



HAL
open science

Comparison of region of interest segmentation methods for video-based heart rate measurements

Peixi Li, Yannick Benezeth, Keisuke Nakamura, Randy Gomez, Chao Li, Fan
Yang

► **To cite this version:**

Peixi Li, Yannick Benezeth, Keisuke Nakamura, Randy Gomez, Chao Li, et al.. Comparison of region of interest segmentation methods for video-based heart rate measurements. IEEE International Conference on Bioinformatics and Bioengineering, Oct 2018, Taichung, Taiwan. hal-01939138

HAL Id: hal-01939138

<https://u-bourgogne.hal.science/hal-01939138>

Submitted on 29 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparison of region of interest segmentation methods for video-based heart rate measurements

Peixi Li¹, Yannick Benezeth¹, Keisuke Nakamura², Randy Gomez², Chao Li³, Fan Yang¹

¹ Le2i EA7508, Arts et Métiers, Univ. Bourgogne Franche-Comté, Dijon, France

² Honda Research Institute Japan Co., Ltd., 8-1 Honcho, Wako-shi, Saitama, Japan

³ State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China

Abstract—Conventional contact photoplethysmography (PPG) sensors are not suitable in situations of skin damage or when unconstrained movement is required. As a consequence, remote photoplethysmography (rPPG) has recently emerged because it provides remote physiological measurements without expensive hardware and improves comfort for long-term monitoring. RPPG estimation methods use the spatially averaged RGB values of pixels in a Region Of Interest (ROI) to generate a temporal RGB signal. The selection of ROI is a critical first step to obtain reliable pulse signals and must contain as many skin pixels as possible with a low percentage of non-skin pixels. In this paper, we experimentally compare seven ROI segmentation methods in the perspective of heart rate (HR) measurements with dedicated metrics. The algorithms are compared using our in-house database *UBFC-RPPG*, comprising of 53 videos specifically geared towards rPPG analysis.

Keywords-remote photoplethysmography (rPPG); Heart Rate (HR); Region of Interest (ROI);

I. INTRODUCTION

THE principle of photoplethysmography (PPG) 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. With the emergence of video-based health care monitoring, remote photoplethysmography (rPPG) has recently been developed [2] as it allows remote physiological measurements only based on the ambient light and a video camera, hence reducing user constraint and without any expensive and specialist hardware requirement. The very low signal-to-noise ratio (SNR), precisely the ratio between the rPPG signal and all possible noises, represents the main difficulty of this methodology. Some research teams have proposed computer vision or signal processing techniques to improve the robustness of the original method [2]. Current rPPG methods are not yet as accurate as the ECG's measurements nevertheless it seems obvious that, in a close future, they will enable a flexible monitoring of human vital signs (e.g. the heart rate, breathing rate) with a significant reduction of user's constraints or eventually extending the

duration of the monitoring.

As discussed in different review papers [1], [3], many rPPG methods implement a general pipeline-based framework: 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. This pipeline-based framework emphasizes the importance of the common first step of ROI segmentation. Several approaches have been proposed for ROI selection in the video stream. In earlier studies, manual selection of the ROI have been used [2]. ROI segmentation can also be based on the result of classical face detection [4] and tracking algorithms and possibly refined with skin pixel classification [5].

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 (e.g. in [6]) that the quality of the ROI has a direct impact on the quality of the rPPG signal. 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 equals to the square root of the number of pixels used in the averaging process [7]. Second, the quality is also affected by the percentage of non-skin pixels in the ROI. All rPPG algorithms suffer from performance degradation when the ROI is not properly selected.

From these observations, we propose to quantitatively evaluate several ROI segmentation methods in the perspective of HR measurements with dedicated metrics. The algorithms are compared using the two datasets of our in-house database *UBFC-RPPG* [8], comprising of 53 videos specifically geared towards rPPG analysis.

In section 2, seven commonly-implemented ROI segmentation methods are described. The video dataset used to compare those methods is described in section 3 while results and conclusion are presented in sections 4 and 5.

