A Non-contact heart rate estimation framework based on photoplethysmography amplitude variation elimination and data fusion

1^{st} Arash Rasti-Meymandi
Department of Electrical Engineering
Iran University of Science and Technology
Tehran, Iran
arash.rasty.ar@gmail.com

2^{nd} Reza Karimzadeh
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
Reza.kma@ee.sharif.edu

3^{rd} Asghar Zarei
Department of Electrical Engineering
Tarbiat Modares University
Tehran, Iran
asgharzarei@modares.ac.ir

4^{th} Aboozar Ghaffari
Department of Electrical Engineering
Iran University of Science and Technology
Tehran, Iran
aboozar_ghaffari@iust.ac.ir

***Abstract*—The evaluation and assessment of Heart Rate Variability (HRV) from contact-based methods have had quite a journey throughout the annals of medical diagnosis. However, non-contact measurement of heart rate (HR) is gaining lots of attention recently. In this paper, we introduce a framework to make a robust contact-less estimation of HR from facial recorded videos. Our framework exploits various color space representations of the video frames and different techniques to extract initial Photoplethysmography (PPG) signals. The acquired PPG signals are then enhanced using the novel PPG amplitude variation elimination technique (AVET) embedded in our framework which is based on their analytic signals. Various HR is estimated from the refined PPGs in order to be fed to the fusion algorithm which is intended to procure a more accurate estimation of the true HR. The evaluation on the publicly available UBFC-Phys dataset shows that the proposed framework has a superior performance compared to the baseline techniques such as ICA and Green methods.**

***Index Terms*—Non-contact PPG,Heart Rate Variability (HRV), Heart Rate (HR), Amplitude variation elimination, Fusion.**

I. Introduction

One of the most pervasive ways of monitoring the physical and emotional status of a person is through Heart Rate (HR) evaluation. This particular physiological measurement has extensive applications such as training aid in exercising or health monitoring in hospitals. HR estimation is commonly acquired through Electrocardiograph (ECG) and contact-Photoplethysmography (cPPG) sensors. However, these methods suffer some limitations in measuring HR due to the wiring and contacted sensors themselves. Recently, a large number of related works have focused on extracting PPG from the images of a recorded video (iPPG). One of the advantages of iPPG is the acquisition of the HR for the pre-term infants in incubators without physical contact. In this application, it is common to use adhesives to attach sensors to the skin of pre-term infants which might cause pain and skin irritation. Hence, the best solution to address such an issue is through the non-contact acquisition of vital signs. The iPPG is also advantageous when it comes to minimizing the spread of a contagious disease. In this regard, procuring PPG without contacting the body has become very demanding these couple of years due to the COVID-19 pandemic situation.

Generally speaking, there are two common ways of extracting iPPG: Remote photoplethysmography (rPPG) [1]–[4], and Ballistocardiographic (BCG) [5], [6]. The first approach exploits the color fluctuation of the skin that is reflected from the face or palm; while in the latter one, the PPG is extracted from the head movement due to the strong flow of bloodstreams regulating in the head that is caused by heartbeat. In comparison, rPPG is more preferable since in most real case scenarios, the subjects have plenty of unrelated head movements that impose a strong noise on a BCG-based design. Although, it is worth mentioning that rPPG extraction has the problem of illumination variation due to the lighting condition. In addition, the relative amplitude of the true PPG is mostly small, which makes it hard to extract a smooth clean PPG from the skin. In rPPG, it is frequent to use RGB cameras in order to capture the intensity variation [7], [8]; however, to enhance the color fluctuation in frames due to the heartbeat, it is preferable to use alternative representations of RGB color model such as HSV, YUV, and YCrCb [9], [10]. Many algorithms have been proposed to acquire a steady PPG from these color space images [11], [12]. The most prevalent ones largely depend on techniques such as global or local spatial averaging [13], ICA [14], or a linear combination of color channels [4]. In [15], authors introduced a new technique to eliminate the unwanted transient motion artifacts such as acute head movement or lips movement during laughter or speaking. Their evaluation showed promising results, especially in the presence of motion artifacts. Authors in [16], used a patchedbased technique to select the best patches of each frame in order to get a stable PPG with less variability of the PPGs amplitude. However, their design was cumbersome due to utilizing the ICA technique for extracting the PPG signal in each patch.

