Automatic Selection of Webcam Photoplethysmographic Pixels Based on Lightness Criteria Photoplethysmographic Pixels Selection Based on Lightness Segmentation

Frédéric Bousefsaf, Choubeila Maaoui, Alain Pruski

Automatic Selection of Webcam Photoplethysmographic Pixels Based on Lightness Criteria

Photoplethysmographic Pixels Selection Based on Lightness Segmentation

F. Bousefsaf*, C. Maaoui, A. Pruski

Laboratoire de Conception, Optimisation et Modélisation des Systèmes (LCOMS), Université de Lorraine, Bâtiment ISEA (Institut Supérieur d'Electronique et d'Automatique), 7 rue Marconi, 57070 METZ Technopôle, France.

E-mail addresses: frederic.bousefsaf@cea.fr (F. Bousefsaf), choubeila.maaoui@univ-lorraine.fr (C. Maaoui), alain.pruski@univ-lorraine.fr (A. Pruski).

* Corresponding author.
Frédéric Bousefsaf
7 rue Marconi, 57070 Metz (France)
Tel.: +33 387 547 652; fax.: +33 387 547 301
E-mail address: frederic.bousefsaf@cea.fr

Automatic Selection of Webcam Photoplethysmographic Pixels Based on Lightness Criteria

Abstract

We propose, in this study, an original method that was developed to remotely measure the instantaneous pulse rate using photoplethysmographic signals that were recorded from a low-cost webcam. The method is based on a prior selection of pixels of interest using a custom segmentation that used the face lightness distribution to define different sub-regions. The most relevant sub-regions are automatically selected and combined by evaluating their respective signal to noise ratio. Performances of the proposed technique were evaluated using an approved contact sensor on a set of 7 healthy subjects. Different experiments while reading, with motion or while performing common tasks on a computer were conducted in the laboratory. The proposed segmentation technique was compared with other benchmark methods that were already introduced in the scientific literature. The results exhibit high degrees of correlation and low pulse rate absolute errors, demonstrating that the segmentation we propose in this study outperform available region-of-interest selection methods.

Keywords

Remote photoplethysmography

Pulse rate

Webcam

Region of interest

1. Introduction

The ability to measure the cardiac pulse wave along with other vital signs using non-contact technologies has been widely studied over the last ten years [1,2]. Webcams correspond to low-cost and remote sensors that can be employed to non-invasively assess a set of specific physiological signals [3]. Important health parameters, like the cardiac pulse rate [2,4-6], the peripheral arterial oxygen saturation [5,7] or the pulse transit time [8] are computed from these biological signals. The quantified markers are continuously monitored and inspected in order to diagnose specific diseases in-hospital. They can also be employed at-home for personal well-being purpose. Standards of measurements recommend the use of contact technologies, like electrocardiographic probes [9], chest straps and pulse oximeters to respectively monitor the heart rate, the breathing rate and the oxygen saturation. However, these contact sensors can cause irritation and discomfort if they are worn over a long period of time. Infants in neonatal intensive care units [10] and patients that present skin ulcers or burns may not be able to wear the contact probes. Herein, non-contact technologies are employed to eliminate these limitations.

Cameras correspond to imaging devices used to sense the cardiovascular pulse wave by photoplethysmography (PPG), an opto-electronic phenomenon that consists in observing color variations on skin. PPG signals are susceptible to motion-induced artifacts [3], particularly when dealing with webcams and ambient light. Machine learning algorithms can be used to improve the pulse rate estimation. The k-nearest neighbor method was employed on features extracted from the power spectral density of preprocessed PPG signals to improve the results [6]. Low-cost webcams [3,6], specific digital cameras [4] and high speed CMOS cameras [11] were used to measure PPG on skin.

Along with behavioral information, physiological parameters are used in affective computing [12,13] to automatically recognize the emotion of the user or to assess stress. Herein, particular vital signs like the Pulse Rate (PR) and the Pulse Rate Variability (PRV) represent reliable inputs to quantify different forms of stress [14-16]. Contact sensors can be limited in some scopes of application where a specialist must install and monitor them [2]. The probes and wires may also generate a bias by interfering with the user, resulting practically by an erroneous estimation [17]. Through imaging systems, PPG signals were already used to estimate mental stress states [18,19].

1.1 Related works: Region Of Interest (ROI) selection

In theory, all regions covered by skin are employable to measure photoplethysmography. In practice, only the face [3-5] or the hand [1,11] regions are recorded to measure these physiological signals. Different image processing techniques were used to define custom Region Of Interests (ROI) based on detected facial features. All the pixels of the ROI are employed to compute the PPG signal by a spatial averaging operation (see eq. 1) that transforms a set of pixels into a scalar to improve the signal to noise ratio. A previous attempt has been presented by Stricker et al. [20] to compare the effectiveness and accuracy of existing segmentation techniques. These methods can be classified in different categories:

$$signal(t) = \frac{1}{N_x . N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} I(x, y, t) \quad (1)$$

N_x and N_y correspond respectively to the width and height of the source frame. $I(x,y)$ represents the pixel intensity in the particular (x,y) location.

1.1.1 Entire face using face detectors

Particular face tracking techniques were employed to segment the region that covers the whole face from the source frame (see Fig. 1 (b) for a typical example). The most popular detector was originally proposed by P. Viola and M. Jones [21], which uses a pre-trained boosted cascade classifier on each frame to locate faces. The algorithm returns the bounding box of the face and particularly the top-left point coordinates along with the width and height of the box. To only consider the pixels that belong to the face, the width of the box can be reduced to 60% of its original value [3].

1.1.2 Defined regions in the image

The cheeks and the forehead [Fig. 1 (b)] usually correspond to uniform areas with low contrast. In comparison with other skin regions, these areas are usually tracked for their enriched signal to noise ratio [10,22-26]. The regions are manually [10,22-24] or automatically segmented by a set of geometrical operations based, notably, on ratios calculated from the Viola and Jones face detector position [25] or in relation with the eyes position [26]. The regions can be tracked using a set of points of interest that are detected by the Speeded Up Robust Features (SURF) algorithm [27] and tracked using the Kanade–Lucas–Tomasi (KLT) method [28].

McDuff et al. [4,18] have excluded the region around the eyes to avoid noise generated by eye blinking. They have retained the frontal and lower (from the noise to the lower part of the jaw) parts of the face. The latter region has also been employed by Kong and al. [7], which have manually segmented a rectangular area. Also, Li and al. have used the discriminative response map fitting algorithm [29] to locate 66 landmarks inside the rectangular face region given by the Viola and Jones detector. A custom ROI that encompass cheeks, nose and mouth was defined using the landmarks location. Once again, the KLT features tracker was used to track the landmarks position.

