

A remote PPG based approach for heart rate estimation

Tran Thi Thao^{*}, Pham Van Truong, Phan Dai Duong

Hanoi University of Science and Technology.

^{*}Corresponding author: thao.tran@hust.edu.vn

Received 29 March 2022; Revised 19 April 2022; Accepted 26 April 2022; Published 19 May 2022.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.79.2022.31-40>

ABSTRACT

Heart rate (HR) estimation is an essential physiological process in the field of biosignal analysis. Recently, remote Photoplethysmography (rPPG) is a pathbreaking development in this field wherein the PPG signal is extracted from non-contact face videos. This study, aiming at building a low cost rPPG system for both heart rate estimation and detecting real face, proposes a non-contact, automatic technique of heart rate measurement from video based on the signal processing techniques. The heart rate is remotely measured from a video recording of a person's face using the camera from a smartphone. At the same time, the heart rate extracted from the videos recorded by a smartphone camera is compared with the data obtained from the PPG optical absorption sensor and achieved a similar result. The system can estimate the heart rates from faces from single subject or multiple participants, and can be applied for anti spoofing face recognition.

Keywords: Heart rate estimation; Non-contact heart rate monitoring; PPG; Face recognition; Real face detection, Camera-based contactless PPG.

1. INTRODUCTION

Heart rate (HR) estimation is an essential physiological parameter in the field of biomedical imaging. Remote Photoplethysmography (r-PPG) is a pathbreaking development in this field wherein the PPG signal is extracted from non-contact face videos. In the COVID-19 pandemic, rPPG plays a vital role for doctors and patients to perform telehealthcare. At this crucial time of a contagious pandemic of COVID-19, contactless physiological signal sensing is of utmost importance. One of the most critical bio-signals that could indicate an underlying medical condition is cardiovascular activity. Heart rate (HR) is the crucial parameter that measures the heart's function and reflects the balance of sympathetic and parasympathetic activity [1, 2].

Photoplethysmography (PPG) with pulse waveforms generated from optical sensors of mobile devices has become a new trend and shows sufficient accuracy for the detection of heart rate and other physiological parameters. An emerging noncontact technique called remote photoplethysmography (rPPG) has been developed for detecting heart rate, which uses digital camera to measure the subtle variations of skin color reflecting the cardiac pulsatile signal due to heart activity pumping blood to and from the face. This method would potentially be applied to mass screen with less cost, as well as to long-term monitoring under appropriate settings.

Developing rPPG is an active area of research due to its potential for instant remote physiological monitoring while ensuring better safety for the healthcare workers, especially during a pandemic of contagious diseases. The contactless nature also provides a quick screening of cardiovascular activity with requiring only a video camera sensors. Traditionally, extracting PPG from a facial video sequence consists of two distinguished parts; detect, localize the bare skin, and apply systematic signal processing techniques on the localized skin to extract the PPG. Several source separation methods

include Independent Component Analysis (ICA) in Poh et al. [3] and McDuff et al. [4]. Singular Spectrum Analysis (SSA) on compressed videos in Zhao et al. [5]. Demirezen and Erdem [6] have shown their success in PPG decoding from the video signal. Such algorithms heavily depend on the skin/face detection aspect and skin signal quality, strength, and variation. Recently, deep learning (DL) frameworks can approximate the function for PPG extraction from skin videos. Huynh et al. [7] have demonstrated the ability of convolutional neural network (CNN) based architecture to detect PPG peaks (acquired at a higher sampling rate) from raw video frames. The authors used CNN and Fully connected feed-forward (FCFF) network in the second step neural network to further localize the ECG signal peak position in the finer scale from the video frames with regression loss in the frame prediction. The authors proposed a network that uses phase information to predict the ECG peak occurrence time and HRV in higher resolution. In a recent study, Hassan et al. [1] focused on reconstructing the entire PPG signal rather than the ECG peaks, not the entire ECG reconstruction. Though having shown promise in extracting PPG information from video sequences. Zhan et al. in [8] shown that DL methodologies depend on video post-processing of network output, and often with high cost.

