# Analysis of minimum face video duration and the effect of video compression to image-based non-contact heart rate monitoring system

## Norwahidah Ibrahim<sup>1</sup>, Mohd Razali Md Tomari<sup>2</sup>, Wan Nurshazwani Wan Zakaria<sup>3</sup>

<sup>1,2</sup>EMCenter, Universiti Tun Hussein Onn Malaysia, Malaysia

<sup>2</sup>Department of Electronic Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Malaysia

## Article Info

## Article history:

Received Aug 27, 2019 Revised Nov 1, 2019 Accepted Dec 23, 2019

## Keywords:

Heart rate Image processing Photoplethysmograph Signal analysis Video compression

## ABSTRACT

Heart rate (HR) is one of important indicator for human physiological diagnosis, and camera can be used to detect it via photoplethysmograph (PPG) signal extraction. In doing so, number of sample images required to measure the HR signal, and quality of the images itself are important to yield an accurate reading. This paper tackles such an issue by analyzing the effect of sampling interval to HR reading in compressed and original video format, obtained in various ranging locations. Technically, important facial points from video stream were estimated by using cascade regression facial tracker. Based on the facial points, region of interest (ROI) was constructed where non-rigid movement is minimal. Next, PPG signal was extracted by calculating the average value of green pixel intensity from the ROI. Following that, illumination variation was separated from the signal via independent component analysis (ICA). The PPG signal was further processed using series of signal filtering techniques to exclude frequencies beyond range of interest prior estimate the HR. From the experiment it can be observed that sampling time of 2 seconds in uncompressed video shows promising HR within the range of 1 to 5 meters.

*This is an open access article under the <u>CC BY-SA</u> license.* 



## **Corresponding Author:**

Mohd Razali Md Tomari, EMCenter, Universiti Tun Hussein Onn Malaysia, 86400, Parit Raja, Batu Pahat, Johor, Malaysia. Email: mdrazali@uthm.edu.my

### 1. INTRODUCTION

Over the past years, many methods introduced to monitor human HR reading since it closely related to human physiological aspects. Recently, the interest is more concentrate on non-contact HR monitoring that especially useful to the patients with burn skin, elderly people that have fragile skin and premature infants that have extremely sensitive skin. One of the most cost effective non-contact devices is based on camera that measure the HR via PPG signal extraction. PPG is a simple non-contact optical measurement technique that can measure pulse activities that connected to human cardiac system from blood flow due to muscle contraction [1]. It was introduced back in 1973 by Hertzman et. all [2] that showed the light transmission variation of a finger could be detected by photoelectric cell. Based on his initial work, further research was conducted and found that, human face video that is recorded by normal camera under ambient light, conta in s useful signal that rich enough to measure the HR [3]. Some of the trend of works that utilize PPG signal to measure HR from colour-based method via web camera can be found in [4-8]. In the camera based HR domains, there are also reported that instead of using three colours channel (RGB), single green channel

provide more accurate outcome for PPG signal extraction because haemoglobin light absorption is most sensitive to oxy-generation changes for green light compared to blue and red lights [5, 9]. Another interesting works that had been introduced is pulse detection via head motion [10] that showed promising outcome for translating subtle head motion into HR estimator.

Almost all of the mentioned works proven to yield promising results, but their main concerns only revolve around motion artifact [11-14] and illumination variation [2-5]. They did not consider sampling time requirements and video compression that will affect the HR reading accuracy. Recently, there is one interesting report made by Yu et. all regarding minimum recording time requirement of input video for image-based monitoring system [15]. Based on their experiments, they stated that if longer video duration is used as an input for image-based monitoring system, the pulse reading accuracy would deteriorate. However, they still not consider the case where the video was compressed especially will be useful in the surveillance camera application. When the video was compressed, image quality of the face will be degraded and consequently will affect the PPG signal especially to the signal shape [16]. Meanwhile Mcduff et. all, stated that video compression degrade the signal to noise (SNR) ratio of PPG signal, thus affecting the accuracy of HR reading [17]. Work by Zhao et all also suggest that there are deterioration in *PPG* signal amplitude, SNR and signal trace due to video compression [18]. Their findings about video duration is very interesting, thus motivated us to analyse more by integrating the minimum time sampling requirements to the compressed video for measureing the non-contact HR reading.