II. ROI SEGMENTATION ALGORITHMS

Although different, most ROI segmentation techniques are based on the result of classical face detection and tracking algorithms and possibly refined with skin pixel classification or more precise ROI definition based on a set of landmarks. In this section, we present seven implemented ROI segmentation methods.

Face detection (*face*)

The easiest way to segment the ROI is to use directly the detected face. It is important to note that all other methods presented in this paper are based on the face detection result. In this experiment, we have used the well-known Viola-Jones face detector [9]. In order to avoid spurious movements of the detected face, we also use Kanade-Lucas-Tomasi tracking [10] algorithm (*cf.* Fig. 1(a)).

Face cropped (*crop*)

The *face* ROI contains a significant amount of non-skin pixels from the background, the hair or the clothes (*cf.* Fig. 1(b)). As suggested by [4], it is possible to simply crop the ROI selecting the center 60% width and height of the box as ROI (*cf.* Fig. 1(b)).

Rule-based skin detection (*rule*)

Because in the context of remote heart rate measurements, we are only interested in skin pixels, another obvious refinement of ROI segmentation method is to perform a pixel-based skin classification on the detected face. Many review papers have been published presenting various color-based skin detection rules (*e.g.* in [11]). It is clearly impossible to implement all possible methods so we have selected a fairly common method [12] in this experiment (*cf.* Fig. 1(c)).

$$\begin{aligned} (R,G,B) \text{ is classified as skin if} \\ R > 95 \text{ and } G > 40 \text{ and } B > 20 \text{ and} \\ \max\{R, G, B\} - \min\{R, G, B\} > 15 \text{ and} \\ |R - G| > 15 \text{ and } R > G \text{ and } R > B \end{aligned} \quad (1)$$

Histogram-thresholding skin detection (*Conaire*)

Macwan et al. [13] uses Conaire et al. [14] in their rPPG algorithm. This skin detector is based on a thresholding of a non-parametric histogram trained using manually annotated skin and non-skin pixels. This method is eventually based on a Look-Up-Table (LUT) and is thus very fast (*cf.* Fig. 1(d)).

Adaptive-range skin detection (*adaptive*)

Instead of using fixed thresholds for detecting skin pixels, it is possible to derive the range of skin pixel values using specific ROI of the face. For example, in this experiment, a small ROI in the center of the face bounding box is used to create a reference skin color specific to the detected face. Similar pixels in the detected face ROI are then segmented to create the ROI mask. The main advantage of this method is that it is person-specific and does not rely on global threshold values (*cf.* Fig. 1(e)).

Graph-cut based skin segmentation (*Graph-cut*)

All previous methods perform pixel based skin detection.

They are usually very fast but are somehow limited because it does not use any spatial information. Another strategy is to formulate the skin segmentation problem as a segmentation problem using for example Graph-cut [15] (*cf.* Fig. 1(f)).

Landmark based skin segmentation (*landmarks*)

Another strategy to select the ROI is to define a polygon based on a set of landmarks. In this experiment, we used the algorithm proposed by Kazemi et al. [16]. We have selected the contour of the face using the detected landmarks as illustrated in Fig. 1(g).

III. VIDEO DATASET

The ROI segmentation algorithms have been evaluated in the framework of remote heart rate measurements. To this end, we have used the two datasets of our in-house database *UBFC-rPPG* [8] comprising of 53 videos. In the first dataset (7 videos), the volunteers were asked to sit still while in the second dataset (46 videos) the subjects were required to play a time sensitive mathematical game that aimed at augmenting their heart rate while simultaneously emulating a normal human-computer interaction scenario. The database which is focused specifically on rPPG analysis was created using a custom C++ application for video acquisition with a Logitech C920 web camera placed at a distance of about 1 meter from the subject. The video was recorded with a frame resolution of 640x480 in 8-bit uncompressed RGB format at 30 frames per second. A CMS50E transmissive pulse oximeter was used to obtain the ground truth PPG data comprising of the PPG waveform as well as the PPG heart rates. The experimental setup with sample images is depicted in Fig. 2. Video frames synchronized with PPG sensor data can be downloaded from our project page¹.