In this study, based on recent research in the world, especially the work of Poh et al, this study proposes a method of measuring heart rate by the non-contact method. Specifically, the study demonstrates the non-contact, automatic technique of heart rate measurement from video based on the signal processing technique. In particular, the study aims to build a low cost rPPG system for both heart rate estimation and detecting real face. The system can be applied for anti spoofing face recognition. First, the method is reviewed and applied to measure heart rate from a video recording of a person's face using the camera from a smartphone. At the same time, the heart rate extracted from the videos recorded by a smartphone camera is compared with the data obtained from the PPG optical absorption sensor and achieved a similar result.

2. MATERIALS

2.1. Methods of measuring heart rate

2.1.1. Measured by electrocardiogram (ECG)

An electrocardiogram (ECG or EKG -Electrocardiogram) is a simple, painless test that records the electrical activity of the heart. The electrical activity of the heart involves electrical impulses through the heart which in turn determine the heart rate and pulse. With each heartbeat, an electrical signal travels from the top to the bottom of the heart. While moving, these signals will cause the heart to contract and pump blood away. This process repeats with each heartbeat. The heart's electrical signals determine the heart's beat rate.

Measuring heart rate by electrocardiogram method gives highly accurate results and is not harmful to the body. However, the equipment of this method has a large size, high cost, and is usually only used in hospitals and clinics. Therefore, using this method would not be in line with the criteria originally set for the product to be small, compact, lightweight, low cost, flexible, and portable. Therefore, an alternative method is necessary.

2.1.2. Measurement of heart rate by optical absorption method (Photoplethysmography)

When the heart expands to pump blood through the body, the pressure in the arteries decreases, and when the heart contracts, the pressure in the arteries increases. This rise and fall in blood pressure will change the amount of blood in organs and parts of the body, thereby changing the amount of HbO₂ in the blood in the arteries, so when light is transmitted through the pulse, the light intensity after passing through the artery will vary synchronously with the heart rate. When a Phototransistor is arranged to receive light after passing through an artery, one can receive an electrical signal that varies in sync with the heartbeat because changing the light intensity on the Phototransistor side will lead to a change in the current flowing through the Phototransistor. The suitable place to place the infrared source and Phototransistor is the fingertips.

Measuring heart rate by use of Photoplethysmography (PPG) is a non-invasive method, does not affect the circulation of blood in the arteries, so it gives reliable and highly accurate results. However, this method will not give accurate results if noise filtering problem is not addressed in the optical sensor well. These interferences can be noise from daylight, light from electric bulbs in the room. Besides, there is the influence of vibration factor as well as the influence of skin pigmentation on the measurement results.

2.1.3. Non-contact measurement

The ability to monitor a patient's physiological signals in a remote, contactless way is a prospect for improving the quality of healthcare. For example, the idea of measuring facial physiology was first put forward by Pavlidis et al. [9] and later proved through the analysis of thermograph videos of the face. Although non-contact methods may not provide the details of the electrocardiographic data that an ECG produces, these methods can provide long-term monitoring of other physiological signals such as heart rate or oxygen by collecting them continuously discreetly and comfortably.

The non-contact heart rate measurement algorithm based on images from the video has appeared and developed since the early 2010s, especially with the popularity of smartphones. This method has been generalized and modeled in the paper [3]. Specifically, the method uses images of human faces via recorded video and automatic face detection techniques along with blind source separation techniques that convert color channels into independent components. It is a low-cost, acceptable accuracy heart rate method for flexible heart rate monitoring applications with video and contactless input.

2.2. Theoretical basis of face recognition

In the face recognition algorithm, determining whether a face is in the frame or not is very important, this will help the recognition algorithm process faster, saving considerable time. The pose of the human face has more influence on face detection than other objects, human faces are a variable subject, they come in many shapes and colors. However, face detection has many benefits. Face recognition is not possible if the face is not separated from the background.