1.1.3 Skin detection

The skin detection acts like a mask which retains the skin pixels and discard the others [Fig. 1 (c)]. Only the skin pixels are considered when forming the PPG signal with the spatial averaging operation (see eq. 2). For example, skin pixels can be detected using the YC_bC_r color space representation of the source frame [30]. The filter is established by setting static thresholds on the 3 channels [31]. This particular filter has been afterwards reemployed in another study proposed by Bal [32]. Sahindrakar et al. [33] have employed the *HSV* color space to gather skin pixels using an empirically defined threshold, followed by the computation of a probability map. The latter is computed using the hue histogram and indicates the probability a pixel have to belong to skin. The *grabcut* method has been employed by Stricker et al. [20] to segment the skin from non-skin pixels. Different morphological operations (erosion, dilation and opening) were used to improve the quality of the mask and reduce artifacts.

$$signal(t) = \frac{\sum_{x=1}^{N_x}\sum_{y=1}^{N_y} I(x,y,t).skin(x,y,t)}{\sum_{x=1}^{N_x}\sum_{y=1}^{N_y} skin(x,y,t)} \quad (2)$$

Just like in eq. 1, N_x and N_y respectively represent the width and height of the frame. *I* corresponds to the pixel intensity and *skin* to the binary mask [only the white pixels in the skin mask are considered, see Fig. 1 (c) for a typical example].

1.1.4 Face registration and background segmentation

Tarassenko et al. [5] have employed face registration algorithms to segment the face from the background and from the upper part of the body. 13 different landmarks that describe particular noise, mouth and eyes location points were extracted and tracked using the KLT features tracker. Non-parametric Bayesian image segmentation is then performed to segment the face from the background. Otherwise, Stricker et al. [20] have proposed to employ a deformable model for extracting different fragments of the face. The latter are coupled to an affine warp operation in order to normalize the face with the trained model.

1.1.5 Grids

A raw frame can be divided into a set of sub-windows to produce a reduced image [Fig. 1 (e)]. Each sub-window of the reduced frame is computed by averaging all the included pixels. This preprocessing operation compromises the spatial resolution of the frame but was employed to improve the signal to noise ratio [34]. Just like native frames, either the entire or a partial set of the sub-windows is averaged to form the PPG signal. Specifically, Sun et al. [23] have fixed the sub-windows size at 10×10 pixels, leading to transform a 640×480 native frame into a coarse image that includes only 64×80 pixels. In a more recent study [11] the same researchers have set the sub-windows size at 8×8 pixels but starting from 256×384 raw images. In addition, Verkrurysse et al. [34] have projected their native 640×480 frames on reduced 50×40 images.

1.2 Aim of the study

Even if the skin is perfectly segmented, all the regions don't contain the same PPG signal power [20]. For example, the use of all skin pixels may include non-relevant areas that are not considered when tracking smaller local regions. In contrast, it has been demonstrated than using larger ROI may be a key to significantly improve the signal to noise ratio. Other factors like contrasts and edges orchestrated by the anatomical definition of the face may generate strong light fluctuations during motion and must also be considered, especially when forming a PPG signal with imaging technologies.

In this study, we present a new method that selects different regions of pixels based on the lightness distribution of the face. The signal to noise ratio of each region is computed using a standard power spectral density analysis**.** The most relevant regions are then automatically selected and combined by evaluating their respective signal to noise ratio. To avoid artifacts generated during lips movements, only the upper part of the face was selected as first ROI. Different experiments while reading, with motion or while performing common tasks on a computer were conducted in the laboratory. The results of the proposed segmentation technique were compared with other benchmark methods that were already introduced in the scientific literature. The reference signal was measured using a finger blood volume pulse sensor.

2. Methods

2.1 Preprocessing

The source frames are temporally (over 5 frames) and spatially smoothed [Fig.2 (a)] using a Gaussian filter. Its kernel is defined by the following formula:

$$G_i = \alpha \times e^{\frac{-(i-(ksize-1)/2)^2}{2.\sigma^2}}$$
$$\sigma = 0.3 \times \big((ksize - 1) \times 0.5 - 1\big) + 0.8 \quad (3)$$

ksize corresponds to the kernel size, empirically fixed at 5 for both the vertical and horizontal filters. The length of the *G* vector corresponds to the kernel size, with $0 \leq i \leq ksize\text{-}1$. σ corresponds to the Gaussian standard deviation and is computed from *ksize*. Also, α corresponds to the scale factor and is computed to respect the condition $\sum_i G_i = 1$.

2.2 Face tracking

The face and its primitives are primarily detected using the *Chehra* tracker proposed by Asthana et al. [35]. The detector was constructed using discriminative techniques and particularly through a deformable generic model that is updated over time. Since the details of the techniques employed or developed by the authors are beyond the scope of this paper, we therefore refer to the original article [35]. From a practical point of view, the detector delivers the position of 49 predefined landmarks from the source frame: 9 for the nose, 10 for the eyebrows, 12 for the eyes and 18 the mouth [Fig. 2 (b)]. All the points are smoothed using a basic averaging operation over the last five frames.

Based on the detected landmarks of the nose and the eyes, we create and adjust an ellipse that covers the cheeks and the forehead regions [Fig. 2 (c)]. The mouth is excluded and not considered in this study, the lower part of the face being subject to motion if the user talks. In addition, the presence of hairiness hides skin pixels, leading to a notable reduction of the PPG signal to noise ratio. Only the pixels included in the ellipsoidal ROI are considered for further analysis [Fig.2 (d)].

2.3 Segmentation based on lightness criteria

All the regions of the face may not possess the same quantity of PPG information. The purpose of this operation consists in splitting the face into different sub-regions to perform a separate analysis of their respective PPG signals. We have chosen to split the frame using a set of criteria based on the lightness distribution of the image, formed by computing its histogram [Fig. 2 (e)] using the *L** lightness component of the CIE *L*u*v** color space. This particular color space is indirectly computed from CIE *XYZ*, a derived version of the Red-

Green-Blue (RGB*)* color space [36]. Depending on the total pixel number, the histogram is split into 5 different subclasses, each class containing the same number of pixels. This step is only performed when the application starts. 5 different masks [only three of them are presented in Fig. 2 (f)] containing different pixels of interest are formed using the low and high limits of each subclass. The photoplethysmographic signals are then formed using a standard spatial averaging operation [34].

2.4 Signal to Noise Ratio (SNR) and combination of the most relevant regions

The PPG signals are independently centered around zero by removing their mean value. A detrending operation based on smoothness priors approach [37] is then performed to remove non-relevant frequencies. The smoothing parameter λ was fixed at 50, which concretely corresponds to a cutoff frequency of 0.75 Hz. The Fast Fourier Transform (FFT) of the resulting signals is then performed to compute the SNR [Fig. 2 (g)]:

$$SNR(i) = \frac{\max\left(PSD(i)\right)}{\sum PSD(i) - \max\left(PSD(i)\right)} \tag{4}$$

Where *i* corresponds to the index of the analyzed PPG signal and varies between 0 and 4. *PSD(i)* is the power spectral density of the i^{th} PPG signal, computed with the FFT algorithm. max(*PSD(i)*) returns the maximum amplitude of the power spectral density between the 0.7…4 Hz frequency band.

