

Regions of interest selection in the tasks of contactless human pulse measurement by analyzing the RGB video stream

Alexander V. Fisunov^{1,*}, Maxim S. Beloyvanov¹, and Issam Samara²

¹Don State Technical University, 346500 Shakhty, Russia

²Higher Institute for Applied Sciences and Technology (HIAST), Damascus, Syria

Abstract. This paper is devoted to improving the accuracy of human pulse measurement by RGB video stream analysis. For this purpose, a study was conducted the influence of the size, location and stability of the region of interest on contactless human pulse measurement results.

1 Introduction

The great interest are the studies of the mental and physical state of a person by contactless methods using digital cameras. These methods can be useful for developing person's condition monitoring system, such as candidate assessment for interviews, vehicle driver monitoring, and others. One of the most important estimated parameter is the heart rate.

Many works describe methods of contactless measurement of the human pulse of the video sequence. They are based on the principle of optical absorption of light by human skin and blood. Human skin contains a large number of small blood vessels, which are filled with blood with each heart cycle, thereby causing a change in the intensity of light reflected from the skin. These changes are not visible to the naked eye, but they can be fixed with ordinary digital cameras. Information about heart rate is contained in the photoplethysmographic signal obtained from the videostream of a person's face. The quality and accuracy of the signal strongly depends on the choice of the area of the face used for the analysis.

2 The statement of the problem

In [1], Muhammad Waqar et al. use Face Tracker [7] Saragih to track faces on video. Areas of the cheeks are selected for analysis since they are less likely to be hidden by hair, beard, or glasses.

According to work [2], Niuesong Niu et al. signal quality can be improved by selecting correct areas of interest. As areas of interest are selected a rectangular area in the center of face, lower area of face, and cheeks. For the detection these areas, an open source detector is used [3] which allows localizing 81 specific points on the face. After obtaining the region of interest, non-skin pixels are deleted.

* Corresponding author: fisunov.a.v@gmail.com

In [4], the 3D Landmark tracker [5] is used to track faces. It is able to identify 7 key points on the face image, namely: the outer and inner corner of the eye, nose, left and right corner of the mouth. Found key points are used to highlight the area above the mouth in order to eliminate the influence of noise caused by lip movement.

The above analysis showed that these methods of extracting areas of interest can partially eliminate the minor movements of participants in the frame. Since the change in light intensity caused by the reflectivity of the skin has a very small amplitude, the distortions caused by movement make it impossible to extract information about the heart rate from the photoplethysmographic signal.

The purpose of this work is to study the influence of the choice of location of areas of interest and their size on the results of heart rate measurement in a non-contact way.

3 The obtained results

For the analysis purposes, several video files with a resolution of 1280x720 pixels and a frame rate of 16 fps were recorded. During recording, the participant sat motionless for two minutes, and possible changes in lighting caused by cloudiness or flickering of the monitor were also excluded. For the analysis, the forehead area proposed in [6] was chosen.

To represent the photoplethysmic signal, it takes place by extracting the average value of the green channel in the area of interest and the resulting values are recorded in a buffer to represent the time signal, and a Butterworth filter with a cutoff frequency of [0.7-2.5] Hz is applied to the signal, which corresponds to the person normal heart rate [42-150]. After that, frequency analysis is applied to extract the dominant frequency in the signal.

Figure 1 shows the results of measuring the pulse with a region of interest on the forehead, with the signal length being selected 1024 counts. The participant's pulse was a priori about 74 beats per minute. When detecting a face on the frames with a Haar cascade, a rectangle of approximately 350x350 pixels was obtained.