2.2.1. Viola-Jones algorithm

Viola-Jones algorithm is popular algorithm used to face detection [10]. The object

identification system is an array of images of known size or a feature and to decide if these properties are derived from an object. For properly and reasonably object identification, Haar Classifier is used in this method. Haar features are used to classify object images. This feature set looks at rectangular regions of the image and aggregates the pixels in this region. The obtained value is used for image classification. One can classify all the images with the same Haar-features in this rectangular region in a certain value range as one type and those out of this range in another. This can roughly split the set of images into images with lots of faces and a few buildings and other buildings. This can be done iteratively to split groups of images.

2.2.2. Adaboost

AdaBoost is a complex nonlinear strong classifier based on the boosting approach introduced by Freund and Schapire in [11]. Adaboost also works on the principle of linearly combining weak classifiers to form a complete classifier. As an improvement on the boosting approach, AdaBoost additionally uses the concept of weights to mark patterns that are difficult to identify. During the training process, for each weak classifier class built, the algorithm will update the weights to prepare for the construction of the next weak classifier: increase the weight of the wrongly recognized samples and decrease the weight of the correctly recognized samples. by the newly built weak classifier. In this way, the later weak classifier can focus on samples that the previous weak classifier did not do well. Finally, the weak classifiers will be combined according to how good they are to create a complete classifier. Viola and Jones use AdaBoost to combine weak classifiers using Haar-like features according to a cascade model.

3. PROPOSED APPROACH

3.1. Working principle

The idea of the non-contact heart rate measurement method comes from the nature of the circulation, that is, when the heart contracts, it will push blood around the body, when the heart expands, blood flows into it, at this time the pressure of the amount of blood in the arteries decreases and when the heart contracts, the pressure in the arteries increases. It is this rise and fall in blood pressure that will change the amount of blood in organs and parts of the body. This change will cause changes in the light on the face that is invisible to the naked eye. Moreover, this change will be detected by analyzing the RGB components of consecutive photos taken from the smartphone's camera. Then by graphing the values of the RGB components, one will get the raw signal form. The value of this signal changes rapidly from frame to frame. These signals will consist of peaks, each of which is equivalent to a beat value. The heart rate will be obtained by counting the peaks of the signal.

However, in practice, detecting the peak of the signal is not an easy thing. This signal is susceptible to various types of noise and overlap during measurement and data acquisition. The noise can be physical noise due to the movement of the patient, noise caused by power sources, environment, errors in calculation, noise from electronic devices during the main data acquisition process. Therefore, one need to process the raw signal obtained from consecutive frames through independent component analysis, spectrum analysis, and signal filtering.

Based on practical experience and from studies in the world, the raw signal from the green channel gives the best and clearest PPG signal. This is very similar to what Poh et al. [3] did. Therefore, the green channel is chosen to process and calculate the heart rate. After obtaining the internal signal, the FFT analysis is applied to the signal which will get the peak of the signal in a period of 10 seconds. Then take the average spectral density over the whole time to obtain the heart rate frequency. Multiplying by 60 one can get the heart rate of the person measuring.