In this work, we exploit various techniques in acquiring the PPG signal. Our framework is enhanced by considering the three-space color representations (RGB, HSV, LAB). The acquired signals from each color model are then utilized to extract multiple PPG signals using ICA and Green [17] methods. We employ a novel technique in extracting HR by eliminating the amplitude variation of the PPG signals (AVET) to better estimate the heartbeat. A novel fusion mechanism is also introduced in order to discard the outliers of the approximated HR and take a more accurate estimation of the true HR.

The rest of this paper is organized as follows. In Section II, we introduce our solution to estimate a robust HR through a series of carefully designed pipelines. Then, we provide some evaluations on the effectiveness of our proposed framework in Section III; and finally, the conclusion and some remarks are gathered in Section IV.

II. Materials and Methods

A. Database

We use the UBFC-Phys dataset [21] to evaluate our proposed framework. It is a public multimodal dataset, in which 56 participants underwent an experiment. During the experiment, subjects were sitting on a chair and filmed while they could freely move their head or blink during the three specified stages such as rest, speech, and arithmetic tasks. All subjects wore a measurement device to record their physiological activities such as blood pressure and HRV. Hence, the HR ground-truth is available for evaluation. The frame rate of the videos in this dataset is 30 Hz. We use 42 subjects from this dataset to conduct our experiments.

B. Proposed Method

Here, we elaborate on the proposed approach. The overall framework of our model is illustrated in Fig. 1. Given a video stream of the subject, we first extract the suitable Region of Interest (ROI). Then, we create two other color spaces from the original RGB. Each ROI in different color representations undergo a series of processing in order to extract multiple PPG signals. Afterward, all acquired signals go through the AVET module, which results in procuring PPG signals with invariant amplitudes. At the end of the pipeline, we apply the peak detection algorithm to extract multiple HR approximations. Then, we employ a fusion mechanism to estimate the best HR. The following is the detailed instruction of our framework implementation.

C. Raw signal extraction

After capturing the RGB images, they are converted into LAB and HSV color space representations. Then, we exploit the off-the-shelf Viola-Jones object detection technique [18] in order to restrict the ROI to the face of the subject. To further isolate the area related to the skin, we also employ the same Viola-Jones algorithm for eye detection to capture the area around the eye. This is due to the fact that eyes do not contribute to the heartbeat and only add unwanted artifacts such as eye blinking. Therefore, they are removed from the ROI. The detection procedure is only applied to the first frame of the captured video. To speed up the process of ROI extraction, we use a Correlation Tracker [19] for the rest of the frames. An illustration of the captured ROI in different color spaces is shown in Fig. 2.

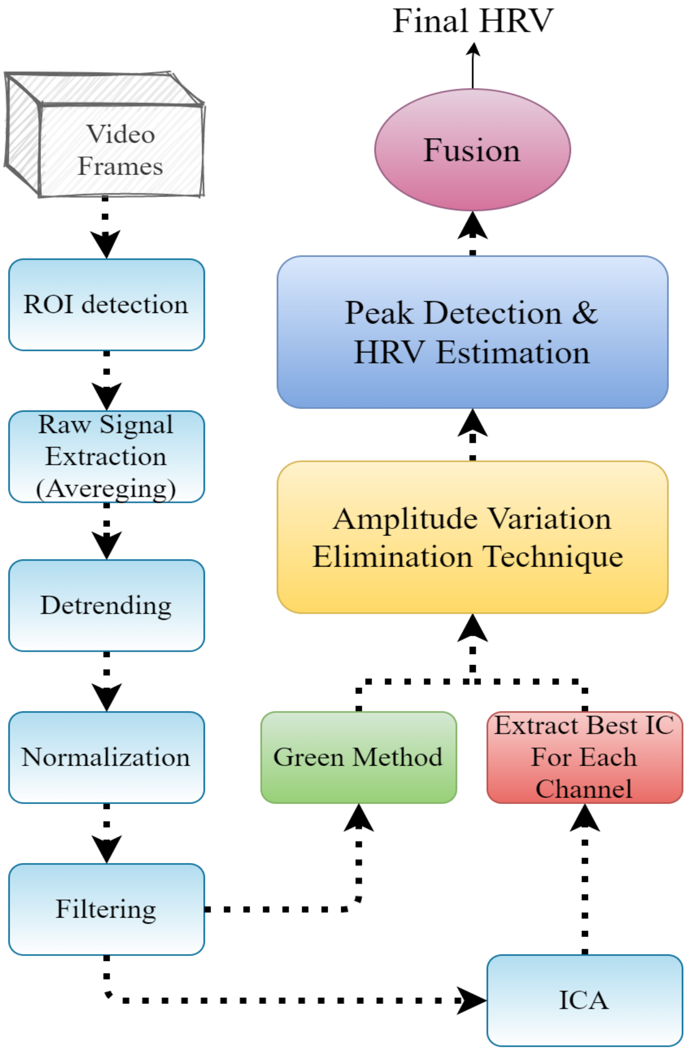
Final HRV
Video Frames
ROI detection
Raw Signal Extraction (Avereging)
Detrending
Normalization
Filtering
ICA
Green Method
Extract Best IC For Each Channel
Amplitude Variation Elimination Technique
Peak Detection & HRV Estimation
Fusion