The ratios are sorted in descending order, the best ratios being at the top of the list. We then combine the sorted masks one after the other, from the best to the worst (eq. 5) and repeat the SNR calculation process. The last combined mask includes all the pixels of the ellipsoidal ROI. We frequently observe an optimum in the evolution of the computed ratios during the combining steps: the SNR typically increases in the first two or three combinations before decreasing. Finally, the best combination is retained [Fig. 2 (h)] to form the PPG signal.

$$combi(n-1) = \sum_{i=0}^{n} mask(i) \tag{5}$$

Where *n* = 1…4. The masks are formed using the lightness histogram (see section 2.3).

3. Materials and experimental procedure

3.1 Experimental procedure

6 experiments (see Table 1) were conducted indoors to evaluate the method on 7 healthy volunteers of both gender and various ages. All participants gave their informed consent before the beginning of a session. The skin types, which correspond to a visual estimation of the participant skin color, are represented using the Fitzpatrick chromatic scale [38]. The skin color is defined between I for white skins and V for black skins. For practical purposes, categories I and II are merged in one set. Following this chromatic scale, four participants were categorized as type I - II, two as type III and one as type IV.

Each experiment in a session was performed on different moments of the day and lasted 1 minute. The light intensity, which was controlled before each experiment, was generally comprised between 800 and 1200 lux. The participants were sitting on a chair in approximately 1 meter from the webcam [see Fig. 3 (a)]. The 42 movies were recorded with a C++ based software and analyzed offline with MATLAB (The MathWorks, Inc.).

The purpose of the first experiment is to evaluate the proposed method in quiet, still and calm conditions. The participants were placed in front of the illumination source (the sunlight entering through windows). The second experiment is similar to the first, except that the illumination source was located perpendicularly to the camera-face trajectory.

In the third experiment, participants had to follow a set of six predefined head rotations: 2 around the vertical axis (yaws, from left-to-right and right-to-left), 2 around the horizontal axis (pitches, from top-to-bottom and bottom-to-top) rotations in addition to a left and a right roll [see Fig. 3 (b)]. The third experiment is similar to the second, except that the participants were asked to perform all the movements more rapidly. The details of the velocity and magnitude for each rotation are presented in Table 1. Practically, these parameters were controlled by computer vision using the *Chehra* tracker [35], which delivers in real time the yaw, pitch and roll values. In addition, markers were physically placed on the scene to limit the rotation angles. A training session was also proposed to the participants just before the operational experiment to avoid misinterpretation and in order to adjust the velocities.

For the two last experiments, the participants were asked to respectively read a text and use the computer as they would in a normal situation [6].

3.2 Materials

A low-cost HD webcam (Lifecam Cinema by Microsoft) was used in these experiments. The resolution of the device was set to 640×480 pixels with an acquisition frequency of 30 frames per second. The three RGB channels are encoded with 8 bits per pixel. It is important to note that auto white balance is disabled in these experiments. White balance locally regulates colors and generates non-desired artifacts in webcam PPG signals. A finger blood volume pulse sensor (BVP-Flex/Pro by Thought Technologies Ltd.) was employed to measure the reference signal with a sampling frequency set at 256 Hz. The light intensity was controlled using a light meter (model TENMA 72-6693 by TENMA) before the beginning of each experiment.

4. Results and discussion

4.1 Data processing

The results of the method presented in section 2 were compared with other ROI benchmark methods proposed in the literature (see Table 2). The entire face position was detected using the face detector originally proposed by P. Viola and M. Jones [21]. The version included in the OpenCV library has been employed in this study. The forehead and cheeks ROI were implemented using a custom algorithm based on the landmarks positions given by the tracker [Fig. 2 (b)].

For fair comparison, we applied the same signal processing algorithm to all the methods. This step is detailed in previous work [19] and is based on the continuous wavelet transform. The latter constructs a time-frequency representation by applying inner products to measure the similarity between an analyzing function and the pulse signals. Briefly, the wavelet representation is filtered in order to remove trends and high frequency artifacts, specifically in the 0.6–4 Hz frequency band. The *Morlet* wavelet has been employed to analyze the time series [19,30]. The signal is reconstructed from the filtered wavelet representation via the inverse wavelet transform. An existing peaks detection algorithm (*findpeaks* function in MATLAB) was employed to find local maxima in order to assess the interbeat intervals and compute the instantaneous pulse rate.

4.2 Selection of the best color component

The PPG signals were formed using several channels computed from different color spaces to assess the component that minimize the root mean square error (RMSE) between its pulse rate series and the reference pulse rate series. The best combination mask (see section 2.4) was used as ROI to compute all the PPG signals.

Specifically, different color and chromaticity components where employed [36,39]:

- a^* and b^* chromaticity components, computed from the CIE $L^*a^*b^*$ color space;
- u^* and v^* chromaticity components, computed from the CIE $L^*u^*v^*$ color space;
- R, G and B color channels;
- C_b and C_r chrominance components, computed from the YC_bC_r color space.

The results are presented in Fig. 4 and summarized in table 3. On the basis of the results presented in this section, the u^* color components presents the better results with a RMSE of 4.81 bpm. When considered separately, the red, green and blue channels deliver the worst results, with a RMSE comprised between 10.18 bpm (for the G channel) and 13.71 bpm (for B). Thus, the u^* component was selected for further analysis.

Pulse rate series that were computed with PPG signals derived from the u^* color component deliver the lowest root mean square error. The CIE $L^*u^*v^*$ color space separates lightness (L^*) from chromaticity components (u^* and v^*). Oxy- and deoxyhemoglobin have a better absorption coefficient between 540 and 577 nm, which corresponds to wavelengths included between green and yellow colors [34]. Thus, PPG fluctuations are more influent in the u^* channel, which represents a red to green color indicator.

When considering the full visible spectrum (typically between 380 and 750 nm) dark skins absorb more incident light than fair skins and, thus, reflect less light to the camera sensor. Consequently, black skins are more sensitive to sensor noise than white skins. In practice, this assumption only affects sensed photoplethysmography by degrading the signal to noise ratio independently of the color components employed. Based on the results presented in this study, the choice of the best color component seems to be uncorrelated with skin color, the u^* channel being the best in most cases. Nevertheless, these statements may be validated in future works with a more detailed analysis, notably by taking a larger number of participants.

In order to improve the general signal to noise ratio, controlling the average brightness of the face by an adaptive procedure may be considered. This adjustment can be achieved by opening the lens aperture and/or by increasing the exposure time. Augmenting the exposure time may result in unwanted motion blur effects. In addition, increasing the exposure time

leads to a reduction of the sampling frequency. This effect may significantly impact pulse rate variability analysis, even if interpolation processes can be employed to improve the time domain resolution [11].

