



Remote Photoplethysmography (rPPG) for Contactless Heart Rate Monitoring Using a Single Monochrome and Color Camera

Xiaocong Ma¹(✉) , Diana P. Tobón² , and Abdulmotaleb El Saddik¹ 

¹ Multimedia Communications Research Laboratory, University of Ottawa,
Ottawa, Canada

{xma064,elsaddik}@uottawa.ca

² Universidad de Medellín, Medellín, Colombia

dptvallejo@gmail.com

<http://www.mcrlab.net>

Abstract. Human vital signs are essential information that are closely related to both physical cardiac assessments and psychological emotion studies. One of the most important data is the heart rate, which is closely connected to the clinical state of the human body. Modern image processing technologies, such as Remote Photoplethysmography (rPPG), have enabled us to collect and extract the heart rate data from the body by just using an optical sensor and not making any physical contact. In this paper, we propose a real-time camera-based heart rate detector system using computer vision and signal processing techniques. The software of the system is designed to be compatible with both an ordinary built-in color webcam and an industry grade grayscale camera. In addition, we conduct an analysis based on the experimental results collected from a combination of test subjects varying in genders, races, and ages, followed by a quick performance comparison between the color webcam and an industry grayscale camera. The final calculations on percentage error have shown interesting results as the built-in color webcam with the digital spatial filter and the grayscale camera with optical filter achieved relatively similar accuracy under both still and exercising conditions. However, the correlation calculations, on the other hand, have shown that compared to the webcam, the industry grade camera is superior in stability when facial artifacts are presented.

Keywords: Computer vision · Heart rate · Photoplethysmography · rPPG · Signal processing

1 Introduction and Related Works

Human heart rate and heart rate variability are the crucial parameters corresponding to the functions of the heart. The speed and volume of the blood pulse

can explicitly indicate various physical conditions of one's body such as emotions, cardio activity levels, stress, fatigue, and heart diseases [1], and thus these parameters are usually measured for quick clinical diagnostics in the first place when necessary. On the other hand, long-term heart rate monitoring is undertaken when abnormal symptoms such as palpitation and extra systole need to be in check on a regular basis. Traditional wearable heart rate monitors, such as FitbitTM and smart watches can measure heart rate and give accurate results, but they are usually dedicated to just one user and need to be placed close against the skin of the user. Impressively, the rPPG technology has taken another approach; by examining the intensity change of a reflective light caused by the change of blood flow on a person's face and applying adequate computer vision and signal processing techniques, determining an approximate value of the heart rate from a distance using a video camera system has been made possible. Due to the nature of video signals, the rPPG method can acquire multi-model vital signs including heart rate, respiratory rate, and facial expression both in real-time and offline. Therefore, it is ideal for the rPPG to become a cost-efficient and user-friendly solution in real-world applications.

Academically, computer vision-based methodologies for front face remote heart rate measurement have already become popular in recent researches. C. Wang in his paper [2] has conducted a survey on multiple rPPG methods that can be classified into either intensity-based methods, which focus on facial light reluctance [3], or motion-based methods, which focus on head movements [4], and he concluded that intensity-based methods are still much more effective in terms of speed and accuracy. In [3,5], the authors have compared Signal to Noise Ratios (SNRs) of the blood pulse in the Red-Green-Blue (RGB) color space model and have shown that the green channel is the best pick for rPPG heart rate detection. In [6], an rPPG heart rate detector on the iOS platform using offline videos had been developed and sufficient usable results were obtained; however, the author did not mention its real-time performance.

It is also noteworthy that rPPG methods can be tailored at various stages of the entire process, such as pre-processing, signal extraction, and post-processing [2]. At the pre-processing stage, Po and his colleagues [7] implemented their system with an adaptive Region of Interest (ROI) selection method based on the detected signal qualities and concluded with improved accuracy at a cost of computational expense in real-time. In [8,9], Rahman's team proposed real-time rPPG systems using Independent Component Analysis (ICA) to combat motion artifacts, but in [10], Demirezen claimed that their work with the nonlinear mode decomposition method achieved better results than ICA. In terms of optical modeling, Sanyal in [11] took another approach by using hue parameters from the Hue-Saturation-Value (HSV) color space model instead of green in the RGB before applying ICA and also summarized a higher accuracy in the performance outcomes. In the post-processing stage, the time domain peak-detection or frequency domain algorithms were the most commonly applied methods used in the past [2], but machine learning and modeling techniques are trending among most recent research. The types of supervised learning methods can include

kNN-based modeling [12], cNN-based modeling [13,14], spatial-temporal modeling [15], adaptive neural network model selection [16], etc. However, obtaining a dataset with appropriate ground truth can be a crucial factor for training accurate models, and real-time performance reduction needs to be addressed considering the complexity of the model trained [2].