Fig. 1: The block diagram of the proposed method.

The best and simplest way of converting the 2D ROI to a 1D raw signal is spatial averaging. This is largely owing to the fact that unwanted noises can be greatly attenuated provided that the noise is AWGN. In any case, special averaging is pretty common and it has been used extensively in the literature. Eventually, this creates nine raw signals to be used in the further procedure, three for RGB, as well as LAB, and HSV.

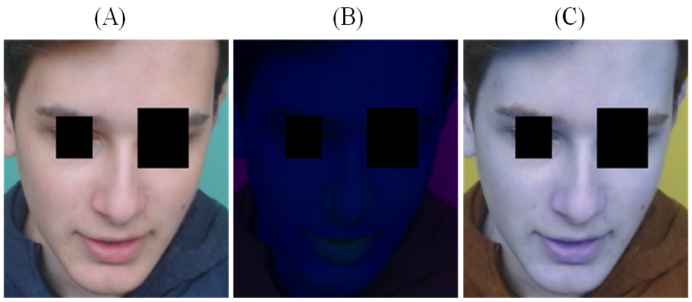

Fig. 2: The extracted ROI in various color space representations from a typical frame in the video. A) RGB, B) Lab, C) HSV.

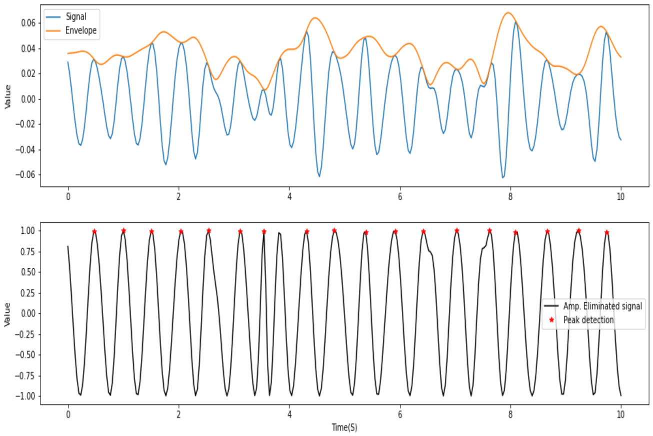

Fig. 3: An illustration of the AVET algorithm performance on eliminating the fluctuation of the PPG signals. The above figure is the PPG signal and its envelope, and the figure below is the resultant signal from the AVET algorithm.

D. Detrending and filtering

The issue with the 1D acquired signals is that they tend to exhibit a sort of undesirable trend which might not be possible to be eliminated using simple filtering. Thus, we implement an advanced detrending mechanism introduced in [20]. Fig. 4 shows the performance of the employed detrending techniques. As seen, the seemingly low frequency and its trend are perfectly captured and removed from the raw signal. Afterward, we normalize the signals to have zero mean and unity standard deviation. We then use a bandpass Butterworth filter of order five with cut-off frequency of [0.5, 3] Hz to extract the range of the heartbeat frequency and eliminate noises.

