UDC 611.1; 519.25

HUMAN HEART RATE MONITORING BASED ON FACIAL VIDEO PROCESSING

N.V. BACH, I.A. BORISKIEVIC

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received February 23, 2023

Abstract. Heart rate (HR) is one of the most important physiological parameters and a vital indicator of people's physiological state, making it important to monitor. Over the last decade, research has focused on non-contact systems, which are simple, low-cost, and comfortable to use. This paper analyses the complexity of each step in the development of a human heart rate monitoring algorithm based on facial video processing. Specifically, the research focuses on the pulse signal extraction step. The proposed algorithm based on the transform of 2D signal to 1D signal, its detrending and window discrete transform are used to improve the accuracy of HR estimation. The experimental results show that the accuracy of human heart rate estimation in terms of MAE and RMSE is equal around 2 bpm.

Keywords: heart rate, facial video processing, remote photoplethysmography, window discrete transform.

Introduction

The human pulse is a rhythmic oscillation of the vessels that correspond to the contractions of the heart and it is one of the most important indicators that helps to track whether everything is good with the heart. Traditional heart rate detection mainly includes two ways: electrocardiograph (ECG) and contact photoplethysmography (cPPG) based on sensors. Due to the limitations of cPPG methods, it is particularly important to study a non-contact HR detection method. The rPPG (remote photoplethysmography) has been proven to be superior because it is non-intrusive. It may be suitable for continuous measurement of heart rate (HR) in many cases, such as neonatal ICU (intensive care unit) monitoring [1], burn victims, driver status assessment [2], online learning [3], provide low-cost solutions for health monitoring applications, another application of rPPG in health monitoring include blood perfusion mapping [4] and monitoring regional anesthesia effectiveness [5]. Although this kind of methods may not be as accurate as an electrocardiogram (ECG) device, they can provide a long-term HR monitoring without being uncomfortable for patients. These technologies can be very helpful in increasing fields like telemedicine, where usability is a key factor.

The choice of a facial ROI (region of interest) acts as the first key step of the system. First, the pulsatile signal strength varies at different locations on the face due to the distribution of capillaries beneath the skin surface. The location of an ROI has a direct impact on the quality of the raw rPPG measurement. Second, the shape of an ROI always leads to unnecessary inclusion of undesired pixels like eyes, mouth, hair, or background pixels, thus, introducing rigid/non-rigid motion artifacts. It is crucial to choose a good ROI to guarantee a higher measurement accuracy. While most rPPG approaches extract pulsed signals by averaging over all skin pixels on ROIs, we propose an algorithm that allows extracting pulse signals from ROIs for increasing HR estimation accuracy.

Our contributions can be summarized as follows:

– we present a human heart rate monitoring algorithm based on facial video processing;

– we propose an algorithm to remove outliers from signal based on Z-score method.

Human heart rate monitoring based on facial video processing algorithm

Proposed method uses video as an input and returns pulse rate as an output. Sequence of steps of the proposed algorithm can be represented as follows:

– detection of person's face on color image;
– extraction region of interest (ROI);
– transform 2D ROI to 1D ROI signal;
– processing of 1D ROI signal;
– computation of power spectrum of 1D ROI signal;
– band pass filtering;
– estimation of person heart rate.

Block diagram of the proposed algorithm to determine HR from a facial video can be represented by the graphical form.