4.3 Comparison between the different methods

Pearson's correlation coefficients along with a statistical analysis and Bland-Altman plots were used to quantify the level of agreement between physiological measurements by the remote and contact techniques. Results of the analysis are summarized in Table 4 and presented in Figs. 5 and 6, where performances of the different ROI-selection methods are revealed. The Bland-Altman plots presented in Fig. 6 represent the differences between estimates from contact and remote measurements against the contact (reference) measurements. Means are represented by dash-dot lines and 95% limits of agreement (± 1.96 SD) by dashed lines.

Based on these beat to beat results, the method we propose in this study presents the most interesting results. The error computed on the instantaneous pulse rate is lower than those derived from the other benchmark methods, independently of motion (Fig. 5). We can observe that the skin detection and the forehead area (benchmark methods III and IV respectively) present a low beat to beat RMSE for all the 6 experiments. These assumptions are also correlated with the Bland-Altman representations, notably when observing the dispersion and the 95% limits of agreement of Fig. 6 (c), (d) and (g).

The skin detection was already used to select pixels of interest in previous studies [20,30,32] and is particularly effective during head rotations. Nevertheless, it has been shown that skin detection based on color models is sensitive to illumination parameters. Thus, adaptive models must be employed to increase the robustness of the system [40]. In contrast, the method we propose in this study is hardware-independent and is fully operational regardless of the camera settings (white balance, gamma, saturation or exposure).

Otherwise and always based on the presented results, the forehead region seems to be quite robust against motion [Fig. 5 (c)] while the cheeks ROI presents a larger error. Important displacements of the cheeks occur when the participant talks or reads. The related motion tends to produce strong variations in the intensity of the pixels comprised in these regions. In contrast, the forehead region is barely affected by these kinds of intensity fluctuations and therefore corresponds to a ROI that is more consistent in order to compute the instantaneous pulse rate by remotely measuring photoplethysmography. From the results presented in Fig. 5, we can also observe that the performances of the forehead and cheeks regions taken simultaneously (benchmark method VI on Fig. 5) are consistently comprised between the performances when considering only the cheeks (benchmark method V) and the performances when considering only the forehead region (benchmark method IV).

As expected, the basic rectangle ROI given by the Viola and Jones face detector delivers the worst results, with a particularly high beat to beat RMSE. The presence of hair and other non-skin pixels tends to add noise in raw PPG signals. Also, our results show that the reduction of the rectangle width, which leads to a better enclosure of the face, produces a lower beat to beat root mean square error (see table 4 and Fig 5). This statement is in correlation with the work proposed before by Poh et al. [3]. The basic and reduced regions delivered by the Viola and Jones face detector tend to produce PPG signals that are particularly sensitive to motion. Strong variations in pixel intensity may appear when the participant talks or moves the head. These fluctuations in intensity are particularly observable in areas located near the edges of the face and inside shadowy regions.

The face was detected and tracked over time using the *Chehra* tracker [35], which is robust to face specifications, notably when user is wearing glasses (both transparent or opaque). The lenses are prone to light artifacts, in particular when specular reflections, which come from light sources or from any other surfaces, are reflected toward the camera. To remove or minimize these unwanted effects, a polarizing filter can be placed right behind the camera lens. The eyeglass frame, which can also generate artifacts, can completely be discarded by the lightness segmentation. Thus and in particular during head rotations, the pixels that compose the eyeglass frame are not considered, the corresponding noise and artifacts being avoided.

In its current form, our method was developed to execute an initial step that selects the different sub-regions (see section 2.3). This step is only performed on the first frame without being refreshed after. Additional work needs to be conducted to update the sub-regions over time, in particular during motion or in presence of important light fluctuations. Also, the distance between the subject and the camera, which directly impacts the pixels number composing the face, is an important parameter that must be considered. Herein, a work has to be conducted in order to determine the number of sub-regions in regard to the number of pixels that composes the face. Otherwise, blind source separation techniques may be used on the red, green and blue signals that belong to each different mask in order to properly recover the underlying PPG signal and compute the instantaneous pulse rate of the subject.

Illumination parameters, like propagation, type (natural and/or artificial), location and diffusion correspond to key parameters that must be considered in these systems based on vision devices. Artificial light, for example, can generate periodic artifacts synchronized to cardiac frequencies [5]. These aliased components tend to decrease the SNR by degrading raw PPG signals. In this paper, we compare ROI selection methods and discuss the relevance of the color spaces, in particular those developed by the CIE. Thus, we choose to keep only a single light source in order to perform fair assessments. Comparing the results under different light sources will be the subject of future work.

Compliance with Ethical Standards

Conflict of interest

Frederic Bousefsaf, Choubeila Maaoui and Alain Pruski declare that they have no conflict of interest.

Ethical approval:

All procedures performed in this study were in accordance with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained from all individual participants included in the study.

References

[1] Humphreys, K., Ward, T., and Markham, C. (2007). Noncontact simultaneous dual wavelength photoplethysmography: A further step toward noncontact pulse oximetry. Rev Sci Instrum 78, 044304.

[2] Kranjec, J., Beguš, S., Geršak, G., and Drnovšek, J. (2014). Non-contact heart rate and heart rate variability measurements: A review. Biomed Signal Process Control, 13, 102–112.

[3] Poh, M. Z., McDuff, D., and Picard, R. W. (2011). Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. IEEE Trans Biomed Eng, 58, 7–11.

[4] McDuff, D., Gontarek, S., and Picard, R. W. (2014). Remote Detection of Photoplethysmographic Systolic and Diastolic Peaks Using a Digital Camera. IEEE Trans Biomed Eng, 61, 2948–2954.

[5] Tarassenko, L., Villarroel, M., Guazzi, A., Jorge J., Clifton D. A., and Pugh, C. (2014). Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. Physiol Meas, 35, 807–831.

[6] Monkaresi, H., Calvo, R. A., and Hong, Y. (2014). A Machine Learning Approach to Improve Contactless Heart Rate Monitoring Using a Webcam. IEEE J Biomed Health Inform, 18, 1153–1160.

[7] Kong, L., Zhao, Y., Dong, L., Jian, Y., Jin, X., Li, B., Feng, Y., Liu, M., Liu, X., and Wu, H. (2013). Non-contact detection of oxygen saturation based on visible light imaging device using ambient light. Opt. Express, 21, 17464–17471.

[8] Shao, D., Yang, Y., Liu, C., Tsow, F., Yu, H., and Tao, N. (2014). Noncontact Monitoring Breathing Pattern, Exhalation Flow Rate and Pulse Transit Time. IEEE Trans Biomed Eng, 61, 2760–2767.

[9] Task Force of the European Society of Cardiology, North American Society of Pacing and Electrophysiology (1996). Heart Rate Variability : Standards of Measurement, Physiological Interpretation, and Clinical Use. Circulation, 93, 1043–1065.

