Imaging Photoplethysmography: What are the Best Locations on the Face to Estimate Heart Rate?

Sibylle Fallet[1], Virginie Moser[2], Fabian Braun[1,2], Jean-Marc Vesin[1]

[1]Swiss Federal Institute of Technology, Lausanne, Switzerland
[2]Swiss Center for Electronics and Microtechnology (CSEM), Neuchâtel, Switzerland

Abstract

The potential of imaging photoplethysmography for cardiovascular monitoring applications has been demonstrated recently. Various processing schemes have been proposed to extract heart rate (HR) from a defined region of interest (ROI) on the face. However, the reasons that motivate the choice of the ROI are often unclear. This study aimed at investigating the spatial distribution of the HR-related information on the subject face, based on a partition into 260 small ROIs. A power spectral density (PSD) analysis was performed to determine the amount of HR-related information in each ROI. Normalized color maps were used to visualize the spatial distribution of the HR-information and clearly showed that color fluctuations due to blood volume changes are always more pronounced in the forehead region. After face segmentation, for the R/G/B/NIR channels, average percentages of power of 27%/43%/27%/36% for the forehead region, 17%/28%/18%/21% for the cheek region and 16%/24%/16%/20% for the whole face were obtained.

1. Introduction

Imaging photoplethysmography has recently emerged as a promising technology. This technique enables the remote sensing of blood volume changes occurring at each heartbeat from facial images of the subjects. A simple camera is sufficient to capture the video sequences, and normal ambient light has proven to be a suitable illumination source [1], making this technique really attractive. It is necessary to process the video sequences to extract the imaging photoplethysmographic (iPPG) signals, which are derived from the very slight skin color changes occurring due to pulsatile perfusion. Most of the time, the skin region of interest (ROI) chosen to compute the iPPG signals is selected on the subject's face. For instance, it has been shown that an ROI encompassing the whole face area can be used to recover the iPPG signals and further estimate HR [2, 3]. However, eye movements and blinking can in-

duce artifacts. For this reason, the authors of [4] decided to exclude the region around the eyes from the whole-face ROI. The selection of the ROI was also questioned in [5], where two ROIs of different sizes were compared. A similar accuracy was reported for an ROI on the forehead and a whole-face ROI, and the authors concluded that the selected forehead ROI was representative of the whole-face region. Additional investigations were carried out in [1], where the iPPG signals derived from different ROIs were visually compared. More specifically, four different ROIs were considered, namely the whole face, the upper head, a forehead rectangle and a single pixel on the forehead. The authors concluded that the selection of ROI is not critical for the determination of HR, but they stressed the fact the signal-to-noise ratio can be reduced by averaging a larger number of pixels. On the other hand, when the whole face is used, some pixels do not contribute to HR signal, decreasing the signal quality. Consequently, for iPPG applications, a compromise has to be made between the number of pixels in the ROI and their relevance.

The aforementioned studies have shown promising results for the quite recent iPPG contactless technology. However, we noted that the reasons behind the choice of the ROI remain obscure. This study aims at providing some rationale for the choice of the ROI in iPPG applications. For this purpose, the power in the HR band for different regions of the face was computed from the power spectral density of the iPPG signals. The results were correlated with the reference HR values, derived from the ECG signals acquired simultaneously with the videos.

2. Methods

2.1. Data

The database is composed of six 4-minute recordings. For each recording, the following data were recorded simultaneously: one-lead ECG and video of the upper body region. The three subjects were lying on the floor and were asked to move the least possible while performing a isometric hand-grip exercise or modulating their respira-

tion according to a given protocol in order to induce HR changes. The videos were recorded using both an RGB camera and a NIR camera and were sampled at 20 frames per second with a resolution of 1.3 megapixels. The reference HR was derived from the ECG waveform. For this purpose, a local maxima detection was first applied in order to detect the R-waves and extract the RR intervals. Then, the RR intervals were uniformly re-sampled at 4 Hz to compute the reference HR.

2.2. Recovery of iPPG signals

For each video sequence, a rectangular box encompassing the whole face of the subject was manually selected. Then, this area was divided into 260 (20x13) small ROIs of equal size (see Figure 1). For each ROI, the iPPGs signals were computed by spatial averaging of the pixels, for each channel and at each frame. As a pre-processing step, the resulting signals were band-pass filtered between 0.6 and 4 Hz using a Butterworth band-pass filter (forward and backward filtering).

Figure 1. Illustration to show how the face area was divided into 260 small ROIs which were indivually processed.