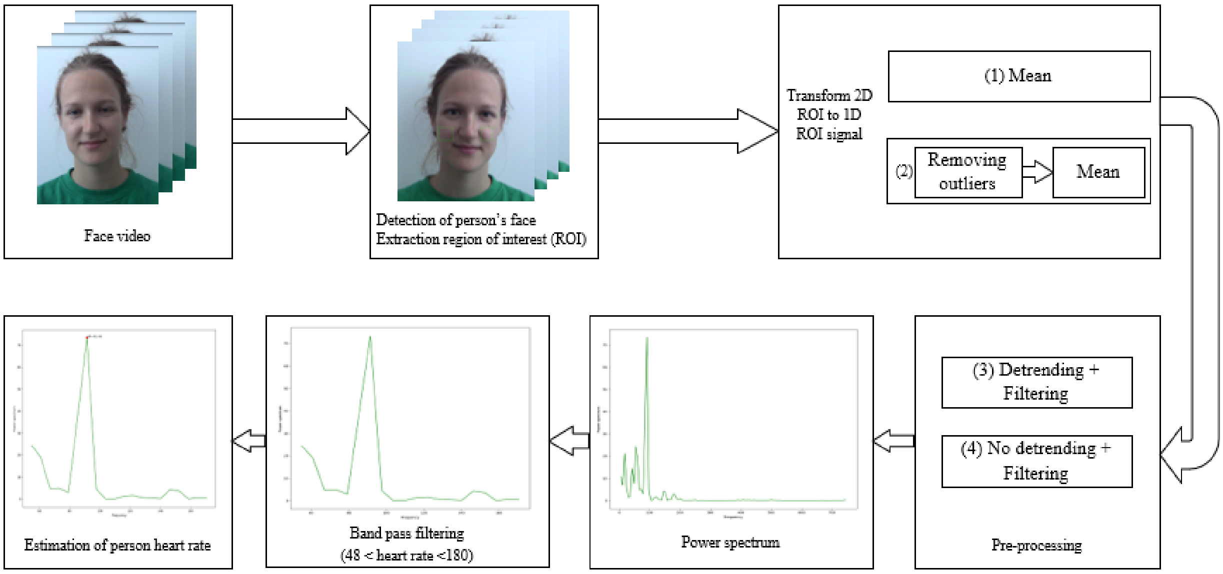

Figure 1.The sequence of steps of the proposed algorithm for estimating heart rate using facial video

Face detection is a crucial preprocessing step for the traditional rPPG methods to measure HR. Its accuracy has a direct impact on the accuracy of HR detection. At present, three mainstream methods Viola–Jones (VJ), histogram of oriented gradients (HOG) and multi-task cascaded convolutional networks (MTCNN) are often used for face detection. Dlib (an open-source library) face detector model is based on HOG feature descriptor and linear SVM (support vector machine)) classification. Dlib's HOG + linear SVM face detector is fast and efficient. Dlib HOG is the fastest method on the CPU (central processing unit) [6]. In this application we use the HOG method to locate a person face.

Extraction region of interest. The ROI in the cheeks was determined through the location of the human mouth and nose. The human eyes, mouth and nose were detected with Dlib HOG method.

Transform 2D ROIs to a 1D signal

For each frame of the video sequence, we obtain ROIs of facial skin pixels. The amplitude of the PPG-signal in light reflected from the skin varies as a function of the wavelength, showing a strong peak around 550 nm and a dip around 650 nm. Because that we use only wavelength 550 nm (green channel) to estimate HR [7].

To transform 2D ROIs to a 1D signal most rPPG approaches extract pulsed signals by averaging over all skin pixels value on ROIs. We propose algorithm to remove outliers from video frame ROI histogram based on Z-score method before extract pulsed signals.

Removal of outliers from ROI histogram based on Z-score method

The Z-score is one of the most commonly used tools in determining outliers. Z-score (Z_{score}) is just the number of standard deviations away from the mean that a certain data point is.

$$Z_{score} = \frac{I_{ROI} - \mu}{\sigma}, \tag{1}$$

where I_{ROI} – matrix of ROI pixels values, $\mu = \sum_{x=0, y=0}^{x<M, y<N} \frac{I_{ROI}(x,y)}{a \cdot b}$ and $\sigma = \sqrt{\frac{\sum_{x=0, y=0}^{x<M, y<N} (I_{ROI}(x,y) - \mu)^2}{a \cdot b}}$ – the mean value and the standard deviation of the ROI image values respectively, (a, b)– the width and height of video frame ROI.

To improve HR estimation accuracy expression (2) is used to detect and remove outliers from matrix of ROI pixels I_{ROI}:

$$\begin{cases} I_{ROI\,OUT} > \mu + Z_{scre} \cdot \sigma \;\; if \;\; Z_{score} < 2 \\ I_{ROI\,OUT} < \mu - Z_{scre} \cdot \sigma \;\; if \;\; Z_{score} > -2 \end{cases} \tag{2}$$

The 1D ROI signal Detrending

Detrending is an important signal processing concept, which is used to remove unwanted trend from the 1D ROI signal that represents as a sequence *s(n)* of discrete mean values of video frame ROIs. We eliminate the signal deviation trend using the adaptive iteratively re-weighted penalized least squares (Airpls) [8].

The window discrete Fourier transform

Before applying discrete Fourier transform (DFT), the PPG-signal that represents as a sequence of discrete values of mean value of video frame ROI is filtered by Hamming window $w(n) = 0,54 - 0,46\cos(\frac{2\pi n}{N-1})$, $0 \le n \le N-1$, N – the window length, n – index of discrete time samples.

The time signal of the pulse wave window is transformed into a sequence of discrete frequency samples by DFT, which is defined by:

$$S_k = \sum_{n=0}^{N-1} s_{FIL}(n) \cdot e^{-i\frac{2\pi kn}{N}}, \tag{3}$$

where S_k – k-th the DFT coefficients, $s_{FIL}(n)$ – n-th filtered value of 1D ROI signal *s(n)*, *N* – the number of the DFT coefficients, k ($k = 0,1,\ldots,N-1$) – frequency index.