The structure of this paper is as follows. Section 2 describes the used methods and materials in this study, Sect. 2.10 presents detailed steps to obtain the final data, Sect. 3 conducts an analysis on the final results, and Sect. 4 draws an open conclusion for future work.

2 Materials and Methods

This work has been implemented as a hybrid system in which either a regular built-in color webcam or a FLIRTM industrial grade grayscale camera can be utilized as its image input sensor based on the detection of the connected hardware. The captured facial image frames were converted to grayscale for further signal extraction and heart rate detection in later stages, and the heart rate results were updated on the screen rapidly in real-time. In addition, a BTChoicTM blood oxygen and dynamic heart rate bracelet was used as a skin attached device for providing the ground truth to our results. Subsection 2.1 provides details regarding our hardware environment, subsection 2.2 describes how the data was collected for our work, subsection 2.3 explains the selection of color space and channel input, subsection 2.4 gives a quick overview of the highlighted signal processing techniques used in our implementation, and subsections 2.5 to 2.10 go through more details on the multiple stages of the entire signal processing process.

2.1 Hardware Setup

The implementation and tests of this project were performed on a PC with an IntelTM i5-8250U processor under MicrosoftTM Windows 10 operating system and Python 3.5 environment. The sensors used for image and data collections were the BTChoicTM skin contact smart wear bracelet, FLIRTM Blackfly S BW industrial camera with RainbowTM H3.5 mm 1:1.6 fixed lens and 500–555 nm light green band-pass filter, and a laptop with a built-in AsusTM USB2.0 HD webcam.

2.2 Data Collection Setup

Data tests were conducted by a group of 20 volunteers aged between 25–40 years with mixed races, skin tones, genders, and various amount of facial hair. Each person was asked to wear the BTChoicTM bracelet to obtain the ground truth heart rate, and was then told to sit still and breathe normally in front of the camera at about 0.5 meters away. Three camera hardware setups were used in our rPPG data collections: a FLIRTM industrial grayscale camera with an optical green filter; a FLIRTM industrial grayscale camera without an optical

filter; an AsusTM HD built-in webcam by itself. Approximately 30 s of data reading was then performed on both the camera and the BTChoicTM bracelet a under controlled light environment in the lab as shown in Fig. 1. The output display seemed to be more stabilized over time, thus we took a visually averaged output as our test result approximately 20 s after the program started. The test subject was then asked to perform a light exercise such as jogging or push-ups for approximately 20 s, and then immediately sit back in front of the camera to take another set of readings.

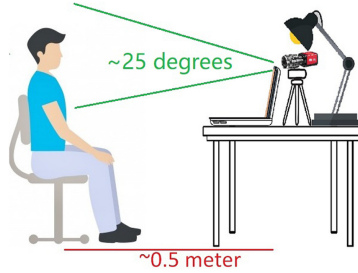


Fig. 1. Data collection setup.

2.3 Selection of Color Space and Channel

Our original plan was to initially implement and pass raw images, a.k.a. data in all RGB channels through an ICA stage for selecting the best signal, and then pass the resulting independent components through the follow-up stages. According to our previous researches, the signals related to the absorption bands for oxy- and deoxyhemoglobin on the facial skin have the strongest signal-to-noise ratio in the yellow and green light color spectrum [2], and even though many other methods have been explored, the green color data in the RGB color space remains to be the most popularly used channel for extracting HR [17]. Our implementation using FastICA agreed with the research results by showing that the projection highly weighted towards the green channel, with fewer contributions from red, and almost none from blue. Thus, with this less significant improvement on accuracy, we decided to reduce the computation expense for better performance in real-time by replacing ICA and applying only the green channel as the selected input in our implementation.

The FLIRTM grayscale camera filters out non-green color spectrum using the equipped optical filter in the analog domain, while the built-in color webcam produces all color channels, from which the green channel was selected in the digital domain. Thus, it was necessary to perform a comparative analysis between these optical and digital filters later in this paper.

