S_i(t) w_i$ where weightings w_i are a function of the main pulsatile component SNR:

$$w_i = \frac{10^{SNR_i}}{\sum_{i \in N} 10^{SNR_i}} \qquad (5)$$

Weights are normalized and in order to conserve the relative contribution of each rPPG signal, weights are defined with the $\log^{-1}(x)$ function (*i.e.* 10^x) The weighting favors TSP that have a high main pulsatile component SNR as these are more likely to represent skin areas. For example, in figure 4, the final rPPG signal is made up of mainly three *tentative* rPPG signals.
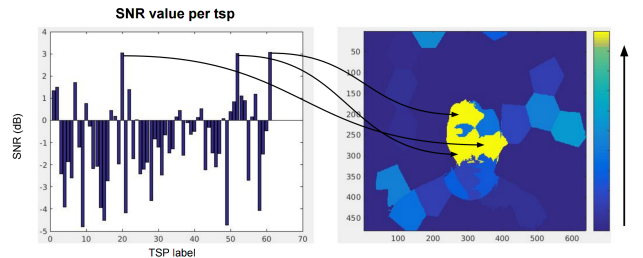

Fig. 4. Examples of SNR values (dB) and its corresponding superpixel.

III. EXPERIMENTS

This section presents the experimental setup for evaluating the proposed method. First, we describe the video dataset. Then, we present the evaluation metrics and finally we compare our implicit ROI selection with regular face detection/tracking and skin detection approach.

A. Video dataset

Videos have been recorded using a simple low cost webcam (Logitech C920 HD Pro) at 30fps with a resolution of 640x480 in uncompressed 8-bit RGB format. During the recording, the subject sits in front of the camera (about 1m away from the camera) with his/her face visible. Participants were asked to sit still but some videos present significant movement (especially at the beginning of the sequence). The dataset is composed of 7 videos (about 16500 frames). All experiments are conducted indoors with a varying amount of sunlight and indoor illumination. To validate the heart rate estimations, PPG is recorded using transmissive pulse-oximeter finger clip sensor (Contec Medical CMS50E) and synchronized with the video. Video frames synchronized with PPG sensor data can be downloaded from our project page[1].

B. Benchmark algorithms and metrics

We compare heart rate estimation of our implicit ROI selection method with classical skin segmentation. To make the comparison with our method fair, chrominance-based method [6] is also used with the skin segmentation method (indistinctly called CHROM or reference method). For the reference method, skin has been segmented based on color feature using the implementation of Conaire *et al.* [21] after face detection [22] and tracking [23]. Examples of skin segmentation results are presented in figure 5. We use the same pre-processing and filtering for both methods.
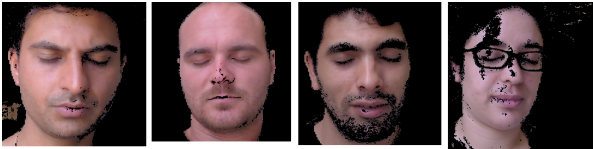

Fig. 5. Skin segmentation result examples for the reference method.

For each video, we estimate heart rate in a sliding window framework. Welch's method is used to obtain the periodogram over a 30 second moving window, with a step size of one second. Heart rate is given by the position of the peaks on the frequency axis. The same heart rate estimation procedure was used on the PPG signal recorded with the contact sensor, on the rPPG signal given by the reference method and the rPPG signal given by our method.

The following metrics are used for comparison:

- **Bland-Altman plots** that measure the agreement between heart rate estimated from rPPG signals and PPG signal.

The lines represent the mean and 95% limits of agreement.
- **Correlation plots** and the Pearson correlation value r^2 between heart rate estimated from rPPG signals and PPG signal.
- **Root mean square error** (RMSE).
- **Precision at 2.5 or 5 bpm**. This metric represents the percentage of estimations where the absolute error is under a threshold (2.5 or 5 bpm).