The DFT power spectrum of 1D signal is defined as

$$P_k = \begin{cases} \frac{1}{N^2}\left|S_0\right|^2, \quad k = 0 \\ \frac{2}{N^2}\left|S_k\right|^2, \quad k = 1,\ldots,N/2, \end{cases} \tag{4}$$

An DFT was applied to the filtered pulse signal, and the heart rate was taken as the frequency where the spectral power was maximal. Then we apply band pass filter with F_l = 0.8 Hz and F_h = 3 Hz, which are 48 and 180 bpm respectively to remove unwanted frequency.

Human heart rate estimation

Human heart rate is calculated per window. The window length is set to 10s and the time distance between the two consecutive frames is equal to 0.04s.

There are usually several peaks in a same frequency domain of power spectrum. Heart rate value in the PPG signal is the position of frequency sample with the highest energy.

Experimental result

We use dataset from https://github.com/vladostan/Dataset-for-video-based-pulse-detection. Open dataset for video-based pulse detection. Includes 30 .mp4 video files and ground truth ECG signals.

Mean absolute error (MAE) and root mean square error (RMSE) [9] is selected to evaluate HR estimation accuracy by the proposed algorithm. When calculating the average error, outliers are discarded. The HR estimation accuracy based on the algorithms depicted in Figure 1, are presented in the table below.

Table 1. **Average heart rate prediction: comparison among different algorithm on different conditions with detrending (3)**

Algorithm	Normal condition			Physical activity		
	ME (bpm)	MAE (bpm)	RMSE (bpm)	ME (bpm)	MAE (bpm)	RMSE (bpm)
(1) Mean	0.21	1.83	2.16	0.75	2.04	2.61
(2) Removing outlier + mean	0.27	1.74	2.01	0.56	1.76	2.21

Table 2. **Average heart rate prediction: comparison among different algorithm on different conditions without detrending (4)**

Algorithm	Normal condition			Physical activity		
	ME (bpm)	MAE (bpm)	RMSE (bpm)	ME (bpm)	MAE (bpm)	RMSE (bpm)
(1) Mean	0.1	1.82	2.2	0.64	2.01	2.55
(2) Removing outlier + mean	0.14	1.75	2.11	0.43	1.87	2.45

It can be concluded from Tables 1 and 2 that the variant of algorithm (2), which removes outliers and applies averaging of 1D signal under different conditions with detrending, provides lower MAE and RMSE value upon estimating HR. The experimental results show that the accuracy of human heart rate estimation in terms of MAE and RMSE is equal around 2bpm due to the proposed algorithms uses outlier removing procedure.

Conclusion

The proposed algorithm is based on transform 2D signal to 1D signal, removing outliers, detrending and window discrete Fourier transform. It allows us to increase the human heart rate estimation accuracy due to removing outliers from video frame ROI based on Z-score method is used. The experimental results show that the accuracy of human heart rate estimation in terms of MAE and RMSE is equal around 2bpm.

References

1. Aarts L.A.M., Jeanne V., Cleary J.P., Lieber C., Nelson J.S. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit – a pilot study // Early Hum Dev. 2013 Dec. Vol. 89, No. 12, P. 943-8.
2. Zheng K., Ci K., Cui J., Kong J., Zhou J. Non-contact heart rate detection when face information is missing during online learning // Sensors. 2020. Vol. 20, No. 24, P. 7021.
3. Taylor W., Abbasi Q.H., Dashtipour K., Ansari S., Shah S.A. A review of the state of the art in non-contact sensing for covid-19 // Sensors. 2020. Vol. 20, No. 19, P. 5665.
4. Rubins U., Erts R., and Nikiforovs V. The blood perfusion mapping in the human skin by photoplethysmography imaging // XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010. IFMBE Proceedings. 2010. Vol. 29, P. 304-306.
5. Rubins U., Miscuks A., Rubenis O., Erts R., and Grabovskis A. The analysis of blood flow changes under local anesthetic input using non-contact technique // 2010 3rd International Conference on Biomedical Engineering and Informatics, Yantai, China. 2010. P. 601-604.
6. Gupta V. Face Detection – Dlib, OpenCV, and Deep Learning (C++ / Python). October 2018. [Online]. Available: https://learnopencv.com/face-detection-opencv-dlib-and-deep-learning-c-python/. [Accessed February 2023].
7. De H. G., Jeanne V. Robust Pulse Rate From Chrominance-Based Rppg // IEEE Transactions on Biomedical Engineering. 2013. Vol. 60, No. 10, P. 2878 – 2886.
8. Zhang Z., Chen S., Liang Y. Baseline correction using adaptive iteratively reweighted penalized least squares // Analyst. 2010 May. Vol. 135, No. 5, P. 1138-46.
9. Pagano T.P., Santos L.L., Santos V.R. Remote Heart Rate Prediction in Virtual Reality Head-Mounted Displays Using Machine Learning Techniques // Sensors. 2022. Vol. 22, P. 9486.