2.4 Applied Key Techniques

To perform feature extraction together with signal processing in real-time, the overall system needs to be optimized so that the image data collection performance will satisfy the demand of correct heart rate determination within the reasonable frequency spectrum. According to the Nyquist Theory, to correctly determine a 3Hz or 180 beats per minute (bpm) heart rate, a video frame rate of at least 6 frames per seconds (fps) is needed. To increase the accuracy and efficiency of the detection, we had utilized OpenCV face detection together with the Dlib facial landmark prediction engine for a fast and accurate facial ROI image extraction. Next, the facial alignment, value outlier correction, and Gaussian average filter were applied at pre-processing to combat lens distortion and motion movement artifact. Once a signal data was extracted, it was then fed through a multi-stage conditioning and shaping process. Finally, temporal filtering, Fast Fourier Transform, and power density selection techniques were used to accelerate calculation and produce end results from the frequency domain, as described in Fig. 2.

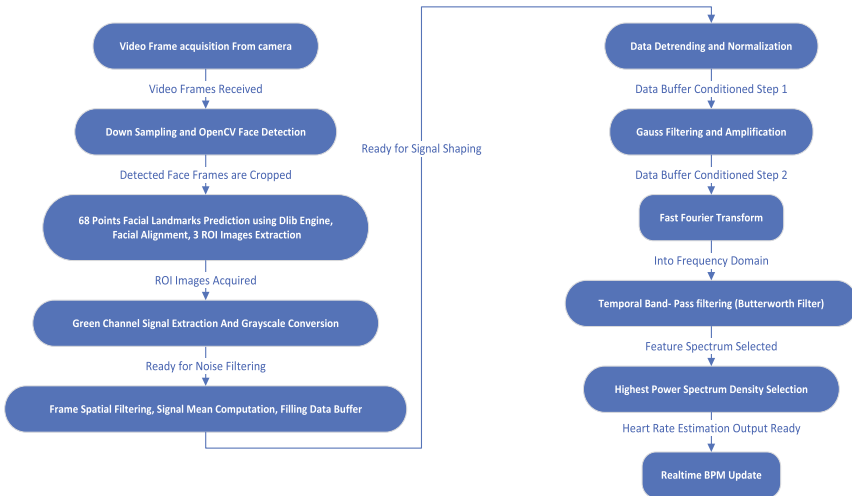


Fig. 2. Signal processing flow chart diagram for heart rate calculation.

2.5 Feature Extraction

The key feature we aimed to extract in our proposed method was the time-series data obtained from spatially averaged grayscale values in multiple ROIs that could represent the slight intensity changes of reluctant light caused by facial blood volume changes, particularly in the green color spectrum.

2.6 Face Detection and Facial Landmark Prediction

The OpenCV library is widely accepted for its fast face detection using Haar Cascade classifier [18]. We have combined this classifier together with a Dlib pre-trained 68-points facial landmark prediction engine to efficiently detect facial bounding boxes and to also crop out selected ROI regions from raw video frames. To speed up the face detection process, the image was downsized to only a quarter of its original size. Once the position of the face box was determined, a face image frame was cropped from the original video frame to retain high resolution, and it was then resized to a 256 by 256 pixels matrix to equalize the size of the input data. Next, to correct the face pose for the landmarks' prediction and minimize the artifacts caused by body movement, a facial alignment process using Imutils library [19] was followed to straighten up the face if the detected face pose was tilted or rotated.

2.7 ROI Selection and Data Collection

To find rich blood vessels, uniformly distributed skin tone, and minimal facial expression movements on the face, three ROIs were selected using rectangular boxes for achieving better signal to noise ratio (SNR) at the following locations: two on the cheeks from each side and one on top of the forehead just above the eyebrows (See Fig. 3). All the selected ROI image data were then cropped and stored for further processing; the ROI image data collected from the FLIRTM camera output was directly reformatted into an 8-bit grayscale data array, while the green channel data from the built-in webcam had to be extracted from the RGB color space before doing so. The formatted data was also filtered through an outlier filter, which replaces any high contrast pixel with the mean value of the array if the standard deviation is 1.5 times higher or lower than the mean value, and was then smoothed by a Gaussian filter for further noise reduction.

An average value for each ROI data was then continuously calculated and fed into a corresponding buffer with a total size of 75. Once 75 useful rPPG frames were collected and calculated, the buffer was filled up and ready for the next process.

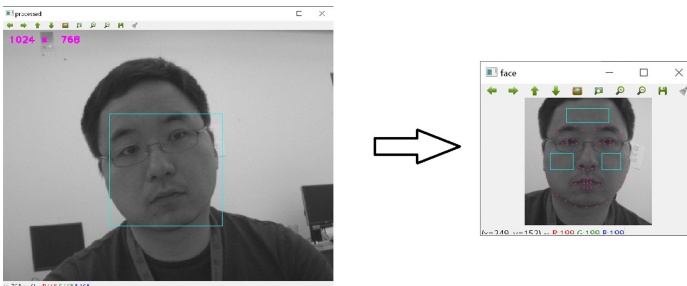


Fig. 3. Face detection, landmarks prediction, and face pose alignment.

