

# Advancements in Contact-free Heart Rate Measurements Using Human Face Videos

Muhammad Waqar  
muw1@aber.ac.uk

Dr. Bernard Tiddeman  
bpt@aber.ac.uk

Department of Computer Science  
Aberystwyth University  
Aberystwyth, UK

---

## Abstract

Recently, several methods have been reported for the contact-free measurement of heart rate from human face videos. Typically, these are based on applying Independent Component Analysis (ICA) to sequences of colour values from tracked face patches and selecting the peak value within a fixed range in the frequency spectrum. Those methods performed well enough under controlled conditions, but in realistic situations their performance degraded significantly. In this paper, we have highlighted issues such as light reflection and participant's movement, and proposed solutions to those factors that reduce the performance of existing methods. We experiment with calculating a robust mean of the skin colour pixels values using a mean shift algorithm and estimating an optimal frequency selection function in the Fourier domain. Preliminary results show that these methods can improve the results of existing methods.

## 1 Introduction

Heart rate is usually expressed as the number of times the heart beats in one minute (bpm), and plays an important role in the measurement of human health. The significance of heart rate is vital in the fields of psychology, medicine and physics etc. [1]. The heart rate is currently measured using different methods such as ECG (electrocardiogram), sphygmology, pulse oximetry etc. ECG is considered to be the most accurate among them but it hinders the participant, as the equipment is in contact with the human body and hence it is uncomfortable. The need for non-contact heart rate measurement is very demanding as it could benefit many digital systems, where the heart rate calculation is required, such as vehicle driving monitoring, PC workspace, patient treatment, sports monitoring etc.

The idea for calculating non-contact heart rate was first introduced in 2007 by Pavlidis et al., who measured the heart rate and breath rate from human face images using a thermal camera [2, 3, 4, 9]. Gatto found cardiac pulse at the forehead using images extracted from the infrared camera videos. He monitored the variations in the blood flow with each pulse cycle that fluctuated the thermal energy released by the body tissue [5]. Takano and Ohta used Charge-Coupled Device (CCD) camera to measure the heart rate from images [6]. They measured the variations of average brightness in the region of interest of the participant's skin. Photoplethysmography (PPG) is the basic principle of most of the recently implemented non-contact heart rate measurement systems. This phenomenon

states that, when a heart beats, it produces a pulse wave that travels across the entire arterial body vascular system. When this wave approaches the face, it causes a small volume change of blood at this point [7]. The naked eye cannot view such a small volume change but we can measure the intensity of the absorbed light by means of images taken from a webcam. This phenomenon presents the most inexpensive way of measuring non-contact heart rate from videos as we can capture colour video from different less expense sources such as a webcam, video camera or mobile phone camera.

Recently, several methods have been reported for the measurement of contact-free heart rate from human faces using a webcam [13, 16]. The performance of those methods was found to be very reliable under controlled conditions but in realistic situations, their performance degrades significantly. Illumination variations and participant's movements are the two major factors that affect the performance of these existing methods. In an HCI environment, if a participant is watching something then the reflected light from the screen, the flicker of the indoor lights, the sunlight variations due to clouds could affect the grey value of the region of interest on the face. Participant's rigid movements, head tilt and facial expressions could also be expected to degrade the overall performance.

In this paper, we have proposed a framework that can reduce the effect of the above mentioned factors. A robust mean algorithm based on mean shift [11] has been introduced that attempts to take mostly skin pixels into account when measuring the skin colour in each frame and ignore any pixels affected by the reflected light. An optimization technique is introduced in order to improve the identification of the correct heart rate peak in the frequency spectrum. We have used Saragih's Face Tracker [8] to track faces from the video images which tolerates slight movements of the participants and ensures less or no loss of frames. We have demonstrated in our preliminary experiments that these methods can reduce the effect of the above factors to some extent and increase the accuracy of contact-free heart rate measurement from human faces under realistic situations.