IV. EXPERIMENTS

A. System framework

For each video frame, the segmented ROI is spatially averaged to obtain the RGB values. The result of this process is a RGB time series. The RGB temporal traces are then pre-processed by zero-mean and unit variance normalization, detrended using smoothness priors approach and band-pass filtered with Butterworth filter. The rPPG signal is then extracted using the chrominance-based method (later called CHROM) [17]. This method applies simple linear combinations of RGB channels and obtains very interesting performance with low computational complexity. Let $y^c(t)$ be the RGB time series obtained after pre-processing, where $c \in \{R, G, B\}$ is the color channel, CHROM method projects RGB values onto two orthogonal chrominance vectors X and Y :

$$\begin{aligned} X(t) &= 3y^R(t) - 2y^G(t), \\ Y(t) &= 1.5y^R(t) + y^G(t) - 1.5y^B(t). \end{aligned} \quad (2)$$

¹<https://sites.google.com/view/ybenzeth/ubfcppg>

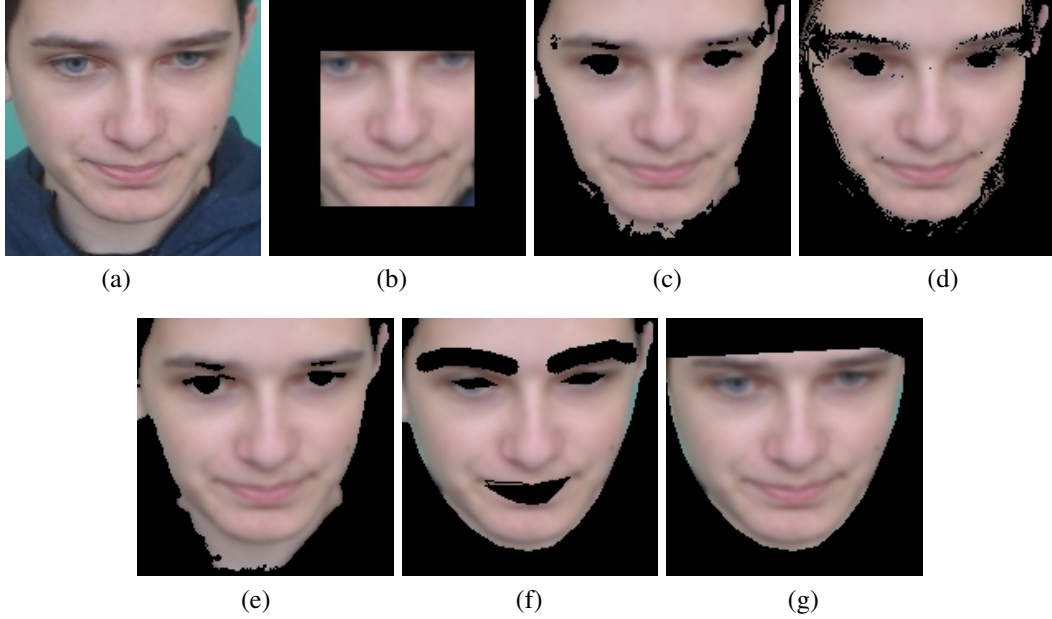


Figure 1. ROI segmentation result examples for (a) face (b) crop (c) rule (d) Conaire (e) adaptive (f) Graph-cut (g) Landmarks.