2.8 Data Buffer Shaping and Conditioning

The data buffer was detrended to avoid the interference of light change during processing, and interpolated by 1 to smooth the change. It was then fed through a hamming filter to become more periodic for spectral leakage prevention, and was eventually normalized according to its average value. Next, the normalized data were amplified by a factor of 10 for boosting small temporal changes according to a pre-designed gain from Eulerian Video Magnification [20], and another Gaussian filter was applied here to further smoothen the data fluctuation.

2.9 FIR Band-Pass Filtering and Power Spectrum Density Selection

The Fast Fourier Transform is utilized to transfer the conditioned data in the buffer from the time domain into the frequency-power density spectrum. Here a 6th order FIR Butterworth filter with a low cut-off at 0.667 Hz and high cut-off at 3 Hz is applied to remove any data from outside the corresponding reasonable spectrum of the human heart rate frequencies, i.e. 40 bpm to 180 bpm. Finally, within the pass band range, the frequency indexes for the highest power density are picked to represent the desired heart rate on each ROI.

2.10 Heart Rate Calculation

The heart rate calculated from each ROI can be slightly different due to the varying SNRs and facial light conditions. The average of the final results was obtained by using a moving average filter with a window size of 20 values. According to our lab test feedback, among all the three results, the one with the least standard deviation over time seemed to have the closest value compared to our ground truth data, and thus, we selected this as the final output on the screen. In real-time, the screen will usually be updated approximately every 2 to 3 s.

3 Results Evaluation and Discussion

In this section, we have provided intermediate outputs at various stages of the entire data process, the end results obtained from multiple voluntary subjects, as well as our limited observation and analysis.

3.1 Intermediate Outputs

Intermediate data outputs were obtained from the FLIRTM grayscale camera with an optical green filter setup. Figure 4 shows a one-minute raw data averaged by the ROI region selected on the forehead of the test subject. From the data, we could see some clear ECG ripples accompanied by large fluctuation and noise DC trends. Next, Fig. 5 shows conditioned data with 8 clear heartbeat peaks in a data buffer filled by a 75-frames window. The detected frame rate was about 10 fps during the test, thus approximately 7 s was required to fill the buffer

window, yielding a heart rate of 68.5 bpm. Furthermore, in Fig. 6, the spectrum power density chart, calculated using the same data from the buffer window, indicates a peak of strong signal at close to 70 bpm. Both bpm's estimated from Fig. 5 and Fig. 6 are close to the ground truth result, 67 bpm, obtained from the contact hand bracket in Fig. 7. In addition, the plot in Fig. 8 gives a continuous bpm output for a length of roughly 50 s after 10 s of the start of the test. The relatively flat line shows a stable reading across the entire testing period. In short, the above figures have shown the capability of our system to produce a reasonable accuracy within 5% as compared to our skin-contacted ground truth.

3.2 Group Test Observation and Results

Table 1 and Table 2 show the detected heart rate results from both the sitting still and exercising conditions. Table 3 compares the percentage error with respect to the hand brace ground truth reading under each condition. Table 4 compares the overall correlation values with respect to the hand brace ground truth reading under each camera setup.

3.3 Evaluation: Sit Still vs. After Exercise

The calculated percentage error values with respect to our ground truth data for the FLIRTM camera with filter, without filter, and Webcam detection under sitting still conditions were 4.9%, 6.0%, and 4.2%, respectively, whereas the values were 9.9%, 10.0%, and 12.4% right after light exercise. The percentage error calculation formula is given by:

$$\text{Percentage Error} = \left| \frac{\text{Camera Reading} - \text{Ground Truth Reading}}{\text{Ground Truth Reading}} \right| \times 100\%$$

By purely examining the numbers, it can be observed that the results from the three setups are fairly close to each other, and the slightly larger deviation after exercise is possibly due to motion artifacts such as heavy breathing. Interestingly, even the built-in color webcam achieved accurate results regardless of having a smaller number of effective pixels, and the difference between the optically filtered grayscale camera and the webcam was negligible at times. This is contradictory to our original thoughts—we expected a higher accuracy output from the grayscale camera since it comes with a high resolution and a more sensitive image sensor. In fact, we found the reduced resolution on selected ROIs to be a contributing factor in stabilizing the output reading because the image output at the data acquisition stage was already heavily compressed and averaged by the camera's internal processor. Since the experiments were conducted under an indoor artificial light environment and each detection result was marked down by visual inspection over a period of roughly 30 s, this did not fully represent the actual performance of the system.