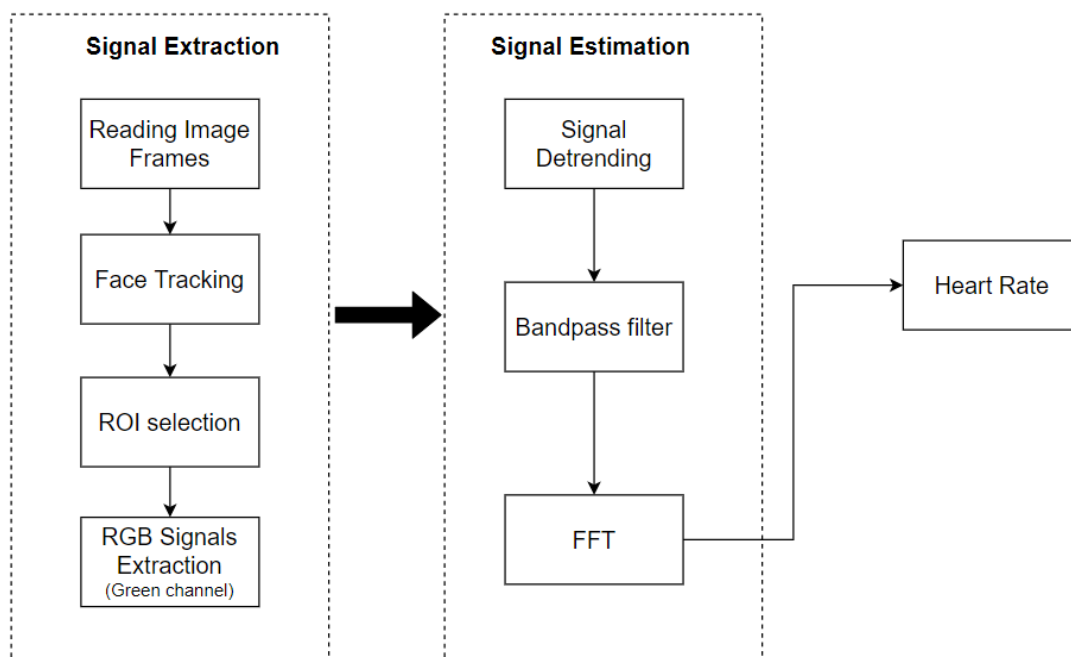


Fig. 1. Heart rate algorithm based on a non-contact method using video images.

3.2. RGB signal extraction

First, the data are acquired from the front camera of a smartphone or laptop with HD720 quality at 30 frames (frames) per second. The next step is to get the gray level of the image. The face is localized by using the Viola-Jones algorithm. By using Haar-like features on the person's face are the horizontal feature of the eyes, the vertical feature of the nose, and the horizontal feature of the mouth. Then, the above features are combined to locate the face. The computer will scan each area of the image, each sub-image area will be checked with features, if there is a Haar-like feature that results in a non-human face, it will be removed and searched until it has enough basic features.

Face positioning will be difficult if there are influences such as light or when there is fast movement. Therefore, in this paper, an algorithm has been applied, which is Compressive Tracking. The algorithm will further assist in determining the face position more accurately by updating the weights at each image scan. As a result, the motion noise can be reduced.

Color values of each pixel of the facial image frames are the most essential part for this experiment. Thus, it was searched a perfect Region of Interest (ROI) over the detected face. The detected face using Viola and Jones method contains some unwanted part which needs to eliminate. To identify the coordinates of the face location in the first

frame a boosted cascade classifier was used for the x and y-coordinates along with the height and width that define a box around the face. Therefore the center was selected as 60% width and 80% height of the box as the region of interest which is free from unwanted parts. Only the ROI was then separated from the entire facial and this ROI is used for further calculations.

The intensity of light at a position of the face has changed, but our eyes do not recognize it. This change in intensity is due to a change in heart rate, so the grayscale graph of the image is the heart rate graph. By taking the gray level of each color channel, one will get the raw signal. From observation, one can see that the strongest PPG signal in the green channel, since blood has the most absorption in the green part of the visible spectrum. Thus, the green channel is chosen for a very simple method. This is very similar to what Poh et al. [3] did.

3.3. Signal filtering and heart rate estimation

When environmental parameters such as temperature or external noise change, the collected RGB signals drift and become noisy. As a result, the signals must detrend. The RGB signal was detrended that uses a smoothness priors approach with a smoothing parameter of $\lambda = 10$ and a cutoff frequency of 0.059 Hz. In addition, because in the image processing process, it is difficult to avoid noise due to external influences, so one need to use 2 Butterworth and bandpass filters with low frequency of 0.4 Hz and high frequency of 3.0 Hz. This eliminates the non-heartbeat frequencies (the desired region is only $0.4 \text{ Hz} < f < 3.0 \text{ Hz}$):

In fact, the raw signal from the green channel produces the best and clearest PPG signal based on practical experience and international studies. As a result, this study processes and calculates the heart rate using the green signal. For each data segment in 256 frames which is equivalent to approximate 8.5 seconds, the Fast Fourier transform (FFT) is utilized to calculate the frequency and from there calculate the heart rate.