[10] Aarts, L. A. M., Jeanne, V., Cleary, J. P., Lieber, C., Nelson, J. S., Oetomo, S. B., and Verkruysse, W. (2013). Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit — A pilot study. Early Hum Dev, 89, 943–948.

[11] Sun, Y., Hu, S., Azorin-Peris, V., Kalawsky, R., and Greenwald, S. (2013). Noncontact imaging photoplethysmography to effectively access pulse rate variability. J Biomed Opt, 18, 061205.

[12] Healey, J. A. and Picard, R. W. (2005). Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. IEEE Trans Intell Transp Syst, 6, 156–166.

[13] Hernandez, J., Morris, R., and Picard, R. (2011). Call Center Stress Recognition with Person-Specific Models. In Affective Computing and Intelligent Interaction, 6974, 125–134.

[14] Hoover, A., Singh, A., Fishel-Brown, S., and Muth, E. (2012). Real-time detection of workload changes using heart rate variability. Biomed Signal Process Control, 7, 333–341.

[15] Melillo, P., Bracale, M., and Pecchia, L. (2011). Nonlinear Heart Rate Variability features for real-life stress detection. Case study: students under stress due to university examination. Biomed Eng OnLine, 10, 96.

[16] Yoshino, K. and Matsuoka, K. (2012). Personal adaptive method to assess mental tension during daily life using heart rate variability. Methods Inf Med, 51, 39–44.

[17] Pavlidis, I., Dowdall, J., Sun, N., Puri, C., Fei, J., and Garbey, M. (2007). Interacting with human physiology. Comput Vis Image Underst, 108, 150–170.

[18] McDuff, D., Gontarek, S., and Picard, R. (2014). Remote measurement of cognitive stress via heart rate variability. presented in the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Chicago, USA, 2957–2960.

[19] Bousefsaf, F., Maaoui, C., and Pruski, A. (2014). Remote detection of mental workload changes using cardiac parameters assessed with a low-cost webcam. Comput Biol Med, 53, 154–163.

[20] Stricker, R., Muller, S., and Gross, H. M. (2014). Non-contact video-based pulse rate measurement on a mobile service robot. IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 1056–1062.

[21] Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA, I–511–I–518.

[22] Holton, B. D., Mannapperuma, K., Lesniewski, P. J., and Thomas, J. C. (2013) Signal recovery in imaging photoplethysmography. Physiol Meas, 34, 1499–1511.

[23] Sun, Y., Hu, S., Azorin-Peris, V., Greenwald, S., Chambers, J., and Zhu, Y. (2011). Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. J Biomed Opt, 16, 077010.

[24] Takano, C. and Ohta, Y. (2007) Heart rate measurement based on a time-lapse image. Med Eng Phys, 29, 853–857.

[25] Feng, L., Po, L. M., Xu X., Li, Y. and Ma, R. (2014). Motion-Resistant Remote Imaging Photoplethysmography Based on the Optical Properties of Skin. IEEE Trans Circuits Syst Video Technol, 25, 879–891.

[26] Lewandowska, M., Rumiński, J., Kocejko, T., and Nowak, J. (2011). Measuring Pulse Rate with a Webcam – a Non-contact Method for Evaluating Cardiac Activity. Presented at the Federated Conference on Computer Science and Information Systems, Szczecin, Poland, 405–410.

[27] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-Up Robust Features (SURF). Comput Vis. Image Underst, 110, 346–359.

[28] Shi, J. and Tomasi, C. (1994). Good features to track. Presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, USA, 593–600.

[29] Li, X., Chen, J., Zhao, G., and Pietikainen, M. (2014). Remote Heart Rate Measurement from Face Videos under Realistic Situations. Presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, USA, 4264–4271.

[30] Bousefsaf, F., Maaoui, C., and Pruski, A. (2013). Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. Biomed Signal Process Control, 8, 568–574.

[31] Mahmoud, T. M. (2008). A new fast skin color detection technique World Acad Sci Eng Technol, 43, 501–505.

[32] Bal, U. (2014). Non-contact estimation of heart rate and oxygen saturation using ambient light. Biomed Opt Express, 6, 86–97.

[33] Sahindrakar, P., Haan, G., and Kirenko, I. (2011). Improving Motion Robustness of Contact-less Monitoring of Heart Rate Using Video Analysis. Graduation Project, Eindhoven University of Technology, Eindhoven, Netherland. Available online: http://alexandria.tue.nl/extra1/afstversl/wsk-i/sahindrakar2011.pdf

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

[35] Asthana, A., Zafeiriou, S., Cheng, S., and Pantic, M., Incremental Face Alignment in the Wild. Presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, USA, 1859–1866.

[36] Poynton, C. (2003). The CIE system of colorimetry. in Digital Video and HDTV: Algorithms and Interfaces. San Francisco, USA: Morgan Kaufmann Publishers Inc., 211–231.

[37] Tarvainen, M. P., Ranta-Aho, P. O., and Karjalainen, P. A. (2002). An advanced detrending method with application to HRV analysis. IEEE Trans Biomed Eng, 49, 172–175.

[38] Fitzpatrick, T. B. (1975). Sun and skin (Soleil et peau). J. Med. Esthétique, 2, 33–34.

[39] Smith, A. R. (1978). Color gamut transform pairs. Presented at the 5th annual conference on Computer graphics and interactive techniques, ACM Press, New York, 12–19.

[40] Hajiarbabi, M. and Agah, A. (2014). Face Detection in Color Images using Skin Segmentation. J Autom Mob Robot Intell Syst, 8, 41–51.

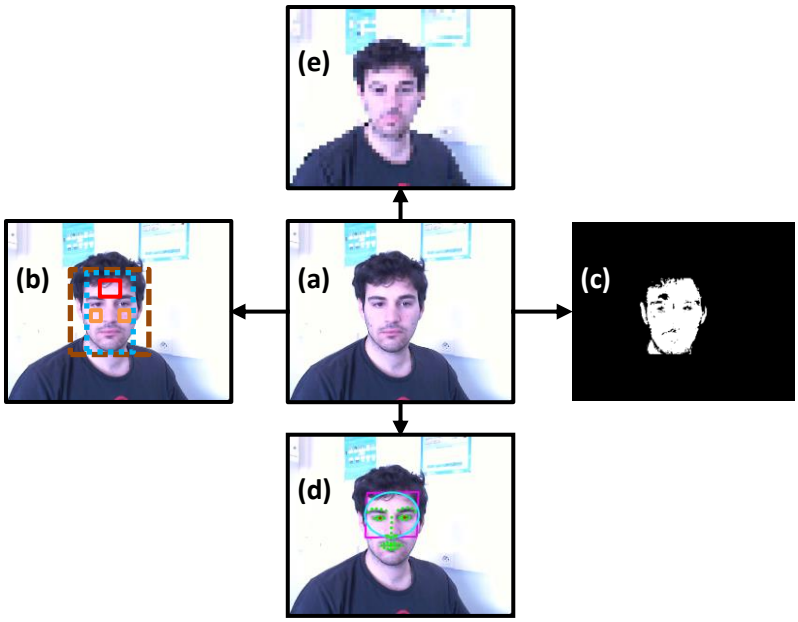

Fig. 1. Different ROI can be defined to measure a photoplethysmographic signal using imaging devices. (a) Source image. (b) Defined regions: original Viola and Jones face detection, before (brown dashed line) and after performing a 40% width reduction (blue dotted line). The forehead (red rectangle) and the cheeks (orange rectangles) can also be segmented to measure the PPG signal. (c) Skin pixels are isolated through a particular mask (white pixels) to properly collect PPG pixels. (d) Custom models can be employed to define particular regions (light-blue ellipse) that exclude the lower part of the face. (e) The source frame is reduced via the computation of local sub-windows to improve the signal to noise ratio.

