Conference Paper

Improving ROI detection in photoplethysmographic imaging with thermal cameras

Author(s):
Scebba, Gaetano; Dragas, Jelena; Hu, Suyi; Karlen, Walter

Publication Date:
2017

Permanent Link:
https://doi.org/10.3929/ethz-b-000165104

Originally published in:
http://doi.org/10.1109/EMBC.2017.8037803

Rights / License:
In Copyright - Non-Commercial Use Permitted
Improving ROI detection in photoplethysmographic imaging with thermal cameras

Gaetano Scebba, Student Member, IEEE, Jelena Dragas, Suyi Hu and Walter Karlen, Senior Member, IEEE

Abstract—Photoplethysmographic imaging (PPGi) enables the estimation of heart rate without body contact by analyzing the temporal skin color changes from video recordings. Motion artifacts and atypical facial characteristics cause poor signals and currently limit the applicability of PPGi. We have developed a novel algorithm for locating cheek and forehead region of interests (ROI) with the aim to improve PPGi during challenging situations. The proposed approach is based on the fusion of RGB and far-infrared (FIR) video streams where FIR ROI is used as fall-back when RGB alone fails. We validated and compared the algorithm against the detection based on single sources, using videos from 8 subjects with distinctively different face characteristics. The subject performed three scenarios with incremental motion artifact content (head at rest, intensive head movements, speaking). The results showed that combining the two imaging sources increased the detection rate of cheeks from 75% (RGB) to 92% (RGB+FIR) in the challenging intensive head movement scenario. This work demonstrated that FIR imaging is complementary to simple RGB imaging and when combined, adds robustness to the detection of ROI in PPGi applications.

I. INTRODUCTION

Photoplethysmography (PPG) is a non-invasive measurement of blood volume changes in the vascular system of peripheral tissues [1]. The PPG signal is commonly measured with a pulse oximeter worn on a finger or earlobe. However, a contactless setup might be more suitable in clinical situations, where sensors that require skin contact are too obtrusive, such as in neonatal monitoring.

The contactless PPG signal can be obtained through an analysis of RGB videos, so called PPG imaging (PPGi) [2] or remote PPG [3]. PPGi of exposed skin parts, such as the face or hands, has two main advantages: no sensor-skin contact is required and the measurements can also be performed with consumer cameras that are ubiquitous and inexpensive [4].

The first step of PPGi processing involves the detection of a region of interest (ROI) and tracking this ROI over time. Typical ROIs used in literature are cheeks [5], forehead [6], and whole face [7], [3]. After a ROI has been detected, skin color changes are captured by spatially averaging pixel values within the ROI of each color channel. The green channel has been identified as the most suitable to extract the PPG [3]. The averaged signal is typically band-pass filtered with cut-offs that enclose the human heart beat range between 0.5 and 4 Hz [7].

PPGi is heavily affected by motion artifacts, variation in illumination, and atypical facial characteristics. Head movements such as tilting and rotation and face movements such as frowning, smiling and speaking can produce relevant motion artifacts [8]. Furthermore, non-uniform spatial illumination from ambient light sources introduces noise to the PPG signal extraction [8]. In addition, this non-uniformity also negatively influences face recognition in RGB images and consequently, the ROI detection [7]. Atypical facial characteristics (e.g., face partially occluded by hair or eyeglasses) prevent an accurate face detection. Viola-Jones [9], one of the most common face detectors, fails when facial characteristics are too different from the norm and when background scenes feature skin-similar patterns [10].

Far-infrared (FIR) cameras can partially overcome the above-mentioned limitations of RGB cameras [11]. FIR cameras detect the thermal radiation emitted from an object in the 3-100 µm band, hence the image quality does not deteriorate with illumination changes. Moreover, FIR images provide higher contrast between the human and the background.

In this work, we investigated if the ROI detection benefits from a fusion of FIR and RGB images. We evaluated if the availability of ROIs increased and the quality of the obtained PPG signal improved with the addition of a second imaging source. This work contributes to the adequate detection of ROI in PPGi applications and provides a novel approach for face detection in FIR imaging.