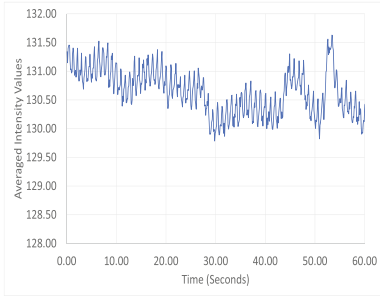


Fig. 4. A 60-s window: averaged intensity values obtained by one ROI Area vs. Time.

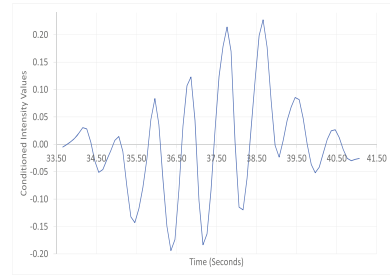


Fig. 5. An 8-s window: Normalized and Conditioned Intensity Values vs. Time.

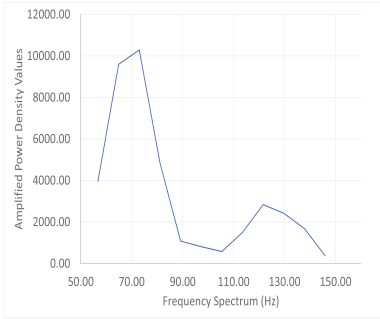


Fig. 6. Frequency spectrum from 50 Hz to 150 Hz: Amplified Power Density Values vs. Heart Rate.

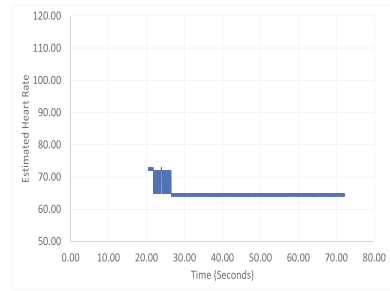


Fig. 7. A 72-s window: Estimated Heart Rate vs. Time.

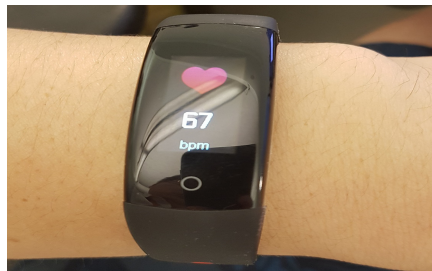


Fig. 8. Reference data: skin contact hand brace for ground truth heart rate detection.

Table 1. Sit still test results (all units are in bpm).

Test subjects	FLIR TM with filter	FLIR TM without filter	Webcam	Hand brace
Subject #1	66	60	62	64
Subject #2*	84	87 (unstable)	87 (unstable)	90
Subject #3	65	60	70	74
Subject #4***	71	71 (unstable)	70	79
Subject #5**	66	63	67 (unstable)	68
Subject #6	72	59 (unstable)	71	71
Subject #7	72	68 (unstable)	72	71
Subject #8	72	78	75	74
Subject #9	74	70	70	71
Subject #10*	60	63	62	62
Subject #11	76	69	72	80
Subject #12	70	71	68	71
Subject #13	88	88	87	84
Subject #14	96	95	96	93
Subject #15	72	78	74	75
Subject #16	72	74	76	75
Subject #17	64	64	68	64
Subject #18	60	64	65	70
Subject #19*	81 (unstable)	82 (unstable)	80 (unstable)	74
Subject #20	76	73	80	73

*with sunscreen and makeup **large amount of facial hair ***failed reference reading

Table 2. Light exercise test results (all units are in bpm).

Test subjects	FLIR TM with filter	FLIR TM without filter	Webcam	Hand brace
Subject #1	95	90	88	91
Subject #2*	98	90	105	74
Subject #3	96	92	92	96
Subject #4***	96	90	93	No reading
Subject #5**	100 (unstable)	58 (incorrect)	55 (incorrect)	86
Subject #6	90	78 (unstable)	90	88
Subject #7	83	84	96	95
Subject #8	90	107	102	96
Subject #9	80	85	80	90
Subject #10*	80 (unstable)	75 (incorrect)	76	110
Subject #11	93	88	85	87
Subject #12	100	94	65 (incorrect)	90
Subject #13	94	97	99	95
Subject #14	106	108	85 (incorrect)	100

(continued)

Table 2. (continued)

Test subjects	FLIR TM with filter	FLIR TM without filter	Webcam	Hand brace
Subject #15	108	110	102	95
Subject #16	90	88	85	95
Subject #17	96	95	101	95
Subject #18	91	88	82	83
Subject #19*	93 (unstable)	72 (unstable)	100 (unstable)	82
Subject #20	80	88	88	86

*with sunscreen and makeups **large amount of facial hair ***failed reference reading

Table 3. Percentage error comparison table.