4. EXPERIMENTAL RESULTS

4.1. Practical implementation

In this study, the data collection, processing, and analysis of both video and video files are performed using OpenCV. The first is the software to recognize faces in each video frame and determine the desired area for each frame, the OpenCV is used to capture the position of the face. Algorithms to identify faces written on OpenCV are algorithms proposed by Viola-Jones and Lienhart-Maydt [10]. The classifier uses the Haar Cascade training set. The classifier uses a simple set of classification patterns and applies them to each desired region. The size of each desired area changes continuously to define faces with different sizes. For each defined face, the algorithm returns the coordinates along with the height and width of the face window. To avoid face segmentation errors affecting the algorithm, face coordinates from the previous frame will be used if no faces are detected.

The desired area is divided into three separate channels red, green, blue (RGB). However, only the raw signal from the green channel is extracted. Then average pixels in the desired face area is used to get the green point in each frame and form the raw signal. Based on the comments in related articles, as well as the actual verification during the

implementation of the project, the green component is chosen because this component often contains the standard PPG signal. Therefore, for simplicity and automation, the green channel is chosen as the desired source signal. After that, the bandpass filter with a low frequency is 0.4 Hz and a high frequency is 4 Hz is used to eliminate the unwanted frequencies that are not the heart rate.

Finally, the FFT is applied to select the signal source to achieve the energy spectrum. The specified heartbeat frequency is the frequency that corresponds to the highest energy level of the spectrum in a frequency range. In this study, from experiments it is observed that setting the operating frequency range in the range [0.4; 3] Hz, is equivalent to [45; 240] heart rate/minute (bpm). At the same time, the same steps are used to get the standard measurement signal from the PPG sensor to compare the correlation between the heart rate results of the two methods.

The interface includes two main sections. The first section displays the video of face position which has been tracked along with heart rate above the forehead. The second part displays the waveform graph of the raw signal, filtered signal, spectrum, and heart rate.

In addition, the software also provides an option to choose the video stream from different sources: laptop camera, phone camera (via URL), video files (.mp4). There is also an option to toggle the graphs to turn the graph on or off.

4.2. Test and evaluation results

4.2.1. Test results

In these tests, one minute video is recored to evaluate the heart rate of each subject and divide it into three time frame: 0-5 (seconds), 5-30 (seconds), and 30-60 (seconds) and then see the differences between the non-contact method (rPPG) and contact method (PPG). Several subjects were tested with different heights, weights, and health indicators. Representative experiment and results of one subject is shown in Fig. 2

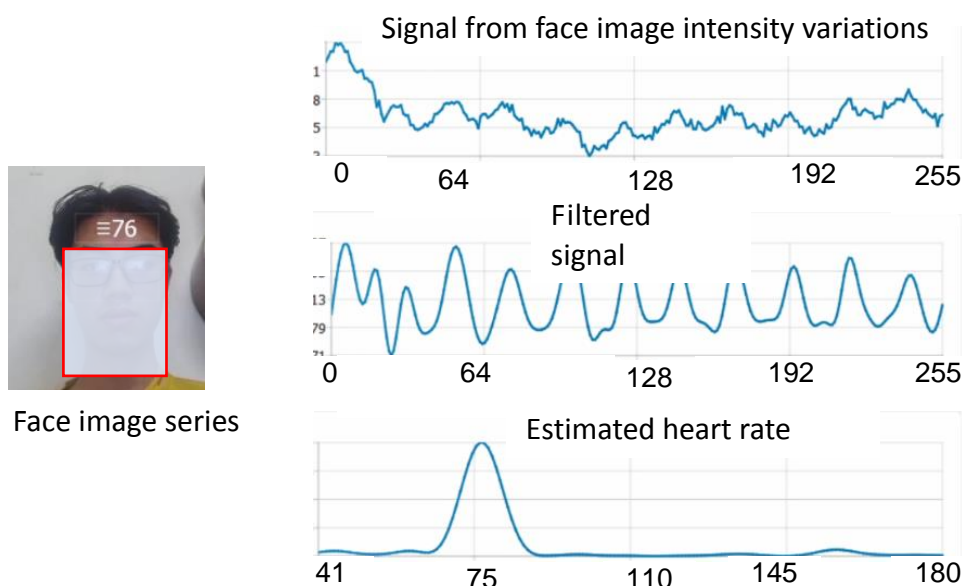


Fig. 2. Representative experiment and results on one subject.