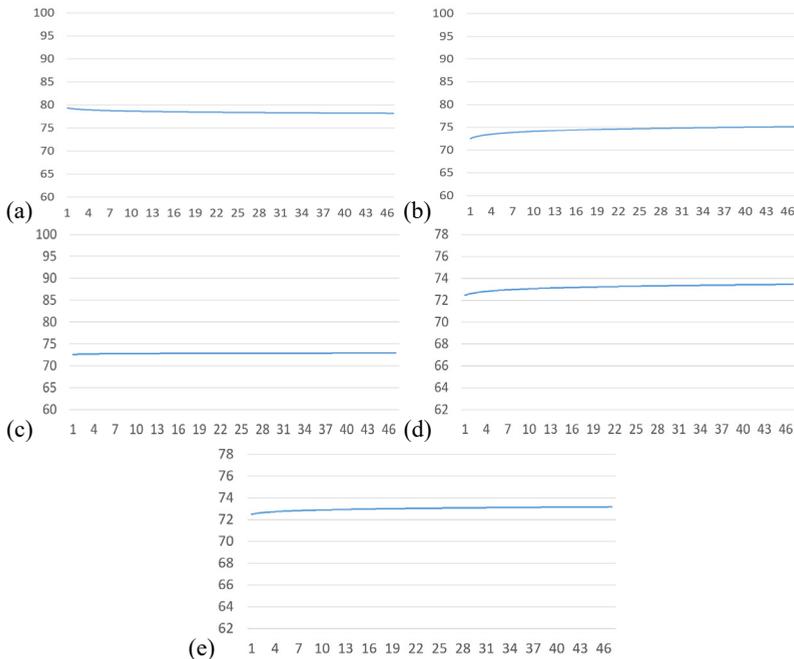


Fig. 1. Heart rate measurement results with different sizes of the selected area (a) 10x10 pixels (b) 20x20 pixels (c) 50x50 pixels (d) 80x80 pixels (e) 140x50 pixels

Figure 1 (a) shows a heart rate measurement graph using a 10x10 pixel area from which it can be seen that the readings differ on average from those previously received at 5-7 beats per minute. Pulse measurements for the 20x20 and 50x50 areas shown in Figure 1 (b) and 1 (c) show an average difference of 3 beats per minute, but the most accurate results are shown in the pulse measurement charts in Figure 1 (d) and 1 (e) for areas of 80x50 and 140x50 pixels, respectively. From this we can conclude that select the areas of interest of small sizes are not a reliable way to extract the signal.

The next stage was a study of the influence of the position of the region of interest on the measurement of human pulse. For this purpose, a rectangle of 80x50 pixels was chosen, which was fixed on the left side of the forehead and pulse were estimated. Then the position of this rectangle was increased by 15 pixels in X axis and the measurements were performed again. Figure 2 (a-e) shows the human pulse values when the region of interest is shifted by 0, 15, 30, 45 and 60 pixels relative to the initial position.