#### 2. SYSTEM OVERVIEW

The framework of this project consists of five main steps which are facial detection and facial tracker algorithm, raw PPG signal extraction from green channel, illumination variation elimination using ICA, signal filtering and histogram analysis. Figure 1 depicted overall block diagram of the system. Initially, facial detection was applied to the recorded videos for localizing human's face in the videos. Next, facial tracker was applied to the detected face region to extract important facial points that later on will be used during PPG signal extraction. The facial tracker produced 49 points based on prominent human facial features, and based on these points reagion for raw PPG signal extraction will be labelled. PPG signal was then extracted from the labelled region using temporal random trace information of the green channel since green channel has a good SNR reading [19]. The extracted of raw PPG signal contains unwanted noise due to environment's illumination and motion artifact. To cater this, combination of Independent Component Analysis (ICA) [20] and series of signal filtering were applied to the raw signal and hence making the signal to be smoother and easier to work with. The refined PPG was then converted to frequency domain for determining the Power Spectrum Density (PSD) that will be utilize for the HR calculation.



Figure 1. Overall system plan

However, relying on single sequence of HR reading is still subject to the measurement variation. To overcome this, a histogram analysis of repetitive HR reading was constructed based on the same ROI with different random traces. Eventually, the HR is estimated from the average of histogram with lowest variation reading.

#### 2.1. Facial detection

Facial detection and facial landmark were used to detect the location and prominent facial features of targeted subject. In this project Viola-Jones (VJ) based facial detection using AdaBoost-based cascade with Haar like features [21] is employed. This classifier works by constructing a strong classifier (positive images) as linear combination weak classifiers (negative images). During detection, a series of classifiers are applied to every image sub-window with different scaling factor. Regions are considered valid if they pass through all the classifier stages. As for facial tracker, combination of Discriminative Response Map Fitting (DRMF) and Monte Carlo parallel linear regression [22] was used. The method works by performing a raw initial guess of facial landmark positions and uses a cascade of regressors to infer the shape as whole and explicitly minimizes the alignment error over the training data. The mathem atical modeling for the facial tracker can be represented as shown in (1), where S=(x1, x2...xp) denotes the coordinate of all p facial landmarks in a bounding box I and rt(...) be the regressor cascade.

$$S^{(t+1)} = S^{(t)} + r_t(I, S^{(t)})$$

(1)

This facial tracker would produce 49 facial points based on important human features which include the rigid and non-rigid points. Sample of the detected 49 points facial landmarks are as shown in Figure 2 where the left side shown the labelled number of facial points and the right side shows its real implementation on actual image. From the detected 49 points, four ROIs were selected to determine the most suitable face area that would results in most accurate HR for near and long distance application. The chosen face areas were as shown in Figure 3. Right cheek was chosen as the first ROI since less non-rigid motion is generated in this area compared to other regions. Another area selected was the center of face because a study claimed that this area considered to be the most suitable area for PPG signal extraction for video based HR system. Next, whole face area was selected since hypothetically, larger ROI means that the possibility to extract the PPG information from a far distance is high. However, 10% horizontal dimension reduction and 20% vertical dimension reduction were applied to this area in order to exclude unwanted background that might affect the PPG signal extraction process. Lastly, the final area selected for this project was lower region of face that includes nose and mouth but excluding eyes and chin area. This area was selected because the non-rigid motion is less and the area dimension is wider compared to the right cheek region. From the selected ROI, random pair temporal green color channel values that indicate the PPG signal is extracted.



Figure 2. 49 Facial landmark produced



Figure 3. Region of interest

## 2.2. PPG signal extraction

PPG signal was extracted from the green channel of the constructed ROI because of green channel contains strongest information for signal extraction due to the light sensitivity of hemoglobin. Since the extracted signal contained unwanted interference, mainly illumination variation, BSS technique known as ICA was used to separate the illumination variation from the true PPG signal. The visual representation of signal extraction is shown in Figure 4 where the top part is the original raw signal, and the bottom part is two signal produced by the ICA which the refined signal and its illumination variation. It can be observed that the polished signal pattern is more or less identical with the raw one as opposed to the predicted noise.

PPG signal using can be modelled using (2) where PPGraw is the true PPG signal and s is the green channel signal and y is the variation of illumination. If the parameter y can be estimated directly, then pure cardiac signal can be obtained easily, however in practice such signal cannot be measured straight away and hence ICA is used. ICA uncover the independent source of the signal by maximizing or minimizing a cost function of the mixing that measure non-Gausianity to uncover the mixture coefficients.