E. PPG extraction

PPG signals are extracted from the Green method and ICA technique. The Green method is simply the green channel of the RGB signal, and it has proven to be an easy yet efficient way of procuring the PPG signal. In ICA, we assume that the three signals of a color space also have three independent components in which one of which is the clean PPG signal. Although this assumption is a bit strict, it can still extract a decent PPG signal. We use FastICA to extract the independent components due to its fast performance. However, one problem with ICA techniques is the permutation issue and the fact that the suitable component corresponding to the desired PPG is not known from the three extracted ones. To tackle this problem, we first acquire the PSD of each component. We choose the component with the most energy in the range of [0.75, 2.5] Hz to be the desired PPG. This is because PPG signals tend to be periodic due to the heartbeat and therefore, have a stronger frequency component in this range compared to the other ones. In the end, four PPGs represented by $p_i(t), i = 1, \ ..., 4$ remain, one from the Green method and three from the ICA techniques used on the three color spaces.

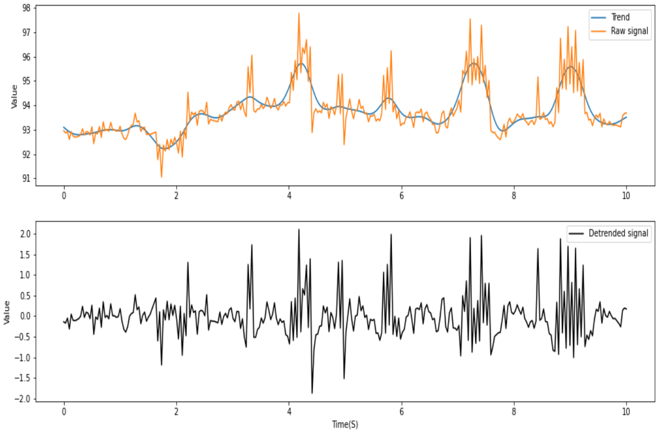

Fig. 4: An example of a raw signal (the orange line on the top panel), trend (the blue line on the top panel), and detrended signal (the black line on the bottom panel).

F. Amplitude Variation Elimination Technique (AVET)

PPG signals, in general, have the issue of amplitude variation due to various reasons from physiological viewpoints to head movement or any device-related phenomena. This characteristic often degrades the quality of the signal and sometimes causes an error in the detection of the pulses. Furthermore, the information regarding HR variation is encoded in the phase of the PPG signals. In this regard, we propose a novel amplitude variation elimination technique (AVET) to equalize the amplitude of the PPG signal. We first acquire its analytic signal using the Hilbert transform. For each extracted PPGs we can represent its analytic signal by

$$p_i^a(t) = p_i(t) + j\mathcal{H}[p_i(t)] \quad i = 1, \ ..., 4, \tag{1}$$

where $\mathcal{H}$ is the Hilbert transform. Moreover, using (1), $p_i^a(t)$ can also be expressed in the polar coordinate as

$$p_i^a(t) = p_i^m(t)e^{j\phi(t)}, \tag{2}$$

where $p_i^m(t)$ denotes the instantaneous amplitude or the envelope and $\phi(t)$ represents the instantaneous phase of $p_i(t)$. It is apparent that the envelope can be extracted by calculating the amplitude of $p_i^a(t)$, i.e., $|p_i^a(t)| = p_i^m(t)$. The equalization of the signal's amplitude is then addressed by

$$\widehat{p}_i(t) = \frac{p_i(t)}{p_i^m(t)}. \tag{3}$$

This can also be readily investigated that by doing so, the analytic signal will be $p_i^a(t) = e^{j\phi(t)}$ which has a uniform amplitude with the time-varying phase of the signal. Therefore, the phase in which the HR is encoded is the only variation in $\widehat{p}_i(t)$ and as a result, HR extraction will be facilitated. The visualization of this transformation is depicted in Fig. 3. We can see that the detection of peaks can be simpler to achieve.