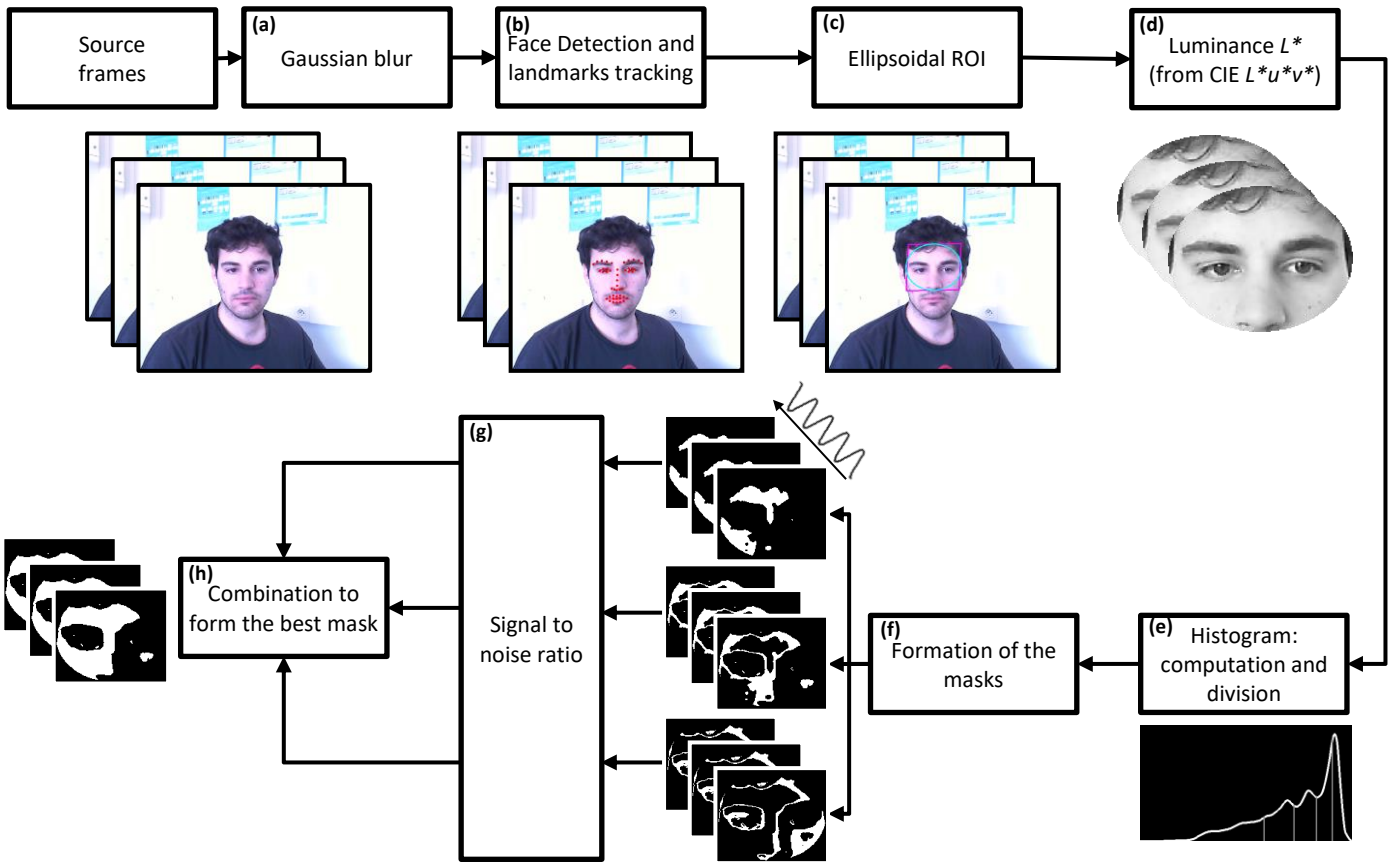

Fig. 2. Overview of the ROI segmentation method.

Table 1. Experiments details

Description
Still and quiet. Source light : sunlight (frontal)
Still and quiet. Source light : sunlight (side)
Slow rotations[a]. Source light : sunlight (side)
Normal rotations[b]. Source light : sunlight (side)
Reading. Source light : sunlight (side)
Using the computer [6]. Source light : sunlight (side)

[a] Yaw: ≈ ±10°/sec (max: ±25°). Pitch: ≈ ±7°/sec (max: ±15°). Roll: ≈ ±10°/sec (max: ±25°).

[b] Yaw: ≈ ±20°/sec (max: ±25°). Pitch: ≈ ±13°/sec (max: ±15°). Roll: ≈ ±20°/sec (max: ±25°).

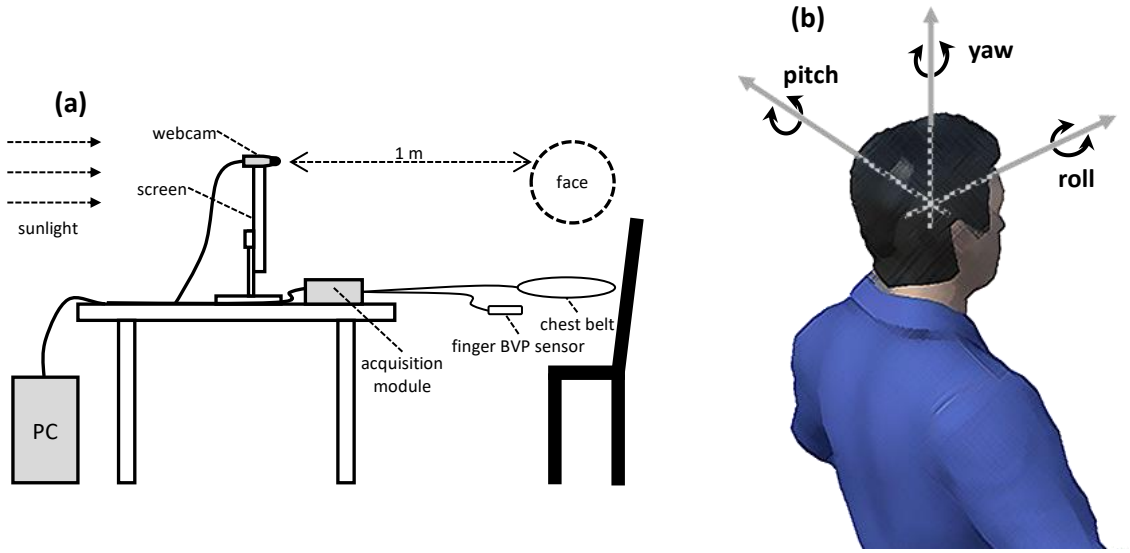

Fig. 3. (a) Experimental setup. The subject is seated at approximately 1 m from the webcam. (b) For the two experiments with motion, the head rotations were controlled following the three yaw, pitch and roll axes.