Heart rate results for four representative subjects are given in table 1.

Table 1. Heart rates using camera and PPG devices of representative subjects.

Subjects	Time frame (second)	Average HR via camera (BPM)	Average HR via PPG (BPM)	$ \Delta HR $ (BPM)
Subject 1	0-5	61.5	55	6.5
	5-30	71.3	65.3	6
	30-60	75.3	75	0.3
Subject 2	0-5	70	62	8
	5-30	77.8	69	8.8
	30-60	84.9	84	0.9
Subject 3	0-5	43.7	53	9.3
	5-30	66.8	72	5.2
	30-60	76	74	2
Subject 4	0-5	62	70	8
	5-30	75.2	77	2.2
	30-60	77.4	77	0.4

The system can also measure heart rate of multiple faces by simultaneously measure heart rate of each individual. Fig. 3 shows the simultaneous heart rate measurements of multiple participants.

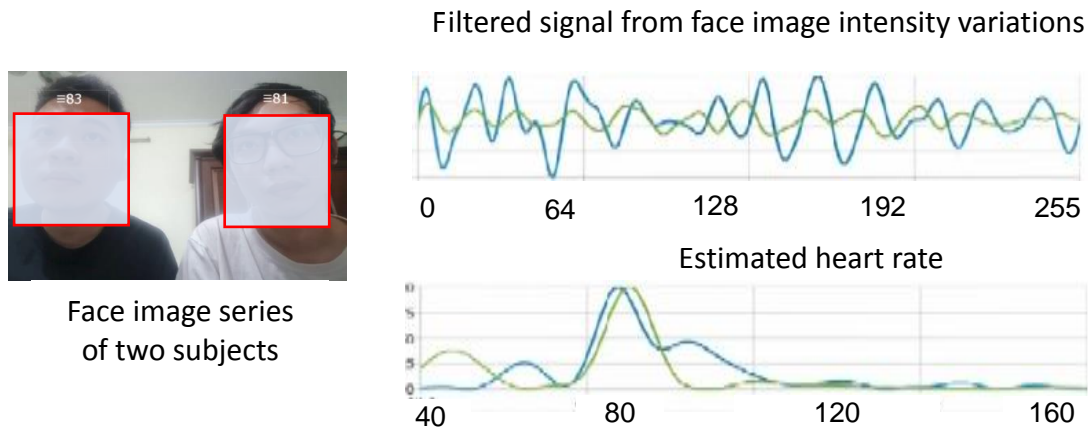


Fig. 3. Simultaneous heart rate of multiple participants.

4.2.2. Evaluation and summary

After aggregating the results from the subjects, the following table summarizes the measured data within 1 minute in table 2:

Table 2. Average heart rate by gender of 10 subjects.

Gender	Average HR via camera (BPM)	Average HR via PPG (BPM)	$ \Delta HR $ (BPM)	Health status
Males	78.4	76.8	1.6	Good
Female	72	74.4	2.4	Good

It can be seen that after 30 seconds, in static state, the heart rate begins to stable and the differences between two methods is reduce to approximate 2.2 and after one minute the heart rate of two methods are likely equal.

However, the results obtained when using the PPG measurement method will achieve very poor quality if it is in the dynamic state, whereas when using the non-contact measurement method, it will recover to a stable level in a short time. The heart rate estimated, and compared with those in the range of human heart rate (45 to 150 bpm) to recognize that the signals are from real person or fake face, for example the photo just from smartphone or printed pictures, that can be applied for anti spoofing face recognition

5. CONCLUSIONS

This study has presented an approach for conducting remote heart rate measurement using low-cost video equipment. The heart rate measurement system has also achieved good accuracy when there is no strong movement, stable processing speed, and can run in real-time. The recognition processing time can be optimized, and the process of estimating heart rate can be improved with hardware upgrade. The future development orientation is to improve the heart rate estimation process to integrate into the system as well as use deep learning algorithms in the recognition process to increase the reliability and the accuracy of the system. Advanced technique can be also applied for anti spoofing face recognition.