C. Results

In the first experiment, we evaluate the robustness of the proposed method to TSP resolution varying the number of temporal superpixels N per frame. N varies from 50 to 1000 (gross to fine segmentation). Figure 6 presents precision at 2.5 and 5 bpm. Results are stable, from 95% to 99% at 5bpm and from 89% to 95% at 2.5 bpm. Smaller number of pixels in a TSP leads to larger quantized RGB errors that seams to be compensated by the weighted average procedure. For the following experiments, we use $N = 200$.
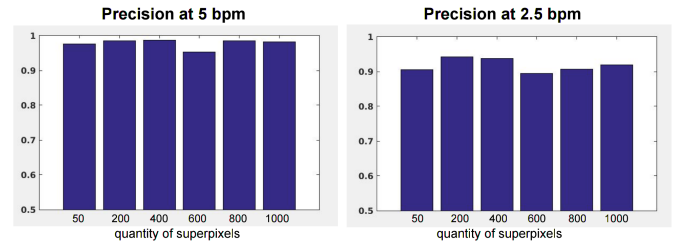

Fig. 6. Precision at 5 and 2.5 bpm calculated on all videos in percentage.

In the second experiment, we compare heart rate estimations with our method and the reference method (called CHROM). Figures 7 and 8 present correlation plots and Bland-Altman plots of the 2 methods. With 200 TSPs per frame, we obtain a Pearson correlation value of 0.9847 compared to 0.9793 for CHROM. From Bland-Altman plots, we can see that the mean bias is -0.36 bpm with 95% limits of agreement 3.4 to 2.7 bpm for our method and the mean bias is -0.55 bpm with 95% limits of agreement 4.1 to 3 bpm for CHROM. We can observe that results are actually very good, even for the reference method. Our method performs slightly better than CHROM. However, the small difference can be explained by the fact that results obtained with the reference are indeed "too good". In table I, we present a summary of all obtained results with a TSP resolution N of 200 and 400 comparedwith the reference. For all metrics, best results are always obtained by our method alternatively with $N = 200$ or $N = 400$.

TABLE I
SUMMARY OF RESULTS

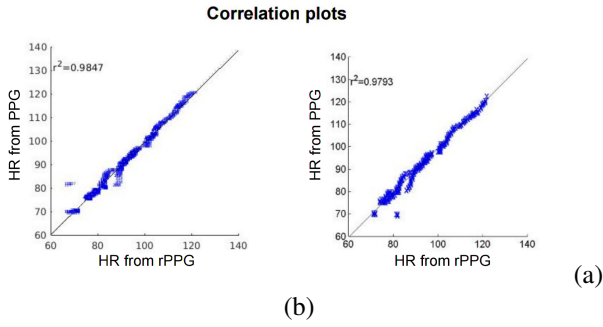| Evaluation metrics	N=200	N=400	CHROM
Precision at 5 BPM	0.9856	**0.9869**	0.9786
Precision at 2.5 BPM	**0.9428**	0.9378	0.9386
RMSE	1.5	**1.43**	2.43
Pearson correlation r^2	**0.9847**	0.9843	0.9793

Fig. 7. Correlation plots obtained from all videos with (a) our method and (b) the baseline CHROM. Abscissa represents heart rate (HR) estimated from rPPG signal and ordinate represents HR estimated from the PPG signal (ground truth).
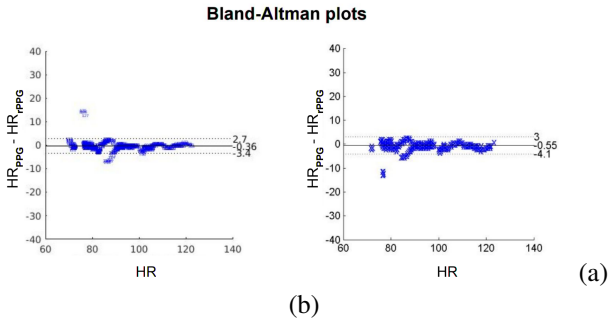

Fig. 8. Bland-Altman plot from all videos with (a) our method and (b) the baseline CHROM. Abscissas represent HR and ordinate represents the difference between HR estimated from PPG (HR_{PPG}) and HR estimated from rPPG (HR_{rPPG}).

IV. CONCLUSION

Remote photoplethysmography technology has a great potential for use in a wide range of clinical assessments, such as homecare, telemedicine and personal healthcare. The selection of ROI is a critical first step of rPPG techniques to obtain reliable pulse signals. In the present study, we have described, implemented, and evaluated a new rPPG method that implicitly select living skin tissue via their particular pulsatility feature. Photoplethysmogram signals are estimated with the weighted fusion of several *tentative* rPPG signals computed on a set of temporal superpixels. Even if the dataset is very simple, the results of this study have demonstrated that the rPPG signals could be remotely estimated without any tedious ROI selection. Our method always performs slightly better than the reference method CHROM.