## 2 Methods

### 2.1 Experimental Setup

Our sample featured 10 participants of both genders and different skin color. All experiments were conducted indoor with the office fluorescent lights as the only source of illumination. The participants were asked to sit on a chair in front of a laptop with built-in webcam at a distance of about 0.5 metres. We played a video clip on the laptop screen that contained different emotional cues and asked the participants to view it. A color video of one minute was recorded for all the participants. The webcam used in the experiments have the following specifications: 1.3 megapixels (mp), 24-bit RGB with 3 channels (8 bits/channel), 15 fps with pixel resolution of 640 x 480. All videos are saved in .AVI format on the laptop with unique identifiers. Since, participants were watching a movie clip therefore their facial features (eyes, nose, mouth, head etc.) were in motion during the recordings. The reflected light from the laptop falls on the participants that also caused variations in the color on the face. For testing and synchronization purpose, we have measured the heart rate of all participants using the Arduino based pulse sensor simultaneously. The pulse sensor was clipped to the participant's fingertip and was connected to the laptop via an Arduino microcontroller. Figure 1 shows our experimental setup.

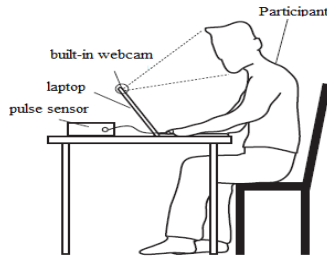


Figure 1: Experimental Setup. [13]

## 2.2 Heart rate extraction methodology

The change in blood volume of the facial vessels during each cardiac cycle changes the amount of light reflected. This tiny change in color can be captured by RGB colour sensor of a webcam. We refer to this small reflected light colour change as a plethysmographic or pulse signal. We have processed and analysed our video data using the software coded in C++/OpenCV and MATLAB R2015a. (The MathWorks, Inc.)

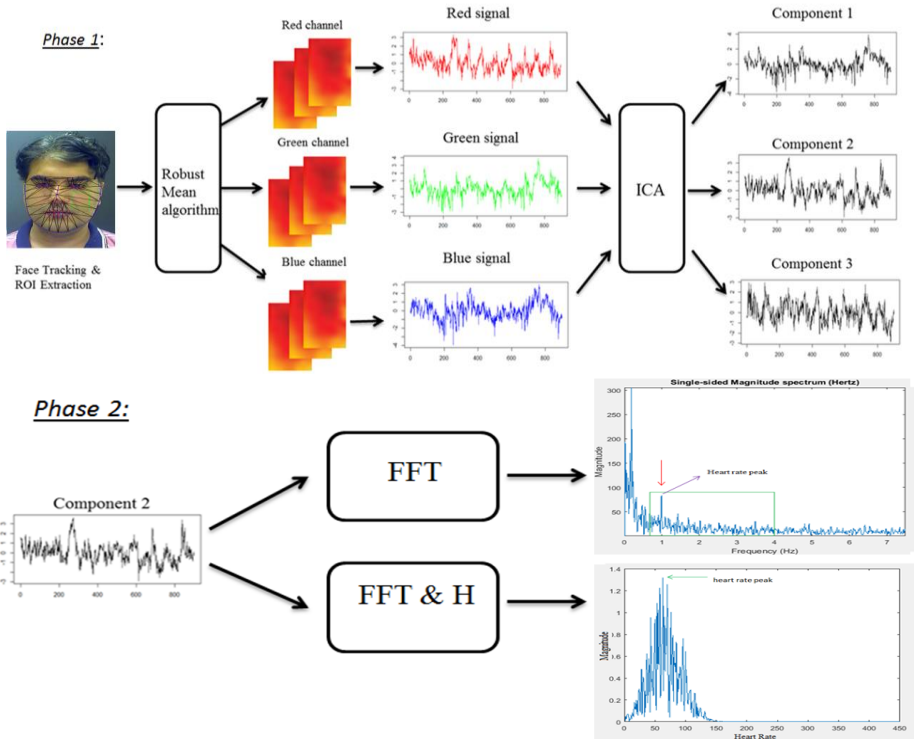


Figure 2: Heart rate extraction methodology. Phase 1: Face is tracked and ROI (green rectangle) extracted. Robust Mean Algorithm applied to ROI. RGB color channels obtained and normalized to RGB raw signals. ICA is implemented on three input signals and three output components obtained. Phase 2: FFT applied on component 2 and the heart rate frequency peak obtained (red arrow shows the pulse sensor heart rate). FFT is applied to component 2 and multiplied by the selection function (H) to obtain the heart rate.