(2)



Figure 4. Visualization pf signal separation process using ICA

PPG signal obtained after the ICA process is a raw signal and still contains a fragment of unwanted noise. Thus some signal filtering processes were applied to obtain refined PPG signal. In this paper, detrending and moving average filters were applied to the signal for reducing slow, non-stationary trend of signal and polishing the random noise, making the signal smooth prior to frequency domain conversion [23, 24]. The filtered PPG signal was converted to frequency domain to determine the power spectrum density (PSD) using Welch method [25] with the constrained frequency spectrum within the range of 0.7Hz to 4 Hz that represent the HR value range from 42bpm to 240bpm. Lastly, the HR is calculated by multiplying the maximal PSD response with 60. The filtering processes for PPG signal were as shown in Figure 5.



Figure 5. Detrending and moving average process

## 2.3. Histrogram analysis

The HR value obtained from a single sequence calculation is still subject to the variation and hence a histogram based analysis is performed to determine the consistent reading against repetition. In this paper 10 repetition of HR reading from the same image sequence is performed. Any HR reading that is significant with the majority in the bin will be labelled as an outlier and eliminated while the remaining will be used to determine an average of HR which shows consistency over the sampled time period.

#### 3. RESULTS AND ANALYSIS

In the experiment, video recording was conducted in an indoor laboratory and normal ambient light was used as lighting source. The videos were recorded with  $1440 \times 1080$  pixels resolution at 60 FPS. The camera-subject distances vary between 1 meter, 3 meters and 5 meters respectively as shown in Figure 6. Pulse oximeter was attached to the participants' finger and reading from this device was made into ground truth for this project. There were two experiments conducted for this paper. The first experiment was to determine the minimum time requires for face video recording used in this project. For this analysis, 2 seconds, 5 seconds and 10 seconds videos were used. Another experiment was conducted to determine the effect of using compress video on the HR accuracy. Original video format is *mov* whereas the compressed video format is *wmv*, a fair comparison was made between HR results that were obtained from original and compressed videos. For this project, the accuracy and the error percentage were calculated using the shown equations.

Percentage of Error = 
$$\left(\frac{|\text{measured value} - \text{actual value}|}{\text{actual value}}\right) * 100\%$$
 (3)

Accuracy (%) = 100% – Percent of Error



Figure 6. General experiment set up for this project

Based on both tables reading, the errors calculated for each recording time with respect to distances did not exceed 50% which means that the system is capable of working properly even with short video sampling time. It can also be seen clearly that the three distances, 2 seconds sampling time provide the most accurate reading with 75% for 1 meter, 94% for 3 meters and 79.9% for 5 meters. As the sampling time increased (5 seconds and 10 seconds), the accuracy significantly reduced with an average of below 80% for various ranging of distances. Thus, based on this result, it was proven that the system able to work properly with acceptable accuracy with two seconds of sampling time. It is also worth to mention that, most of commercial pulse oximeter device required more than 5 seconds to obtain the HR reading. However for the non-contact camera based system, our results showed that reading with 2 seconds is enough.

Even though the accuracy from Table 1 and Table 2 is relatively high (above 80%), it was executed using compressed video format (*wmv*) to speed up the processing time. In the next experiment, we showed the effect of uncompressed video to the HR reading. In principle, with the uncompressed video image the processing time will be increased since density of the pixels in the picture fragment is slightly higher.

For this analysis, original images from the video frames (mov format) were used and compared with the uncompressed one (wmv). The time assigned was two seconds for both videos format since previously we have shown that this is optimal time sampling for calculating the HR. Technically 120 image frames from the video was used as input stream to the system. The result for this analysis was as shown in Table 3.

Based on the result obtained from Table 3, the reading of HR reading from original video showed great consistency and did not varies much from the ground truth. The experiment was repeated five times to determine the reading variation for the estimated HR. For experiment that used the compress video, there are fluctuations in the HR reading. For example, the HR reading of compressed video for 3 meter distance, the differences between the first reading and the second reading was inconsistence. However, there was no fluctuation in results obtained from original video. This clearly showed that using compressed video for as input for this system affected the HR reading accuracy. This happened because when the video format was changed, compression occurred to the videos and perhaps there were information loss during the compression process which caused the HR to be inconsistence and inaccurate.