Acknowledgment: This research is funded by the Hanoi University of Science and Technology (HUST) under project number T2021-PC-005.

REFERENCES

- [1]. Z. Hasan, S. R. Ramamurthy, and N. Roy, "CamSense: A camera-based contact-less heart activity monitoring," *Smart Health*, vol. 23, pp. 1-13, 2022.
- [2]. J. M. Arnold, D. H. Fitchett, J. G. Howlett, E. M. Lonn, and J.-C. Tardif, "Resting heart rate: a modifiable prognostic indicator of cardiovascular risk and outcomes?," *Canadian Journal of Cardiology*, vol. 24, pp. 3A-8A, 2008.
- [3]. M. Z. Poh, D. J. McDuff, and R. U. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, pp. 10762-10774, 2010.
- [4]. D. McDuff, S. Gontarek, and R. W. Picard, "Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 2948-2954, 2014.
- [5]. C. Zhao, C. L. Lin, W. Chen, and Z. Li, "A novel framework for remote photoplethysmography pulse extraction on compressed videos" in *Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1299-1308.
- [6]. H. Demirezen and C. E. Erdem, "Remote photoplethysmography using nonlinear mode decomposition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 1060-1064.
- [7]. S. Huynh, R. K. Balan, J. Ko, and Y. Lee, "VitaMon: measuring heart rate variability using smartphone front camera," in the *17th conference on embedded networked sensor systems*, 2019, pp. 1-14.
- [8]. Q. Zhan, W. Wang, and G. Haan, "Analysis of CNN-based remote-PPG to understand limitations and sensitivities," *Biomedical Optics Express*, vol. 11, pp. 1268-1283, 2020.

- [9]. I. Pavlidis, N. L. Eberhardt, and J. Levine, "Human behavior: Seeing through the face of deception," *Nature*, vol. 415, pp. 35, 2002.
- [10]. P.A. Viola, and M.J. Jones., "Robust real-time face detection," *International Journal of Computer Vision* volume vol. 57, pp. 137–154, 2004.
- [11]. Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119–139, 1997.

TÓM TẮT

Ước lượng nhịp tim từ xa dựa trên phương pháp hấp thụ quang học

Tính toán và ước lượng nhịp tim là một bước quan trọng trong phân tích tín hiệu y sinh. Ngày nay, phương pháp tính nhịp tim từ xa dựa trên phương pháp hấp thụ quang học (rPPG) sử dụng chuỗi hình ảnh từ khuôn mặt đang phát triển mạnh mẽ. Trong nghiên cứu này, với mục tiêu là xây dựng một ứng dụng rPPG giá rẻ cho cả mục đích là ước lượng nhịp tim, và cả phát hiện khuôn mặt giả mạo nhờ vào nhịp tim đó, bài báo xây dựng một phương pháp tự động, không tiếp xúc có thể đo nhịp tim từ video và xử lý tín hiệu. Video có thể thu thập thông qua hình ảnh khuôn mặt từ camera của điện thoại thông minh. Đồng thời, nghiên cứu cũng so sánh nhịp tim đo được từ rPPG bằng hình ảnh với nhịp tim đo được từ PPG thông qua thiết bị gắn vào tay ở cùng thời điểm, và cho thấy các giá trị nhịp tim khá khớp nhau. Hệ thống đề xuất có thể ước lượng nhịp tim từ hình ảnh khuôn mặt một người hoặc nhiều người cùng một lúc, và có thể sử dụng để phát hiện sự giả mạo khuôn mặt.

Từ khóa: Ước lượng nhịp tim; Giám sát nhịp tim không tiếp xúc; Tín hiệu quang phổ-PPG; Nhận dạng khuôn mặt; Phát hiện khuôn mặt trực tiếp; Đo PPG dựa trên camera.