Figure 2, presents an overview of the steps involved in extraction of a heart rate from the recorded videos using robust mean and optimal selection function. Tracking is an important phase in order to extract accurate region of interest (ROI) from our video data. We have used Saragih's Face Tracker [8] to track faces from the video images. This approach is more robust and it tolerates slight movements of the participants and ensures less or no loss of frames. The issue of movement arti-facts has been resolved up to considerable extend by applying this tracking algorithm and it allows participants to move slightly, while watching the movie clip on the laptop. We have then chosen cheeks as our region of interest because it is least likely to be corrupted by hair, beard or glasses. The ROI (indicated as green rectangle in Figure 2) was extracted using OpenCV library function. We have used the Mean Shift algorithm [11] with a flat kernel to make our mean more robust so that it can ignore non-skin pixels values and taking into account only the skin pixels values, so that the effect of reflected light from the laptop has been eliminated from the region of interest in each frame of a video. The estimated skin color from each frame yields three raw signals  $x_1(t)$ ,  $x_2(t)$  and  $x_3(t)$ , which were then normalized as follows:

$$x'_i(t) = \frac{x_i(t) - \mu_i}{\sigma_i} \quad (1)$$

for each  $i=1, 2, 3$  where  $\mu_i$  and  $\sigma_i$  are the mean and standard deviation of each color trace ( $x_i$ ) respectively. The raw signals  $x_i$  are now transform to normalized raw signals  $x'_i$  that has a zero mean and unit variance.

The ICA (Independent Component Analysis) was introduced at this point and the normalized RGB signals were decomposed into three independent components. In our study, we have used joint approximate diagonalization of eigenmatrices (JADE) algorithm of ICA [10], which is capable of removing noise occurred by participants movement. The returned independent components were in random order but we have selected the component with the highest peak in its power spectrum as the desired pulse signal. The second component qualifies this condition and was considered for further processing.

FFT (Fast Fourier Transform) was applied to the source signal in order to get the power spectrum. We have designated heart rate frequency as the frequency with the highest peak in the power spectrum of a signal within operational frequency band. The operational range was set to [0.7-4] Hz that corresponds to [40-240] bpm, which covers almost all acceptable heart rate measurements.

Our experiments have shown that detection of the peak within a fixed window is not reliable. We have developed a selection function (inspired by the MOSSE filter optimization process [12]) to identify the correct heart rate peak and measure the heart rate of a participant. In this regard we have designed a selection function which is given below:

$$H = \frac{\sum_i G_i \cdot F_i}{\sum_i F_i \cdot F_i} \quad (2)$$

where  $H$  is the estimated selection function,  $G$  is a Gaussian function centered on the known heart rate of the  $i^{\text{th}}$  video clip with width  $s$ , and  $F_i$  is the  $i^{\text{th}}$  training source signal (component 2), all in the Fourier power domain.

The heart rate of a new input signal  $F_j$  is then calculated as follows:

$$\text{Heart Rate} = HF_j \quad (3)$$

Where  $F_j$  is the power spectrum of the desired source signal and  $H$  is the selection function.

We have trained our system by combining all the participants, FFT of component 2 signals and the true heart rate obtained by Arduino pulse sensor.

### 3 Results

Using the methodology detailed in Section 2, we have tested our results for two cases. First, we extracted RGB signals from our ROI and applied ICA on them to obtain three components and the component with the highest peak in its power spectrum (component 2) was considered as the desired pulse signal. We then applied fast Fourier Transform (FFT) to the component 2 signal to get the power spectrum of it in frequency domain and hence obtained the heart rate frequency. In the second case, we applied robust mean algorithm to our ROI and the resultant RGB signals are then exposed to ICA to yield the component 2 signal. Finally FFT is applied to obtain the power spectrum of the component 2 in frequency domain. Figure 3, presents the components 2 obtained for both cases and Figure 4 shows the power spectrum in frequency domain of their component 2 respectively.