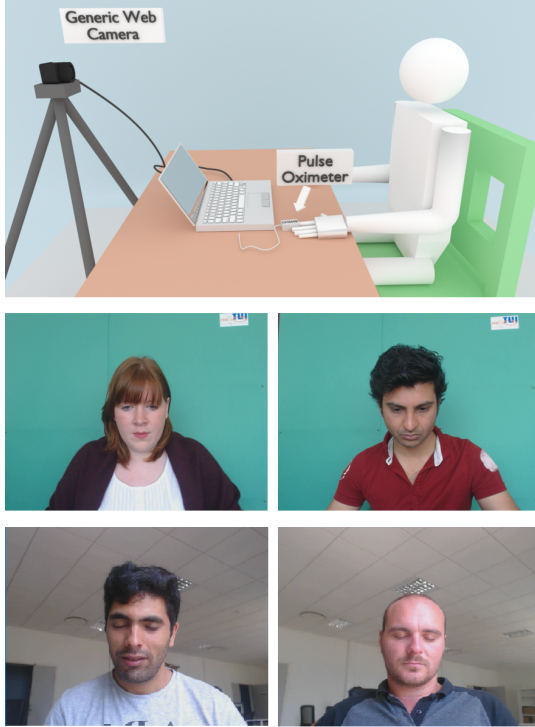


Figure 2. Experimental Setup (top) and sample images from the *UBFC-RPPG* database.

The pulse signal S is finally calculated with $S(t) = X(t) - \alpha Y(t)$ where $\alpha = \sigma(X)/\sigma(Y)$. Because X and Y are two orthogonal chrominance signals, PPG-induced variations will likely be different in X and Y , while motion

affects both chrominance signals identically.

For each video, we estimate heart rate in a sliding window framework. Welch’s method is used to obtain the periodogram over a 20 seconds 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 and on the rPPG signal given by each evaluated ROI segmentation methods.

B. The Evaluation Metrics

To evaluate the performance of the ROI segmentation method in the perspective of heart rate measurements, the following metrics are used:

- **Mean Absolute Error (MAE)** in beats per minute (bpm) is calculated between heart rate estimated from rPPG signals HR_{rPPG} and heart rate estimated from PPG signals HR_{PPG} with $|HR_{rPPG} - HR_{PPG}|$, averaged per video.
- **Root mean square error (RMSE)** between HR_{rPPG} and HR_{PPG} .
- **Signal-to-Noise Ratio (SNR)** is calculated as the ratio of the power of the main pulsatile component and the power of background noise, computed in dB due to the wide dynamic range.
- **Precision at 5 bpm.** This metric represents the percentage of estimations where the absolute error is under a threshold (2.5 or 5 bpm).

C. Results

Averaged results are presented in Table I. First, it is possible to observe that the ROI selection has indeed a large

impact on the final heart rate measurement precision. One example for this idea is that the SNR obtained with *face* is only 2.11 dB while *adaptive* has a SNR of 4.15 dB. Second, it is also possible to observe that the overall ranking changes from one metric to another one. *Graph-cut* for example is the best method according to the MAE but performs poorly according to the SNR. From this observation, it is very difficult to give a final ranking of ROI segmentation method. However, we can observe that *face* obtains consistently the worst result. Then, more complex methods such as *Graph-cut* or *landmarks* obtain interesting results but are significantly computationally more expensive than pixel-based methods. *crop* obtains surprisingly good results in this experiment. One possible explanation is that in all videos, the volunteers remained silent and never opened their mouths. The mouth area is very important in the *crop* ROI (cf. Figure 1(b)). As a consequence, even if it is not the best for all metrics, the histogram thresholding method proposed by Conaire et al. [14] is an interesting ROI selection. It obtains consistently good performance with all metrics and interestingly, this method is very fast because it is based eventually on a simple LUT.

Table I
THE AVERAGE EVALUATION VALUES FOR ROI DETECTION

	Precision5	MAE	RMSE	SNR
<i>face</i>	0.883	4.43	7.19	2.11
<i>crop</i>	0.938	2.35	3.86	2.98
<i>rule</i>	0.909	3.91	5.34	3.11
<i>Conaire</i>	0.931	3.02	4.99	3.02
<i>adaptive</i>	0.913	3.63	5.50	4.15
<i>Graph-cut</i>	0.941	2.34	3.82	2.95
<i>landmarks</i>	0.922	3.21	4.95	2.87