Table 2. Benchmark methods

Method n°	Description
I	whole face [20,22]
II	whole face, but reduced to 60 % of its original width [3]
III	skin detection [20,30,33]
IV	forehead area [26]
V	cheeks regions [25]
VI	forehead and cheeks regions

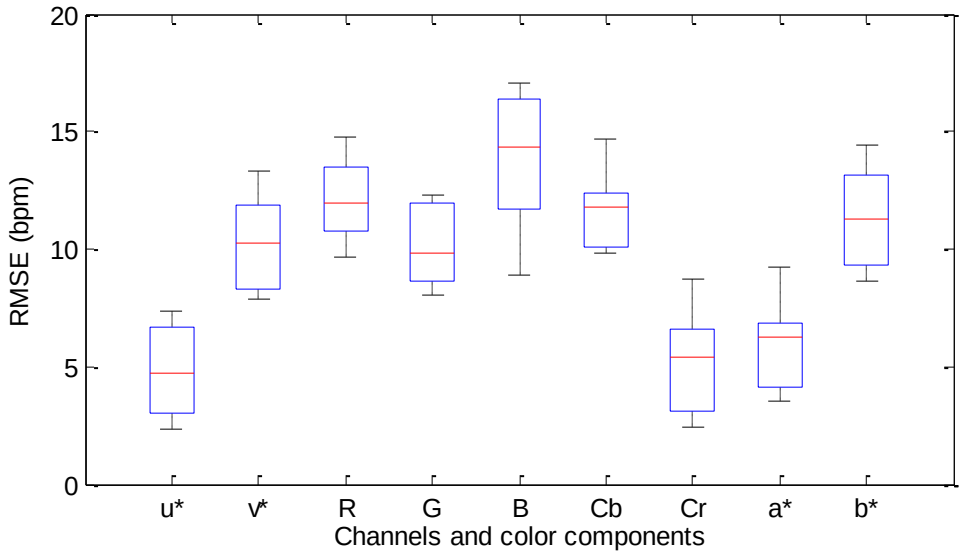

Fig. 4. Boxplot representing the root mean square error for each color components between the reference measurements and the method we propose in this study. The representations integrate the data of the 7 participants for all the 6 experiments.

Table 3. Statistical analysis of the different color components for all the subjects and all the experiments. The instantaneous pulse rate series of the contact sensor were used as references to compute the presented results.

components	RMSE (bpm)						Correlation
	mean	min	25^{th} percentile	median	75^{th} percentile	max	
*u**	**4.81**	**2.34**	**3.04**	**4.72**	6.67	**7.38**	**0.67 ± 0.15**
*v**	10.24	7.91	8.26	10.26	11.84	13.29	0.26 ± 0.11
R	12.1	9.66	10.74	11.98	13.53	14.79	0.2 ± 0.08
G	10.18	8.08	8.68	9.86	11.93	12.28	0.3 ± 0.08
B	13.71	8.9	11.71	14.35	16.39	17.1	0.14 ± 0.14
Cb	11.64	9.8	10.07	11.78	12.37	14.65	0.18 ± 0.06
Cr	5.09	2.43	3.1	5.43	**6.58**	8.75	0.65 ± 0.15
*a**	5.84	3.56	4.16	6.27	6.86	9.22	0.57 ± 0.14
*b**	11.42	8.68	9.32	11.31	13.13	14.45	0.23 ± 0.08