Subjects	Distance (m)		<i>HR</i> reading for respective duration ( <i>bpm</i> )			
		Ground Truth (bpm)	2 seconds	5 seconds	10 seconds	
	1	91	86	70	79	
1	3	94	70	94	65	
	5	94	85	111	116	
2	1	65	91	115	88	
	3	75	61	92	56	
	5	75	89	76	82	
	1	70	88	92	109	
3	3	65	71	71	72	
	5	69	104	77	64	
4	1	69	76	85	64	
	3	68	76	85	71	
	5	70	68	100	107	
5	1	61	97	82	82	
	3	56	67	68	67	
	5	67	107	86	58	
	1	67	76	88	76	
6	3	68	62	92	77	
	5	66	73	82	85	
7	1	80	97	67	61	
	3	68	83	80	61	
	5	75	67	67	97	
8	1	56	70	74	88	
	3	59	64	73	62	
	5	59	62	92	86	

Table 1. HR Reading (Bpm) for assigned time

 Table 2. Accuracy Summary from Table 1

Range (m)	2 5		10					
	Accuracy (%)	Accuracy (%)	Accuracy (%)					
(HR accuracy results for near distance)								
1	75.00	66.40	70.00					
3	94.00	80.60	84.40					
(HR accuracy results for farther distance)								
5	79.90	75.50	73.30					
Average	82.97	74.17	75.90					

Table 3. Bpm readings for compressed and original video with 2 seconds sampling time

Video Tuno	Distance (m)	Ground Truth (bpm)	$1^{st}$	2 <sup>nd</sup>	3 <sup>rd</sup>	$4^{\text{tn}}$	5 <sup>th</sup>
video Type			Repetition of Estimated Reading ( <i>bpm</i> )				
	1	65	92	94	94	101	91
Compressed Video (wmv)	3	75	82	118	74	77	124
	5	75	68	79	69	74	98
	1	65	67	67	67	67	67
Original Video (mov)	3	75	74	74	74	74	74
	5	75	60	60	60	60	60

#### 4. CONCLUSION AND FUTURE WORK

This paper investigates time sampling requirements for calculating HR from compressed and uncompressed video samples. The PPG signals were extracted from eight different subjects that was recorded at 60 FPS from three different distances of 1 meter, 3 meters and 5 meters and hence producing 24 video samples. In the first experiment, three sampling time were analyze which are 2 seconds, 5 seconds and 10 seconds. Averagely from all various distances sampling time of 2 seconds yield 83% accuracy and beyond this point the accuracy level deteriorate significantly with below than 80%, and hence we conclude that 2 seconds sampling time is optimal to measure the HR that obtain from camera. In the second experiment we conclude that original uncompressed video or high quality videos will yields accurate and stable HR reading, but at a cost of longer processing time. In future, the system can be improved by optimizing the processing time for real implementation. It can be done by utilizing faster facial point's extractor module. Apart from that, accuracy of the system can also be improved by tuning the signal processing part for adapting more robust motion and illumination artifact.

#### ACKNOWLEDGEMENTS

The authors would like to thank to Ministry of Education (MOE) and Universiti Tun Hussein Onn Malaysia(UTHM) for supporting this research under Fundamental Research Grant Scheme (FRGS) (Vot. No 1582).