Further developments include the evaluation of the proposed technique on a larger dataset with more challenging scenarios (motion, illumination variation, compression noise etc.). Also, we will continue the developments to handle cases where multiple persons are in the scene and we will evaluate our implicit ROI selection method with other rPPG algorithms.

ACKNOWLEDGMENT

REFERENCES

[1] Y. Sun and N. Thakor, Photoplethysmography Revisited: From Contact to Noncontact, From Point to Imaging, IEEE Trans. on Biomedical Engineering 63 (3) (2016) 463 - 477.

[2] X.F. Teng and Y.T. Zhang, The effect of contacting force on photoplethysmographic signals, Physiological Measurement 25 (5) (2004) 1323-1335.

[3] W. Verkruysse, L. O. Svaasand, J. S. Nelson, Remote plethysmographic imaging using ambient light, Optics express 16 (26) (2008) 21434-21445.

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

[5] B.S. Kim and S.K. Yoo, Motion artifact reduction in photoplethysmography using independent component analysis, IEEE Trans. on Biomedical Engineering 53 (3) (2006) 566-568.

[6] G. de Haan and V. Jeanne, Robust pulse rate from chrominance-based rppg, IEEE Trans. on Biomedical Engineering 60 (10) (2013) 2878-2886.

[7] W. Wang, S. Stuijk and G. de Haan, A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation, IEEE Trans. on Biomedical Engineering (2015).

[8] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers and Y. Zhu, Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise, Journal of Biomedical Optics 16 (7) (2011).

[9] M. Z. Poh, D. J. McDuff and R. W. Picard, Advancements in non- contact, multiparameter physiological measurements using a webcam. IEEE Trans. on Biomedical Engineering 58 (1) 2011 7-11.

[10] H.E. Tasli, A. Gudi and M. Uyl, Remote PPG based vital sign measurement using adaptive facial regions, In IEEE International Conference on Image Processing (2014) 1410-1414.

[11] W. Wang, S. Stuijk and G. de Haan, Exploiting Spatial-redundancy of Image Sensor for Motion Robust rPPG, IEEE Trans. On Biomedical Engineering 62 (2) (2015) 415-425.

[12] F. Bousefsaf, C. Maaoui, A. Pruski, Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate, Biomedical Signal Processing and Control 8 (6) (2013) 568-574.

[13] A.R. Guazzi, M. Villarroel, J. Jorge, J. Daly, M.C. Frise, P.A. Robbins and L. Tarassenko, Non-contact measurement of oxygen saturation with an RGB camera, Biomedical Optics Express 6 (9) (2015) 3320-3338.

[14] A.A. Kamshilin1, E. Nippolainen, I.S. Sidorov, P.V. Vasilev, N.P. Erofeev, N.P. Podolian and R.V. Romashko, A new look at the essence of the imaging photoplethysmography Scientific Reports 5 (2015).

[15] W. Wang, S. Stuijk, and G. de Haan, Unsupervised Subject Detection via Remote PPG, IEEE Trans. On Biomedical Engineering 62 (11) (2015) 2629-2637.

[16] X. Ren and J. Malik, Learning a classification model for segmentation, In Proc. IEEE Conference on Computer Vision and Pattern Recognition 1 (2003) 10-17.

[17] J. Chang, D. Wei and J. Fisher, A video representation using temporal superpixels, In Proc. IEEE Conference on Computer Vision and Pattern Recognition (2013) 2051-2058.

[18] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, SLIC Superpixels Compared to State-of-the-art Superpixel Methods, IEEE Trans. on Pattern Analysis and Machine Intelligence 34 (11) (2012) 2274-2282.

[19] P. Neubert and P. Protzel, Superpixel benchmark and comparison, In Proceedings Forum Bildverarbeitung, (2012) 1-12.

[20] M.P. Tarvainen, P.O. Ranta-Aho and P.A. Karjalainen, An advanced detrending method with application to HRV analysis, IEEE Trans. on Biomedical Engineering 49 (2) (2002) 172-175.

[21] C.O. Conaire, N. O'Connor and A.F. Smeaton, Detector adaptation by maximising agreement between independent data sources, IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum (2007).

[22] P. Viola and M. Jones, Rapid Object Detection using a Boosted Cascade of Simple Features, IEEE Conference on Computer Vision and Pattern Recognition 1 (2001) 511-518.

[23] C. Tomasi and T. Kanade, Detection and Tracking of Point Features, Carnegie Mellon University Technical Report CMU-CS-91-132 (1991).