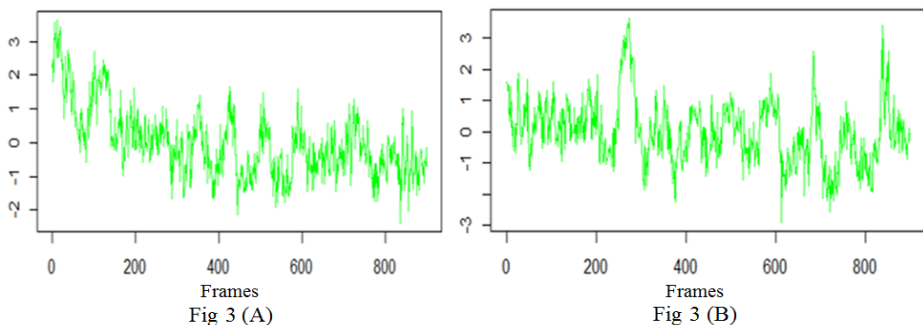


Figure.3: Figure 3(A) shows component 2 obtained without the use of robust mean algorithm. Figure 3(B) shows component 2 obtained by using robust mean algorithm.

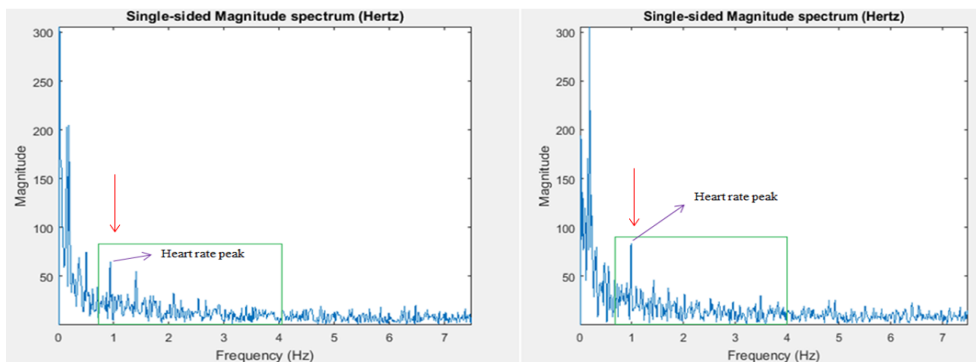


Figure 4: FFT of component 2 without robust mean (left), FFT of component 2 with robust mean (right). The red arrows indicate the heart rate as measured with the Arduino sensor. The purple arrows indicate heart rate frequency peak measured from video. Green rectangle indicates the operational frequency range.

Figure 4 clearly shows that the peak in the power spectrum of component 2 signal become more prominent and closer, to the red arrow (ground truth heart rate) after applying the robust mean algorithm, at a certain frequency, which corresponds to the estimated heart rate. Table 1 shows the estimated heart rates with and without the use of robust mean algorithm and their error rate for each case in comparison with the heart rate obtained by means of an Arduino pulse sensor. The results show that error rate for the robust mean algorithm is found less than the non-robust mean of the ROI and quite acceptable to the original heart rate.

Videos	Pulse sensor (bpm)	Estimated (bpm)	Robust Mean (bpm)	Estimated Error (%)	Robust Mean Error (%)	Selection Function (bpm)	Selection Function Error (%)
1.	58	55.98	58.99	2.12	0.99	60	2.0
2.	65	62.10	63.85	2.90	1.15	62	3.0
3	75	73.50	76.10	1.50	1.10	74	1.0
4	67	65.20	65.30	1.80	1.70	65	2.0
5	72	69.80	71.50	2.20	0.50	67	5.0
6	62	59.60	60.80	2.40	1.20	58	4.0
7	57	59.20	56.75	2.20	0.25	57	0
8	75	74.10	76.20	0.90	1.20	71	4.0
9	66	64.01	65.10	1.99	0.90	62	4.0
10	71	69.30	71.50	1.70	0.50	74	3.0
Average				1.971	0.949		2.80

Table 1: Comparison of heart rate obtained from video with and without the use of robust mean algorithm, and by means of Selection function.

Figure 5 illustrates the outcome of the selection function, we applied to the FFT of component 2 signal. Table 1 shows heart rate obtained by using selection function and its error as compared to the arduino heart rate measurement. The error found was more than the robust mean method and therefore it require further improvement.

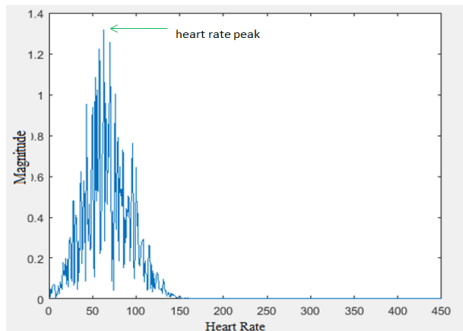


Figure 5: Heart rate peak obtained as a result of selection function.