	Percentage error with respect to reference		
	FLIR TM with Filter	FLIR TM without filter	Webcam
Sit still	4.9%	6.0%	4.2%
After exercise	9.9%	10.0%	12.4%

3.4 Evaluation: Correlation for Each Camera Setup

Figure 9, 10, 11, 12, 13 and 14 show the scatter plots for comparison with linear regression lines and the correlation values calculated based on the result data selection and camera setup. The red dots and lines represent the results by the FLIRTM camera with a filter on, blue represents results by FLIRTM without a filter, and pink represents results from the webcam. It may be hard to distinguish which setup would have an advantage in accuracy as compared to the reference by simply looking at the plot at the first place, however, by a closer examination, one can see that the correlation values give the information about which comparison is the best, and this confirms that the webcam performs less accurately compared to the FLIRTM camera.

The test results from subjects with heavy facial artifacts, makeup, sunscreen, and failed reference readings are treated as data outliers due to their instability in this experiment. In Table 4, we calculated the Pearson Correlation coefficient values in two groups, either using all test results or using results without outliers. The Pearson Correlation coefficient lies within the range of +1 to -1, with a number 0 considered to have no association between the two data sets. The readings from the FLIRTM camera with optical filter generated the highest correlation value of 0.75, which indicates a strong accuracy as compared to the

Table 4. Calculated correlation values with respect to hand brace ground truth.

	Correlation values with respect to reference		
	FLIR TM with Filter	FLIR TM without Filter	Webcam
All data	0.7517	0.7286	0.5934
Data outliers excluded	0.8871	0.8870	0.8102

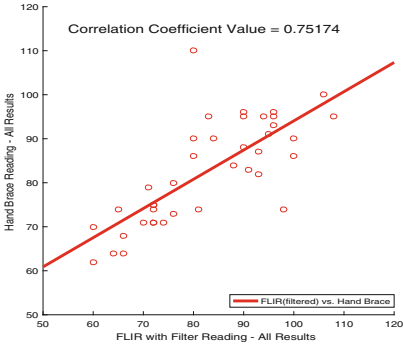


Fig. 9. Scatter plot - FLIRTM with filter (All Data).

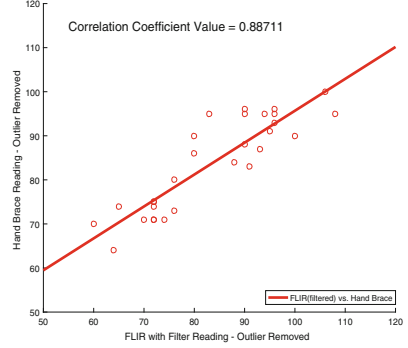


Fig. 10. Scatter plot - FLIRTM with filter (Outliers Removed).

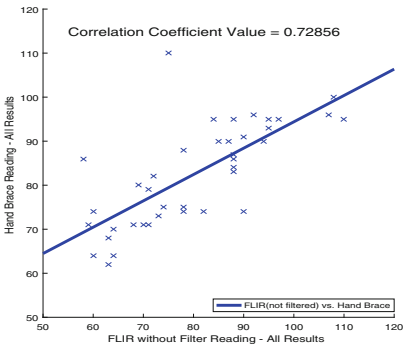


Fig. 11. Scatter plot - FLIRTM without filter (All Data).

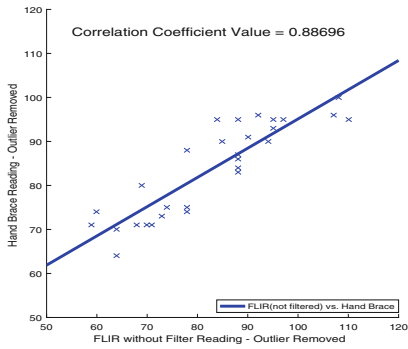


Fig. 12. Scatter plot - FLIRTM without filter (Outliers Removed).