2.3. Power spectral density analysis

A 10-second centered sliding window (50% overlap) was used to compute the power spectral density (PSD) of the iPPG signals, using a parametric spectral estimation method based on the Yule-Walker method. An order of 30 was empirically chosen. For each window and each channel, the percentage of power within a frequency band of 0.2 Hz centered at the local true HR was computed. The local true HR was defined as the average true HR (derived from the ECG) in the corresponding 10-second window. Figure 2 shows the PSD for a 10-second segment of the first se-

quence. In this figure, the grey area indicates the HR band (i.e. a frequency band of 0.2 Hz centered at the true HR) for which the percentage of the power was computed. The resulting power percentages were finally averaged over the duration of the sequence.
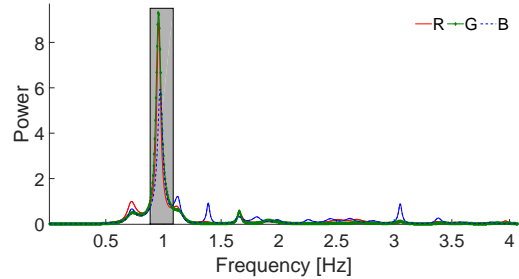

Figure 2. Parametric PSD for a 10-second segment for a forehead ROI of the first sequence (R/G/B channels). The grey area represents the true HR frequency band (mean true HR $\pm$ 0.1 Hz)

2.4. Face segmentation

In order to facilitate the interpretation of the quantitative results, some ROIs were clustered and their results were averaged. More specifically, the following four subregions were considered: the forehead, the cheeks, the whole face and the whole face except forehead and cheeks. For this purpose, a segmentation mask was used (the same for all subjects). This mask is illustrated in Figure 3.
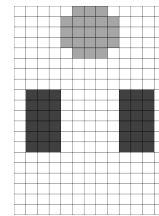

Figure 3. Segmentation mask. The grey part corresponds the forehead area, and the black parts correspond to the cheeks.

3. Results

Table 1 reports the average power percentage in the HR band for each sequence and each channel (R/G/B/NIR). These percentages are shown for the forehead, the cheeks, the whole face and the whole face except forehead and cheeks regions. These results are also represented on a boxplot (Figure 4). Figure 5 is an example to illustrate the time evolution of the power for different face regions for the first sequence. The spatial distribution of HR-related

information for the R/G/B/NIR channels is shown in Figure 6. In order to give a similar weight to every subject, the values have been normalized for each sequence, according to the largest power percentage for the green channel. In these color maps, the red regions indicate a larger power in the HR band.
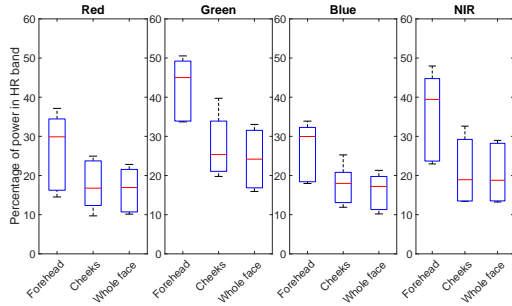

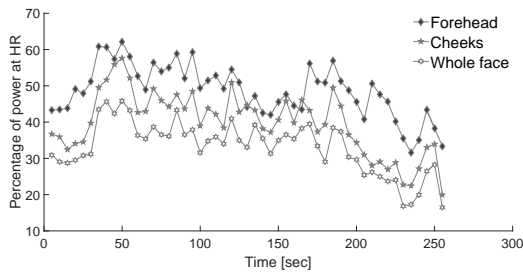Figure 4. Power percentages for face regions and channels.

Figure 5. Example of the time evolution of the power at HR frequency in the different zones for the first sequence.

4. Discussion

Figure 4 and Table 1 indicate that, for all channels, the forehead region was the one that contained the largest power at the HR frequency. This effect was observed for all sequences and for all channels. This is also confirmed by the average color maps (see Figure 6), in which the forehead part is clearly distinguishable. The color maps also indicates that cheekbones may be suitable ROIs. Interestingly, it is possible to notice that there was some power at HR on the neck (visible because the subjects were lying down). This is not really surprising as a lot of blood vessels pass through the neck. Concerning the cheeks regions, the power at HR frequency was larger than for the whole face 83% of the time. However, these differences a were not very large. On the other hand, the difference between the whole face region and the forehead region was very marked. Indeed, the whole face region contains, on average, only 59% of the forehead information. It should also
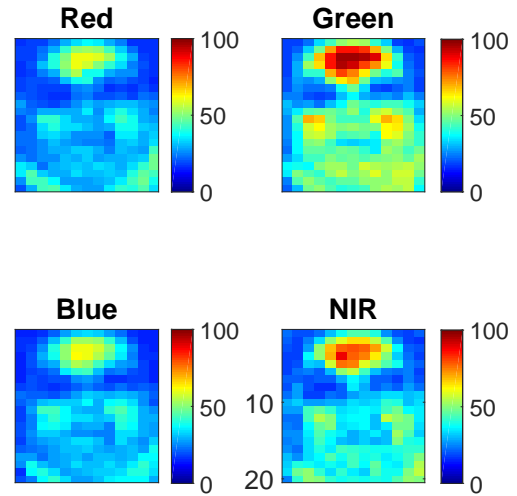