V. CONCLUSION

ROI segmentation is a critical first step in all remote PPG algorithms to obtain a reliable pulse signal. The ROI must contain as many skin pixels as possible with a low percentage of non-skin pixels. From this observation, we present in this paper a comparative study of seven different ROI selection methods. These methods were implemented and evaluated in the perspective of heart rate measurements. The experiments were done with a low cost camera and a contact PPG as the ground truth. 53 videos were recorded and the averaged results of each method were calculated. The results show that the ROI selection has indeed a large impact on the final heart rate measurement precision and that the histogram thresholding method proposed by Conaire et al. [14] obtains in average very good performance with all metrics. Moreover, this method is interestingly very fast because it is based eventually on a simple LUT.

Even if color-based ROI selection method are fast and reliable, these methods do not consider the distinct pulsatility feature of informative ROI, *i.e.* only skin tissue generates pulsatility. In future work, this particular property may help in improving ROI segmentation methods. Second, it has been

shown that the rPPG signal is not distributed homogeneously across the skin, as a consequence the regular spatial averaging of the segmented ROI may not be optimal and offer good opportunities to improve current ROI segmentation techniques.

REFERENCES

- [1] Sun, Y. & Thakor, N. Photoplethysmography revisited: from contact to noncontact, from point to imaging. *IEEE Trans. on Biomedical Engineering* **63**, 463–477 (2016).
- [2] Verkruysse, W., Svaasand, L. O. & Nelson, J. S. Remote plethysmographic imaging using ambient light. *Optics express* **16**, 21434–21445 (2008).
- [3] McDuff, D. J., Estepp, J. R., Piasecki, A. M. & Blackford, E. B. A Survey of Remote Optical Photoplethysmographic Imaging Methods. *int. conf. of the IEEE Engineering in Medicine and Biology Society*.
- [4] Poh, M. Z., McDuff, D. J. & Picard, R. W. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans. on Biomedical Engineering* (2011).
- [5] Wang, W., Stuijk, S. & de Haan, G. A novel algorithm for remote photoplethysmography: Spatial subspace rotation. *IEEE Trans. on Biomedical Engineering* (2015).
- [6] Bousefsaf, F., Maaoui, C. & Pruski, A. Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomedical Signal Processing and Control* **8**, 568–574 (2013).
- [7] Guazzi, A. R., Villarroel, M., Frise, M. C., Robbins, P. a. & Tarassenko, L. Non-contact Measurement of Oxygen Saturation With an RGB Camera. *Journal of Biomedical Optics* **45**, 1764–1771 (2015).
- [8] Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A. & Dubois, J. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters* (2017).
- [9] Viola, P. & Jones, M. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, I–I (2001).
- [10] Lucas, B. D., Kanade, T. *et al.* An iterative image registration technique with an application to stereo vision. *Proceedings DARPA Images Understanding Workshop* 121–130 (1981).
- [11] Kakumanu, P., Makrogiannis, S. & Bourbakis, N. A survey of skin-color modeling and detection methods. *Pattern recognition* **40**, 1106–1122 (2007).
- [12] Kovac, J., Peer, P. & Solina, F. Human skin color clustering for face detection. In *EUROCON 2003. Computer as a Tool. The IEEE Region 8*, vol. 2, 144–148 (IEEE, 2003).
- [13] Macwan, R., Benezeth, Y., Mansouri, A., Nakamura, K. & Gomez, R. Remote photoplethysmography measurement using constrained ica. *IEEE int. conf. on E-Health and Bioengineering* (2017).
- [14] Conaire, C. O., O’Connor, N. E. & Smeaton, A. F. Detector adaptation by maximising agreement between independent data sources. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1–6 (2007).
- [15] Hu, Z., Wang, G., Lin, X. & Yan, H. Skin segmentation based on graph cuts. *Tsinghua Science & Technology* **14**, 478–486 (2009).
- [16] Kazemi, V. & Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1867–1874 (2014).
- [17] de Haan, G. & Jeanne, V. Robust pulse rate from chrominance-based rppg. *IEEE Trans. on Biomedical Engineering* **60**, 2878–2886 (2013).