#### REFERENCES

- [1] J. Kranjee, S. Beguš, G. Geršak and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical Signal Processing and Control*, vol. 13, pp. 102-112, September 2014.
- [2] A. B Hertzman and CR Spealman, "Comparative Estimation of Blood Supply of Skin Areas from Photoelectrically Recorded Volume Pulse," *Experimental Biology and Medicine*, vol. 38, no. 4, pp. 562-564, 1938.
- [3] W. Verkruysse, L. Svaasand and J. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, vol. 16, no. 26, pp. 21434-21445, Dec. 2008.
- [4] C. Li, C. Xu, C. Gui and M. D. Fox, "Distance Regularized Level Set Evolution and Its Application to Image Segmentation," in *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3243-3254, Dec. 2010.
- [5] M. Poh, D. J. McDuff and R. W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," in *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7-11, Jan. 2011.
- [6] A. Qayyum, A. S. Malik, A. N. Shuaibu and N. Nasir, "Estimation of non-contact smartphone video-based vital sign monitoring using filtering and standard color conversion techniques," 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, pp. 202-205, 2017.
- [7] Q. Fan and K. Li, "Non-contact remote estimation of cardiovascular parameters," *Biomedical Signal Processing and Control*, vol. 40, pp. 192-203, Febuary 2018.
- [8] Q. Zhang, Q. Wu, Y. Zhou, X. Wu, Y. Ou and H. Zhou, "Webcam-based, non-contact, real-time measurement for the physiological parameters of drivers," in *Measurement*, vol. 100, pp. 311-321, 2017.
- [9] K. Matsumura, P. Rolfe, J. Lee and T. Yamakoshi, "iPhone 4s Photoplethysmography: Which Light Color Yields the Most Accurate Heart Rate and Normalized Pulse Volume Using the iPhysioMeter Application in the Presence of Motion Artifact?," in *PLoS ONE*, vol. 9, no. 3, pp. e91205, March 2014.
- [10] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, pp. 3430-3437, 2013.
- [11] X. Li, J. Chen, G. Zhao, and M. Pietikainen. "Remote Heart Rate Measurement from Face Video under Realistic Situations," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, pp. 4264-4271, 2014.
- [12] M. Kumar, A. Veeraraghavan and A. Sabharwal, "Distance cePPG: Robust non-contact Vital Sign Monitoring using a Camera," *Biomedical Optic Express*, vol. 6, no. 5, pp. 1565-1588, Apr. 2015.
- [13] A. Lam and Y. Kuno, "Robust Heart Rate Measurement from Video Using Select Random Patches," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, pp. 3640-3648, 2015.
- [14] R. Y. Huang and L. R Dung, "Measurement of Heart Rate Variability using off the shelf smart phones," in *Biomedical Engineering online*, vol. 15, no. 11, pp. 1-16, January 2016.
- [15] Yu, Y., Raveendran, P. and Lim, C., "Dynamic heart rate measurements from video sequences," in *Biomedical Optics Express*, vol. 6, no. 7, pp. 2466-2480, July 2015.
- [16] S. Hanfland and M. Paul, "Video Format Dependency of PPGI Signals," 20th International Student Conference on Electrical Engineering, May 2016.
- [17] D. J. McDuff, E. B. Blackford and J. R. Estepp, "The Impact of Video Compression on Remote Cardiac Pulse Measurement Using Imaging Photoplethysmography," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, pp. 63-70, 2017.
- [18] C. Zhao, C. Lin, W. Chen and Z. Li, "A Novel Framework for Remote Photoplethysmography Pulse Extraction on Compressed Videos," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, pp. 1380-1309, 2018.

- [19] T. Tamura, "Current progress of photoplethysmography and SPO2 for health monitoring," in *Biomedical Engineering Letters*, vol. 9, no. 4, pp. 21-36, Febuary 2019.
- [20] B. Holton, K. Mannapperuma, P. Lesniewski and J. Thomas, "Signal recovery in imaging photoplethysmography," in *Physiological Measurement*, vol. 34, no. 11, pp. 1499-1511, October 2013.
- [21] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 2001, pp. I 511-I 518, Febuary 2001.
- [22] P. Dollar, P. Welinder, P. Perona, "Cascaded Pose Regression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 1078-1085, 2010.
- [23] M. P. Tarvainen, P. O. Ranta-aho and P. A. Karjalainen, "An advanced detrending method with application to HRV analysis," in *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 2, pp. 172-175, Feb. 2002.
- [24] H. Lee, J. Lee, W. Jung and G. Lee, "The Periodic Moving Average Filter for Removing Motion Artifacts from PPG Signals," in *International Journal of Control, Automation, and System*, vol. 5, no. 6, pp. 701-706, December 2007.
- [25] [P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," in *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70-73, June 1967.

## **BIOGRAPHIES OF AUTHORS**



Norwahidah Ibrahim was born in Johor, Malaysia in 1993. She received the B.Eng degree in Electronics (Mechatronics) from University Tun Hussein Onn Malaysia in 2016. Currently, she is doing her M.Eng degree in Electrical at the same university. Her current research interest is computer vision, image processing and signal processing.



Razali Tomari was born in Johor, Malaysia, in 1980. He received the B.Eng degree in Mechatronics from the Universiti Teknologi Malaysia, in 2003, M.Sc in intelligent system from the Universiti Putra Malaysia, in 2006. And PhD degree in computer vision and robotic from the Saitama University, Japan in 2013. In 2003, he joined Faculty of Electrical Engineering, University Tun Hussein Onn as a tutor and later on become a Senior Lecturer in 2013. His current research interest includes computer vision, pattern recognition, smart wheelchair and sensing technology.



WN Wan Zakaria received B.Eng (2007) in Electronics and Mechanical Engineering from Chiba University and MSc (2008) and PhD (2012) from Newcastle University. She is currently a lecturer in Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia. Her current interests include Medical Robotics System specifically on development of robot force control, Image Processing and Computer Aided Diagnosis, and development of Wearable Device. She is author and co-author of several journal papers and conference proceedings.