Figure 6. Normalized color maps averaged over all subjects to represent the amount of power at the true HR.

be noted from Table 1 that the absolute power percentage results are very variable between the three subjects, despite similar experimental conditions. This suggests that some characteristics such as skin color, skin thickness and blood perfusion probably play an important role in the quality of the extracted iPPG signals.

Concerning the variability between the different channels, our results show that the green channel has the strongest iPPG signals (see Figure 4). These results are in accordance with previous findings [1], and may be explained by the position of the hemoglobin absorption peaks.

Figure 5 illustrates the time evolution of the power percentage in different face regions for the first sequence. It can be noted that the three curves follow the same trends, indicating that it would be of a dubious interest to use the different regions in a complementary way.

It should be mentioned that the main limitation of this preliminary study is the size of the dataset. Still, the same tendencies were observed for all sequences and all subjects, reinforcing the hypothesis that the results are reproducible. The obtained results call into question the choice of the ROI. Indeed, most often the whole face is used [2, 3], but better performance might be achieved using only the forehead. A previous study pointed out considerations about the ROI size and its impact on the signal-to-noise ratio [1]. Our study highlighted the fact that, in addition to the size, the location of the ROI also matters. Indeed, the selected ROIs of equal sizes were behaving differently, even the ones containing only skin.

Table 1. Average percentage of power in the HR band for the different regions. (a) forehead; (b) cheeks; (c) the whole face; (d) the whole face except the forehead and the cheeks

Sequence #	Red				Green				Blue				NIR			
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
subject1 { 1	32.5	24.9	22.8	21.5	48.7	39.7	33.0	30.2	32.3	25.3	21.3	19.5	44.2	32.6	28.2	25.8
subject1 { 2	37.1	23.7	21.6	19.6	50.5	33.9	31.5	29.2	31.7	20.8	18.9	17.2	48.0	29.2	29.0	27.0
subject2 { 3	14.5	9.7	10.7	10.5	33.7	19.8	16.9	14.6	18.0	11.9	11.3	10.5	23.7	13.4	13.2	12.1
subject2 { 4	16.2	12.3	10.1	9.1	33.9	21.5	15.9	13.1	18.4	13.1	10.2	8.8	23.0	15.6	13.5	12.2
subject3 { 5	27.3	13.1	14.4	13.3	41.3	21.1	20.9	18.8	28.2	15.3	15.5	14.3	34.6	13.5	15.6	14.0
subject3 { 6	34.4	20.4	19.5	17.7	49.2	29.2	27.4	24.8	33.9	20.7	19.8	18.1	44.7	22.3	22.0	19.5
Av.	**27.0**	**17.4**	**16.5**	**15.3**	**42.9**	**27.5**	**24.3**	**21.8**	**27.1**	**17.8**	**16.2**	**14.7**	**36.4**	**21.1**	**20.3**	**18.4**

5.   Conclusions

These preliminary results suggest that the forehead part is the most appropriate region to estimate HR, followed by the cheeks. These regions should be tracked in priority in iPPG applications. Further work should be performed to validate these conclusions on a broader population, including subjects with dark skin and for different subject postures.

Acknowledgements

References

[1] Verkruysse W, Svaasand LO, Nelson JS. Remote plethysmographic imaging using ambient light. Opt Express 2008; 16:21434–21445.

[2] Poh MZ, McDuff D, Picard R. Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. IEEE Trans Biomed Eng 2011;58:7–11.

[3] Holton BD, Mannapperuma K, Lesniewski PJ, Thomas JC. Signal recovery in imaging photoplethysmography. Physiol Meas 2013;34:1499.

[4] McDuff D, Gontarek S, Picard R. Improvements in Remote Cardiopulmonary Measurement Using a Five Band Digital Camera. IEEE Trans Biomed Eng 2014;61:2593–2601.

[5] Lewandowska M, Ruminski J, Kocejko T, Nowak J. Measuring pulse rate with a webcam - A non-contact method for evaluating cardiac activity. In 2011 Federated Conference on Computer Science and Information Systems (FedCSIS). 2011; 405–410.

Address for correspondence:

Sibylle Fallet
EPFL SCI STI JMV - ELD 224 - Station 11
1015 Lausanne, Switzerland
sibylle.fallet@epfl.ch