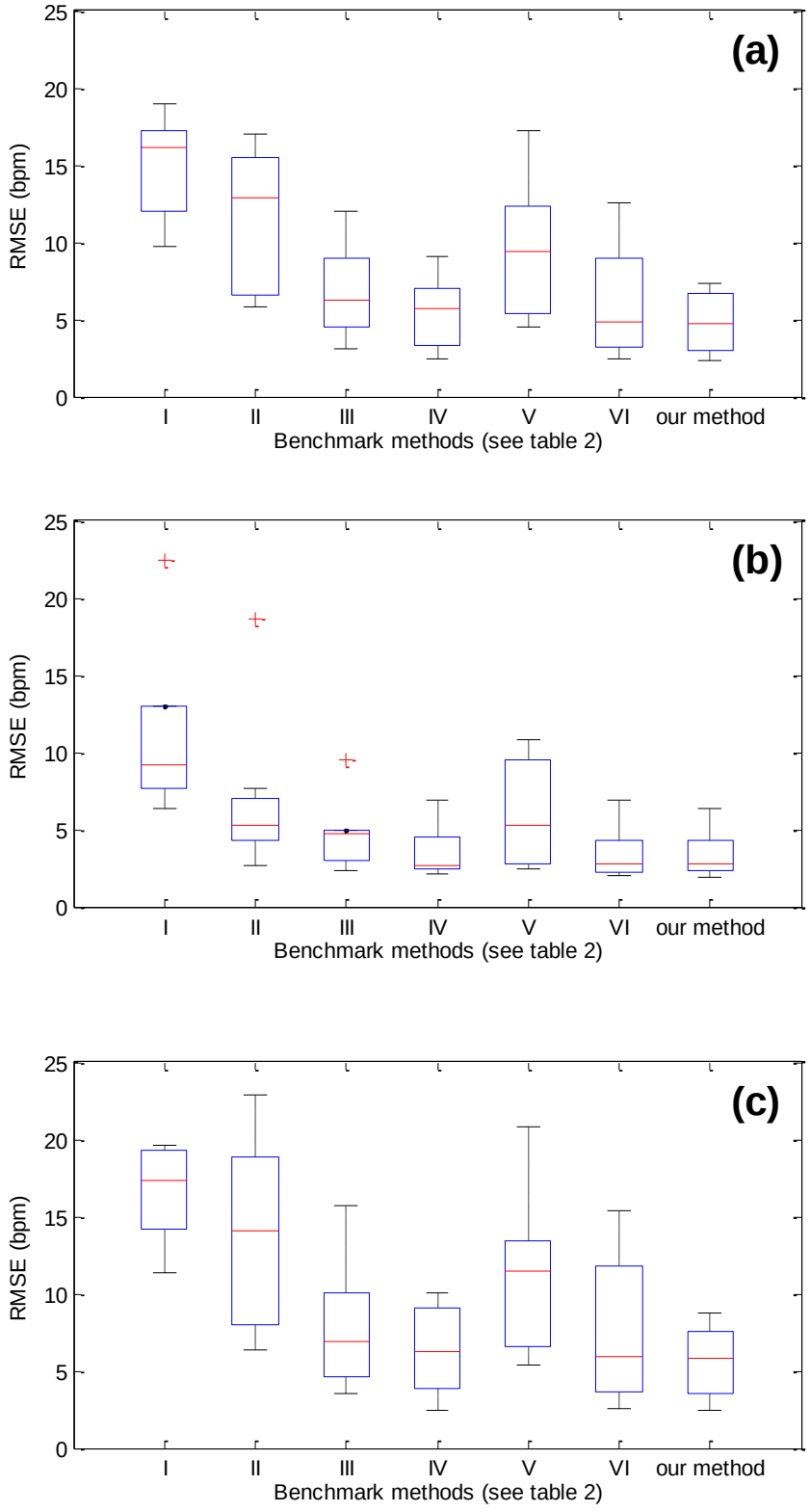

Fig. 5. Boxplots representing the root mean square error for the pixel-detection method proposed in this study along with the other benchmark methods. The representations integrate the data of the 7 participants, (a) for all the 6 experiments, (b) for the 2 experiments without motion and (c) for the 4 experiments with motion.

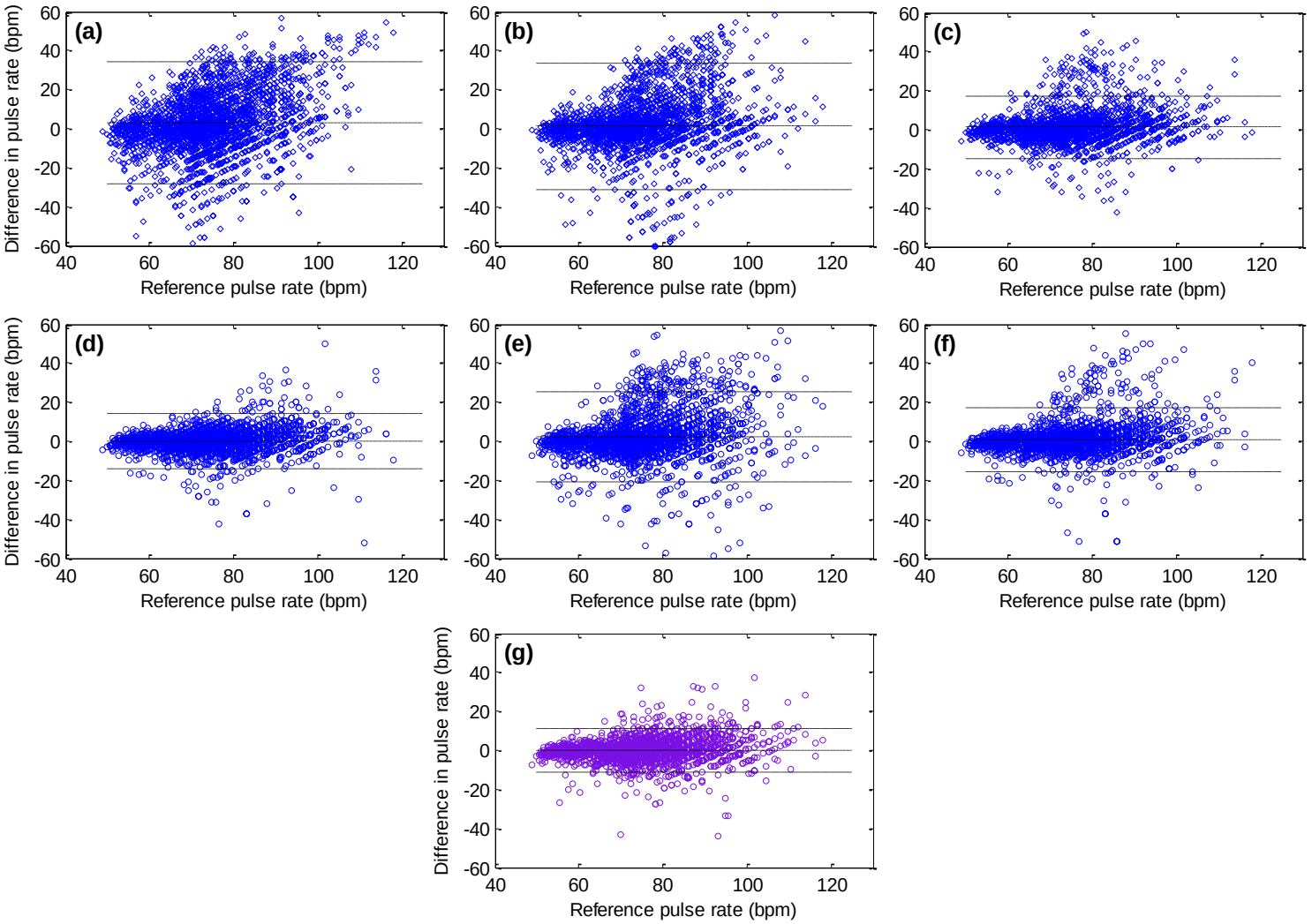

Fig. 6. Beat to beat Bland-Altman plots showing the differences in pulse rate between the webcam and the contact measurements plotted against the contact measurements. Means are represented by dash-dot lines and 95% limits of agreement (± 1.96 SD) by dashed lines. Figures (a) to (f) represent the data for the I to VI benchmark methods respectively (see table 2) while figure (g) represents the data for the method proposed in this study. These representations integrate the data of the 7 participants for all the 6 experiments.

Table 4. Statistical analysis of the different ROI selection methods for all the subjects and all the experiments. The instantaneous pulse rate series of the contact sensor were used as references to compute the presented results.

Methods	RMSE (bpm)			Bland-Altman statistics (bpm)			Correlation
	mean	min	max	mean bias	-95% LoA[a]	+95% LoA	
I	14.81	9.71	18.99	2.77	-28.53	34.07	0.27
II	11.38	5.78	17.01	1.17	-31.14	33.47	0.53
III	6.93	3.11	12.03	1.22	-14.68	17.12	0.66
IV	5.41	2.45	9.03	**-0.01**	-14.07	14.05	0.75
V	9.39	4.46	17.27	2.18	-20.79	25.14	0.56
VI	6.09	2.47	12.56	0.81	-15.83	17.45	0.77
Our method	**4.81**	**2.34**	**7.38**	0.16	**-10.95**	**11.26**	**0.78**

[a] Limits Of Agreement