II. ALGORITHMS

We implemented three ROI detection algorithms with different input sources: 1) RGB only, 2) FIR only and 3) RGB+FIR (Fig. 1). All three algorithms used the green channel of the RGB image to estimate the PPG signal, computed as a mean ROI pixel intensity [3]. Forehead, left, and right cheek were selected as ROIs because of their large exposed skin area. For this work we did not implement a ROI tracker into the algorithm sequence as to better observe the performance of the ROI detection. In a real-time algorithm this approach would be too computational intensive and a feature tracker such as Kanade-Lucas-Tomasi [12] would be implemented.

A. RGB Algorithm

A Viola-Jones object detector [9] was implemented to detect a face in the RGB frames. Empirically obtained
thresholds were used to segment the face using a multi-segmentation Otsu algorithm [13]. The algorithm labeled the largest segment above the eyes-region as forehead and the two largest segments below as cheeks (Fig. 2a). In case the Viola-Jones detector failed, the corresponding frame was discarded.

B. FIR Algorithm

We implemented a single-threshold segmentation using the Otsu algorithm to discriminate the face area from the background in the FIR image. The periorbital regions were located using a 95\% threshold of the pixel intensity, as they are the warmest area of the face due to the high vascularization [14]. The largest continuous areas above and below the periorbital region were labeled as forehead ROI and cheek ROIs (Fig. 2b). In case the periorbital region could not be detected, the frame in question was discarded.

C. RGB+FIR Algorithm

We combined the two previously described algorithms to increase reliability of ROI identification. First, the ROI was detected from the RGB source. In case of detection failure, the ROI was detected using the FIR source. The PPG signal was then estimated using the mean of all pixels within the detected ROI.

III. METHODS

A. Hardware setup

RGB videos were recorded with a PlayStation Eye camera (Sony Corp, Tokyo, Japan) and FIR videos were recorded with an uncooled ICI 7640 FIR camera (Camera Inc, Beaumont, USA). The camera lenses were aligned vertically with 5 cm spacing. Both cameras were connected via USB 2.0 to the same computer.

RGB videos were captured with Matlab R2015b (MathWorks Inc, Natick, USA) with a frame rate of 30 Hz and a resolution of 640×480 pixels. FIR videos were captured with IRFlash (Camera Inc, Beaumont, USA) with a frame rate of 10 Hz and a resolution of 640×480 pixels.

Further data processing was done offline after the recording sessions. We used the image acquisition tool from Matlab to synchronize the two video streams at 10 Hz and perform the image registration. We removed conflicting geometrical information due to the different positions of the cameras by computing a geometrical transformation model based on selected point features from the RGB and FIR frames. Then we applied the PPGi algorithms on the multispectral video.

B. Experimental protocol

After institutional ethics board review and informed written consent, healthy volunteers were recruited. Subjects were placed at 1 m distance from the cameras and then exposed to three scenarios, each lasting 2 minutes. At rest (scenario 1) subjects were asked to sit still and watch a movie shown on a large screen in front of them. This way, ideal conditions with minimal motion artifacts were achieved. With intensive head movements in scenario 2, significant motion artifacts were generated. Subjects moved their head following a pre-defined sequence of tilts and rotations. Head tilts in 4 directions were followed by 45° head rotation to left and right. We controlled the subjects movements by pointing into the direction to which the subjects should move their head. Each position was maintained for 5 s before returning to the starting position. Finally, videos of subjects speaking were acquired (scenario 3) as they continuously read a text from a screen placed behind the cameras.

C. Analysis

We introduced two measures for estimating robustness of the PPGi algorithms. 1) For evaluating the performance of ROI detection, we defined the ROI detection rate (%) as follows:

\[
ROI\ detection\ rate = \frac{\#ROI\ detected}{\#number\ of\ frames} \times 100
\]
This metric quantified the rate of success in the detection of a ROI in each frame.

2) The signal to noise ratio (SNR) in dB of the obtained PPG signal was calculated to evaluate its quality. We computed the ratio of the energy of the PPG signal in the 0.9 to 1.7 Hz band and the remaining energy contained in the spectrum, defined as:

$$SNR = 10 \log_{10} \left( \frac{\sum_{f=0.9}^{1.7} PSD(f)}{\sum PSD(f) - \sum_{f=0.9}^{1.7} PSD(f)} \right)$$

where PSD was the power spectral density of the derived PPG signal estimated by Welch’s method. The SNR of the PPG signal was computed only for the subjects whose ROI detection rate exceeded 33%, as lower rates contained too few data points.