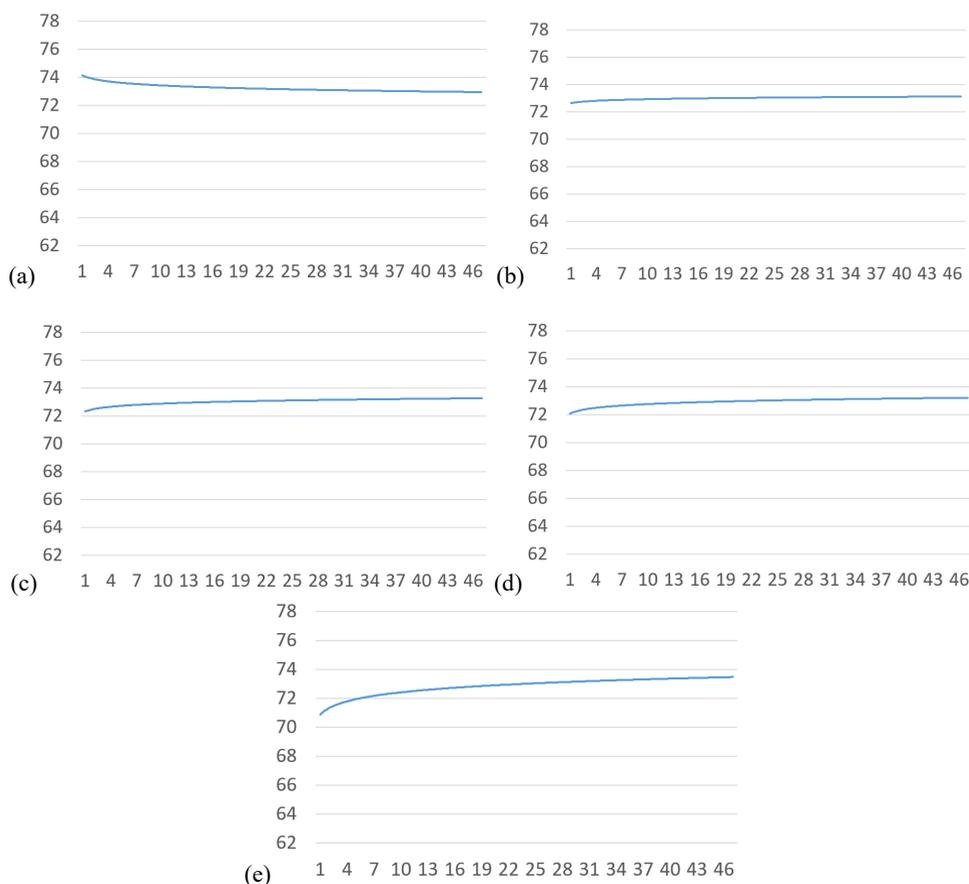


Fig. 2. Pulse measurement results with different ROI position shifting.

The Figure 2 show that the initial displacement of the region of interest does not affect the measurement accuracy. Binding the coordinates of the region of interest to particular points on the face can lead to a «shaking» of ROI coordinates from frame to frame. An experiment was conducted in which the region of interest was subject to an offset of 3 pixels in arbitrary directions in each frame. The interest is due to the possibility of using various face landmarks trackers and their influence on the measurement results. The 10x10 area was excluded from consideration, since previous experience shows the inexpediency of using the

area of such size. Figure 3 shows the results of measurements of the pulse on the forehead when the region of interest is shifted by 1-3 pixels against initial position at random direction in each frame.

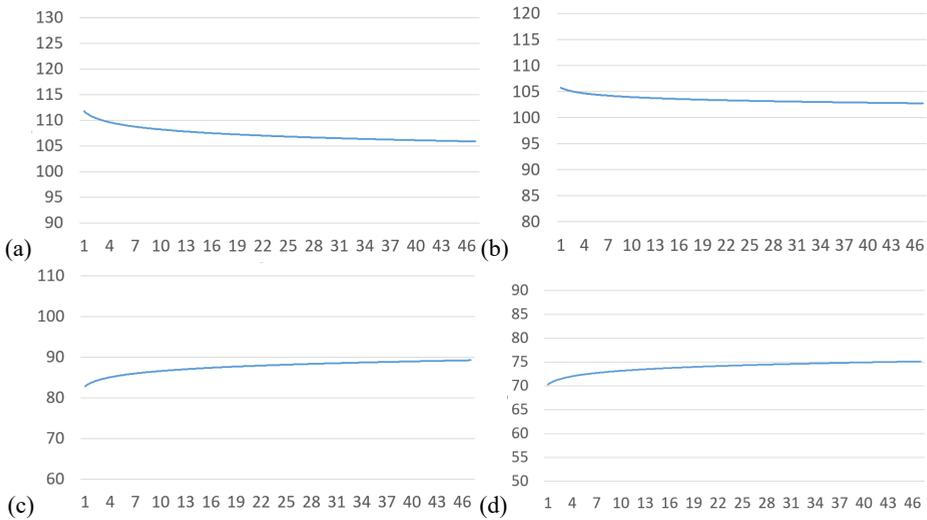


Fig. 3. Pulse measurement results with coordinate shaking and different sizes of the selected area (a) 20x20, (b) 50x50, (c) 80x50, and (d) 140x50 pixels.

From Fig. 3 (a) and 3 (b) it can be seen that the measurement results obtained using the 20x20 and 50x50 regions are most susceptible to distortion. The Figure 3 (c) shows that when using the 80x50 area, the measurement error decreases, but the results are not reliable, since the deviation from a priori known pulse value averages more than 10 beats per minute. The graph shown in Figure 3 (d) shows that the results of pulse measurements are less prone to distortion when using large areas region of interest.