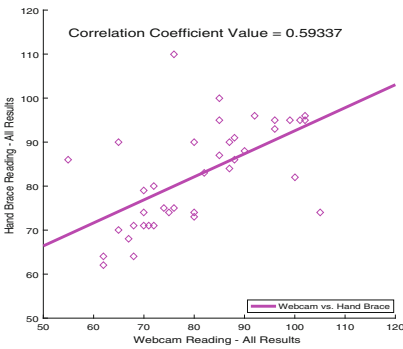


Fig. 13. Scatter plot - webcam (All Data).

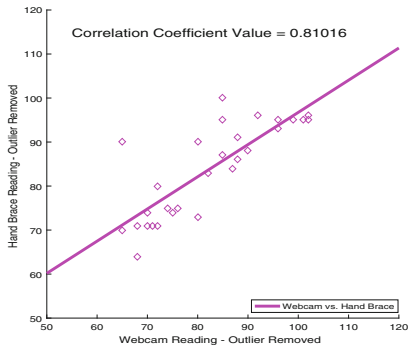


Fig. 14. Scatter plot - webcam (Outliers Removed).

reference reading among all the three setups. The lowest correlation value was 0.59 from the color webcam, which still indicates a good performance but with relatively large deviations. Interestingly, once we exclude the data obtained from the outliers, the correlation values increased to a much stronger level: 0.8871 for FLIRTM with filter, 0.8870 for FLIRTM without filter, and 0.8102 for the webcam. This would be an indication that the accuracy of the proposed system can be heavily reduced due to the presence of the above artifacts and should be properly dealt with while incorporating it in real-world applications.

In reality, when the optical filter was equipped, we noticed that the actual heart rate outputs had more stable readings and less fluctuation over time, and this was reasonable because the camera was receiving signals only from the green color spectrum where the signal for blood volume intensity change was strong. The results color webcam and the unfiltered grayscale camera appeared to be more unstable most of the time, and sometimes even incorrect due to the heavy noise inside the selected spectrum band. Additionally, it was obvious that the webcam struggled to produce stable results when we dimmed the light output in the room. The smaller aperture on the webcam with less effective pixels contributed much more noise as compared to the industrial FLIRTM camera, thus making this setup less desirable in a darker environment.

We also noticed that several factors could affect and reduce the accuracy of the results, and these should be taken into consideration in both laboratory and real-life environments:

- **Makeup and sunscreen:** Large fluctuations or even incorrect readings were observed in the heart rate of people wearing heavy makeup and sunscreen on their faces, due to the block of the reflective light change on blood oxygen saturation that was received by the camera.
- **Facial hair:** The heart rate of people with large amounts of facial hair can sometimes be hard to detect.
- **Length of the face:** People belonging to different races may have distinct facial length variations, and this may contribute to size variation of the ROIs selected by the Dlib facial landmarks engine.
- **Head movements:** Heads with regular horizontal movements introduced more fluctuation as compared to that of vertical movements. People who sat still had the most stable readings.
- **Sitting distance:** Even though the subjects were asked to sit at a distance of approximately 50 cm from the camera, this distance could vary slightly upon the actual execution during the tests.
- **Physical heart condition:** People with stronger heart conditions could have more blood volume pumped into their facial blood vessels, thus creating a higher signal to noise ratio, resulting in more stable detection.
- **Exercise intensity:** Exercises conducted by subjects depended on their personal preferences and the load to each of their hearts can also be different.
- **Speed of rest recovering:** Some people are able to calm down quickly after exercise, while others may take more time for it. Usually, heart rate readings gradually decrease after exercise, but the time to take a measurement could vary from person to person.

As future work, we will explore how to overcome the limitation of why the lightness of skin tone affects the accuracy of the system.

4 Conclusion

Multimedia in medicine has gone a long way in providing efficient clinical assistance. While a single image can provide rich information for a physical exam [21], a series of video sequences has proven to be much more useful for an in-depth clinical analysis [22]. In this paper, we have proposed a real-time rPPG system that is capable of detecting the vital signals of the heart rate by using either an optically filtered grayscale camera or a digitally filtered color camera sequentially from a distance. Then, we compared the experimental results on several subjects under a pre-setup indoor environment with the help of data collected by a skin-contact reference device. The obtained results show that while the built-in color camera can be great for handy heart rate detection given enough luminance on a sit still subject, the pre-optically filtered grayscale camera is more robust to facial and motion artifacts and can outperform a regular camera under less ideally lighted condition. In addition, we have also identified several factors that may reduce the accuracy of our proposed rPPG system, such as makeup and sunscreen, which heavily block light reflection from the face. Therefore, our future work will be focused on improving the accuracy of this study, dealing with more difficult reading scenarios such as sunscreen and heavy makeup, as well as combining other approaches possibly in machine learning fields for achieving better results.