## 4. Conclusion and Future Work

In this paper, we addressed issues of illuminations and movement distractions while measuring the heart rate from face videos using a webcam. We proposed robust mean of skin colour pixels values using a mean shift algorithm and estimated an optimal frequency selection function in the Fourier domain of the source signal. Preliminary results of our work proved that the heart rate frequency peak get more prominent and closer to the ground truth heart rate frequency after we applied robust mean to our video data. The robust mean algorithm reduced the effect of illumination and extracted mostly skin color pixel from the ROI. We have used Saragih's face tracking approach to minimize the loss of frames resulted by the participant movements. These techniques contribute towards the above mentioned issues for the measurement of heart rate from videos. The optimal frequency selection function, we used to measure heart rate from videos, require improvement in order to reduce its error rate.

In our future work, we will try to measure heart rate from videos under more challenging environment. We will try to measure the heart rate outdoor in the presence of sunlight and varying illumination lights. Some useful work has been done by using video magnification approaches [14, 15] to reduce the issue of motion in the images, such approaches if utilized, could improve the heart rate measurement from videos.

## References

- [1] Deter, "General Internal Medicine" Stuttgart, Vandenhoeck & Ruprecht, (2007).
- [2] I. Pavlidis, J. Dowdall, N. Sun, C. Puri, J. Fei and M. Garbay, "Interacting with Human Physiology", Computer Vision and Image Understanding, vol no. 108, pp. 150-170, (2007).
- [3] I. Pavlidis, M. Garbay, N. Sun and A. Merla, "Contact-free Measurement of Cardiac Pulse Based on the Analysis of Thermal Imagery", IEEE Transaction Biomed.,pp. 1418-1426 (2007).
- [4] I. Pavlidis and J. Fei, "Thermistor at a Distance: Unobtrusive Measurement of Breathing", IEEE Transaction Biomed., pp. 988-998(2010)
- [5] R. G. Gatto, "Estimation of Instantaneous Heart Rate using Video Infrared Thermography and ARMA Models", Ph.D. thesis, University of Illinois at Chicago, 2009.
- [6] Chihiro Takano and Yuji Ohta, "Heart Rate Measurement based on a time-lapse image," Medical Engineering and Physics, vol. 29, no. 8, pp. 853-857, (2007).
- [7] J. Allen, "Photoplethysmography and its application in clinical physiological measurement." Physiol. Meas., vol. 28, pp. R1-R39, Mar.(2007).
- [8] Jason Saragih. (2015). FaceTracker. Available:<https://github.com/kylemcdonald/FaceTracker>. Last accessed 10/03/2015.
- [9] M.Garbey, N. Sun, A.Marla, and I. Pavlidis, "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery," IEEE Trans.Biomed. Eng., vol. 54, no. 8, pp. 1418-1426, Aug.(2007).
- [10] J. Cardoso, "High-order contrasts for independent component analysis (ICA)", Neural Comput. 11(1), 157-192, (1999).
- [11]OpenCV.(2015). MeanShift. Available:[http://docs.opencv.org/master/db/df8/tutorial\\_py\\_meanshift.html](http://docs.opencv.org/master/db/df8/tutorial_py_meanshift.html). Last accessed 05/06/2015.
- [12] David S. Bolme, J. Ross Beveridge, Bruce A. Draper and Yui Man Lui . Visual Object Tracking using Adaptive Correlation Filters. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference pp. 2544-2550, (2010)
- [13] M.-Z.Poh, d.j.McDuff, and R.W.Picard, Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt. Express, vol no. 18, pp. 10762-10774. (2010)
- [14] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Fredo Durand, William T. Freeman. Eulerian Video Magnification for Revealing Subtle Changes in the World. ACM Transactions on Graphics, Vol 31, Number 4 (2012)



- 
- [15] Mohamed A. Elgharib, Mohamed Hafeeda, Fredo Durand, William T. Freeman. Video Magnification in Presence of Large Motions. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2015
- [16] Yong-Poh Yu, Ban-Hoe Kwan, Chern-Loon Lim, Siaw-Lang Wong, P. Raveendran. Video-based heart rate measurement using short-time Fourier transform. Intelligent Signal Processing and Communications Systems (ISPACS), 2013