IV. RESULTS

A. Subjects and Data

Two female and six male subjects between 25 and 27 years of age were recruited. Four subjects were wearing eyeglasses, two had the forehead partially covered by hair, and one male had cheeks partially covered by a beard. A total of 48 minutes of video were recorded and available for analysis.

B. ROI detection rate

The median ROI detection rate at rest for the left cheek increased from 95.8% (RGB) to 100% (RGB+FIR) and the failure of detecting the majority of ROIs was reduced for two subjects (#4 and #8) (Fig. 3). During intensive head movements in scenario 2, the median ROI detection rate increased from 75% (RGB) to 92.8% (RGB+FIR) with only subject #5 featuring poor ROI detection rates. During the speaking scenario, the median ROI detection rate increased from 88.3% (RGB) to 100% (RGB+FIR), where all subjects showed good ROI detection rates. Figure 3 is also representative for the other studied ROIs, i.e. right cheek and forehead, in which similar improvements in ROI detection rates were observed.

C. SNR

The SNR was calculated only for 6 out 8 subjects, as two subjects had ROI detection rates below 33%. High variability in the SNR of the estimated PPG signal was observed between the three different ROI detection algorithms for all three measurement scenarios (Fig. 4).

V. DISCUSSION

We have presented a novel approach for ROI detection based on the fusion of visible and thermal imaging cameras. We evaluated the performance of ROI detection algorithms based on RGB input, FIR input, and the combination of both inputs. The performance was evaluated using videos of a diverse group of subjects who performed challenging tasks, such as intensive head and mouth movements. The proposed RGB+FIR algorithm for cheeks and forehead ROI detection was compared to the RGB- and FIR-only algorithms. We demonstrated an improvement in detection rate by using a combined ROI detector operating both in the visible and in the FIR spectrum.

The full advantage of fusing multispectral image sources was demonstrated in the scenario in which the subjects performed intensive head movements. Here, the algorithm that combined two cameras clearly outperformed the algorithms using RGB- and FIR-only cameras. The median values of ROI detection rate at rest (scenario 1) and during speaking (scenario 3) did not exhibit considerable difference between the tested algorithms. However, the combined RGB+FIR detector outperformed the conventional RGB detector in two subjects (#4 and #8). These two subjects were challenging
cases as they had the face partially covered by either hair or eyeglasses.

Three subjects in the study had particularly challenging facial characteristics: subject #4 had the face partially covered by hair, disturbing the detection of the forehead ROI; subject #5 was wearing eyeglasses that reflected light and hindered the RGB face detection; subject #8 had a beard and brown skin that both prevented cheek detection. In these three subjects, the advantage of using the combined algorithm compared to RGB-only was clearly noticeable.

No clear preference over different algorithms emerged from the SNR analysis of the PPG signal. The ROI detection that was individually performed in each frame, did not guarantee a constant ROI pixel area over time. In some consecutive cases, this lead to sudden changes of average pixel intensity, which negatively affected the quality of the PPG signal. Adding a ROI tracking algorithm will reduce these sudden changes, which could help improve the SNR of the PPG signal. In this work, we purposely excluded a tracking algorithm to better analyze the performance of the ROI detectors. The quality of the PPG signal could further be improved with a more sophisticated method than the mean ROI pixel intensity calculation.

We used SNR to evaluate the quality of the obtained PPG signal. The SNR alone is not a good indicator of the quality of the PPG, which partially could explain the high SNR variability in our data. Therefore, further research should include an evaluation of the derived heart rate compared to a reference measurement.

Adding a FIR camera increased the complexity of the system, requiring additional image registration and calculation of two separate ROI detection algorithms. However, the recent reductions in cost and size of uncooled FIR cameras make this technology widely accessible. The adoption of FIR cameras could significantly improve the reliability of PPGi systems in health applications and therefore, the extra engineering effort and costs can be justified.

VI. CONCLUSION

In this work, we demonstrated that the adoption of a FIR camera adds robustness to the detection of ROI when compared to previously used methods that include only RGB cameras. Particularly in challenging situations where head movements are significant or where facial features are atypical or obstructed, the use of multispectral information containing a FIR component is advantageous. The effects of improved ROI detection rates on the robustness of PPG signal estimation and its derived vital signs, like heart rate, require further investigation.

REFERENCES