Acknowledgments. Our work is fully supported by the Multimedia Communication Lab at the University of Ottawa and the ICSM2019 conference committee. We would also like to sincerely appreciate and express our gratitude to every participant for the time and effort they invested in conducting our tests.

References

1. Monfredi, O., et al.: Biophysical characterization of the underappreciated and important relationship between heart rate variability and heart rate. *Hypertension* **64**(6), 1334–1343 (2014)
2. Wang, C., Pun, T., Chanel, G.: A comparative survey of methods for remote heart rate detection from frontal face videos. *Front. Bioeng. Biotechnol.* **6**, 33 (2018)
3. Verkruyse, W., Svaasand, L.O., Nelson, J.S.: Remote plethysmographic imaging using ambient light. *Opt. Express* **16**(26), 21434 (2008)
4. He, D.D., Winokur, E.S., Sodini, C.G.: A continuous, wearable, and wireless heart monitor using head ballistocardiogram (BCG) and head electrocardiogram (ECG). In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2011)
5. Freitas, U.S.: Remote Camera-Based Pulse Oximetry. *eTELEMED* (2014)
6. Kwon, S., Kim, H., Park, K.S.: Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2012)

7. Po, L.-M., Feng, L., Li, Y., Xu, X., Cheung, T.C.-H., Cheung, K.-W.: Block-based adaptive ROI for remote photoplethysmography. *Multimedia Tools Appl.* **77**(6), 6503–6529 (2017). <https://doi.org/10.1007/s11042-017-4563-7>
8. Rahman, H., Ahmed, M., Begum, S., Funk, P.: Real time heart rate monitoring from facial RGB color video using webcam. In: *The 29th Annual Workshop of the Swedish Artificial Intelligence Society (SAIS)* (2016)
9. Fan, Q., Li, K.: Non-contact remote estimation of cardiovascular parameters. *Biomed. Signal Process. Control* **40**, 192–203 (2018)
10. Demirezen, H., Erdem, C.E.: Remote photoplethysmography using nonlinear mode decomposition. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2018)
11. Sanyal, S., Nundy, K.K.: Algorithms for monitoring heart rate and respiratory rate from the video of a user's face. *IEEE J. Transl. Eng. Health Med.* **6**, 1–11 (2018)
12. Monkareji, H., Calvo, R.A., Yan, H.: A machine learning approach to improve contactless heart rate monitoring using a webcam. *IEEE J. Biomed. Health Inform.* **18**(4), 1153–1160 (2014)
13. Qiu, Y., Liu, Y., Arteaga-Falconi, J., Dong, H., Saddik, A.E.: EVM-CNN: real-time contactless heart rate estimation from facial video. *IEEE Trans. Multimedia* **21**(7), 1778–1787 (2019)
14. Tang, C., Lu, J., Liu, J.: Non-contact heart rate monitoring by combining convolutional neural network skin detection and remote photoplethysmography via a low-cost camera. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2018)
15. Niu, X., et al.: Robust remote heart rate estimation from face utilizing spatial-temporal attention. In: *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)* (2019)
16. Wu, B.-F., Chu, Y.-W., Huang, P.-W., Chung, M.-L.: Neural network based luminance variation resistant remote-photoplethysmography for driver's heart rate monitoring. *IEEE Access* **7**, 57210–57225 (2019)
17. Chen, D.-Y., et al.: Image sensor-based heart rate evaluation from face reflectance using Hilbert-Huang transform. *IEEE Sens. J.* **15**(1), 618–627 (2015)
18. Kaehler, A., Bradski, G.R.: *Learning OpenCV 3: Computer Vision in C with the OpenCV Library*. O'Reilly Media, Sebastopol (2017)
19. Ucar, M., Hsieh, S.-J.: Board 137: MAKER: facial feature detection library for teaching algorithm basics in Python. In: *2018 ASEE Annual Conference & Exposition, Salt Lake City, Utah, June 2018. ASEE Conferences* (2018). <https://peer.asee.org/29934Internet>. Accessed 31 Jul 2019
20. Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J.V., Durand, F., Freeman, W.T.: Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.* **31**(4), 1–8 (2012)
21. Cai, J., Wang, X., Jiang, X., Gao, S., Peng, J.: Research on low-quality finger vein image recognition algorithm. In: *2019 International Conference on SmartMultimedia* (2019)
22. West, C., Soltaninejad, S., Cheng, I.: Assessing the capability of deep-learning models in Parkinson's disease diagnosis. In: *2019 International Conference on Smart-Multimedia* (2019)