In addition to the forehead, often in the work areas of the cheeks are selected. Pulse measurements were made in the forehead, left and right cheek areas, the results are shown in the figure 4.

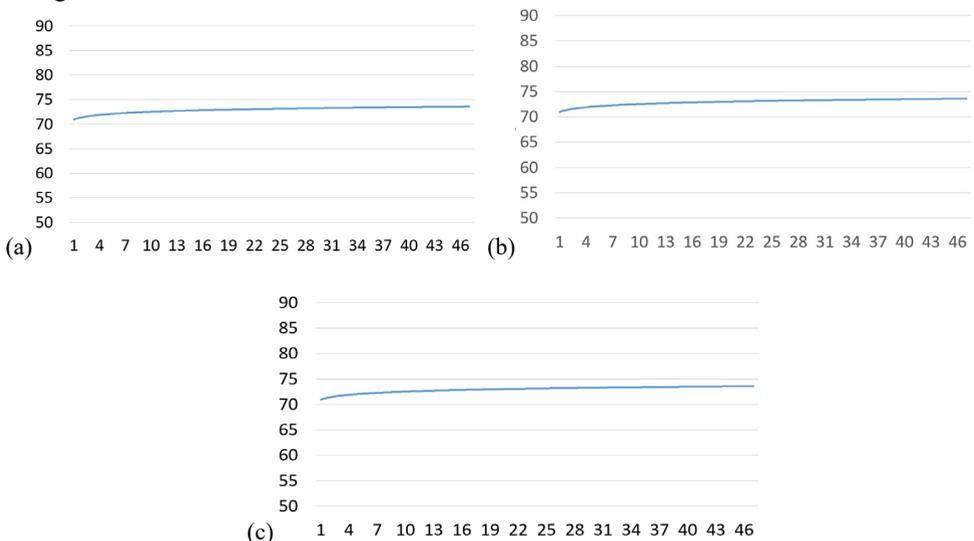


Fig. 4. Pulse measurement results using areas of (a) the forehead, (b) the left and (c) the right cheek.

The graphs shown in Fig. 4 show that the results of pulse measurements with in the cheek area do not differ from the results obtained from the forehead area. This suggests that the cheek areas are also reliable areas for extracting the photoplethasmographic signal and analyzing the heart rate.

5 Conclusions

The study of the influence of the size of the region of interest on the results of pulse measurements showed that when choosing a region of small size, the results are subject to strong distortion and their use is not advisable.

Of all the areas in question, the least distorted subject to the random offset introduced by the area of interest is the largest area of 140x50 pixels. This suggests that with the use of face trackers, it is necessary to select the largest area, since it is least likely to be susceptible to position distortions introduced by the tracker.

When comparing the results of measurements of the pulse on the forehead and cheeks, rendering have minor differences and lie within the 1st stroke per minute, it is therefore possible to use different areas at the same time to confirm the correctness of the algorithm, as well as to use each of the areas inability of one of them distortions introduced by ambient lighting or the subject's movement in the frame.

References

1. M. Waqar, B. Tiddeman, Proc. 7th UK Br. Mach. Vis. Work, 7, 1 (2015)
2. X. Niu, H. Han, S. Shan, X. Chen, In 2017 IEEE International Joint Conference on Biometrics (IJCB) IEEE, 642 (2017)
3. J. Zhang, S. Shan, M. Kan, and X. Chen. In European conference on computer vision Springer, 1 (2014)
4. Heart Beat Rate Estimation from Video Pavel Cerny CTU-CMP-2014-07 (2014)
5. J. Cech, V. Franc, J. Matas, A In 2014 22nd International Conference on Pattern Recognition, 2173 (2014)
6. M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak. n 2011 federated conference on computer science and information systems (FedCSIS), 405 (2011)
7. S. Jason FaceTracker Available: <https://github.com/kylemcdonald/FaceTracker> (2015)