G. Fusion technique

In the final stage of the proposed framework, four enhanced signals ($\widehat{p}_i(t)$) are exploited to approximate the HR. We utilize a common peak detection technique to capture the peaks. A constraint on the peaks' selection is also embedded to increase the robustness of the algorithm. We consider 180 Beat Per Minute (BPM) to be the maximum heart rate. This means that the minimum interval between two successive beats should not be less that $\frac{60}{180} \approx 0.3$ sec. Accordingly, the peak detection algorithm does not consider the second peak if the interval is less than 0.3 sec. Furthermore, a time window of 10 sec is selected to have a relative high resolution in HRV estimation. The HR in the time window is then calculated by averaging the time interval between the two consecutive peaks denoted by Δ_{avg} and using $HR = \frac{60}{\Delta_{avg}}$. We then shift the window on the refined PPG signals with the step size of 1 sec in order to capture HRV.

Here we need an outliers detection technique to better estimate HR. First, the mean and the standard deviation (std) of HR of four acquired HR is calculated. Then, a threshold will be determined to outcast those HR that are outside of the predetermined range from the average. We empirically set the threshold to be $Th = 1.2std$. Thus, if an HR is outside the range $(mean - Th, mean + Th)$ it will be disregarded. Afterward, we use the remaining HR to calculate the new average HR which is to be more accurate than the previous one. It should be noted that this technique can be generalized to situations where there are more than four estimated HR and the steps in which HRs are disregarded can be more than one.

III. EXPERIMENTAL RESULTS

In order to have a fair comparison between different techniques, we consider three metrics as follows:

$$MAE = \frac{1}{N} \sum \|HR_{est} - HR_{gt}\| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{N} \sum \left(HR_{est} - HR_{gt}\right)^2} \tag{5}$$

$$r = \frac{N \sum (HR_{est} HR_{gt}) - \sum (HR_{est}) \sum (HR_{gt})}{\sqrt{\left[N \sum HR_{est}^2 - (\sum HR_{est})^2\right]\left[N \sum HR_{gt}^2 - \left(\sum HR_{gt}\right)^2\right]}} \tag{6}$$

where HR_{est} denotes the estimated HR, HR_{gt} presents the HR ground-truth and N is the total number of subjects used in this experiment. In addition, the correlation coefficient metric is denoted by r in (6). Throughout the experiments, we compared our proposed model with the baseline techniques such as the Green method and ICA technique applied to the three color space representations.

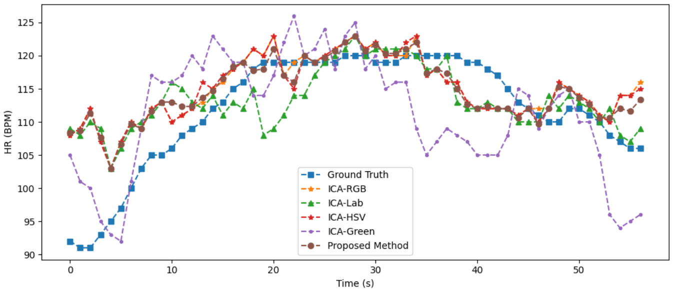

Fig. 5: Estimated HRV using different techniques including ground truth (obtained using a commercial device), the RGB channel (orange line), the LAB channel (green line), the HSV channel (red line), the Green method (purple line), and the proposed technique (brown line).

TABLE I: The evaluation of the proposed framework compared to the baseline methods based on MAE, $RMSE$, and r metrics.

Metrics	**Methods**				
	Proposed method	**ICA-RGB**	**ICA-LAB**	**ICA-HSV**	**Green**
MAE	3.02	3.46	3.13	3.50	8.27
RMSE	3.94	4.41	4.02	4.56	9.76
Correlation Coef.	0.95	0.86	0.92	0.90	0.45

A. HRV evaluation

In order to show the robustness of our proposed framework compared to the baseline methods, we calculated HR through time. Fig. 5 illustrates the estimated HRV different methods in 60 sec. It is seen that the proposed method has done a better job in following the HRV of the ground-truth compared to the Green and ICA techniques.

B. HR evaluation

For the next experiment, we have applied various techniques in HR estimation on the UBFC-Phys database. Table I shows the result of the algorithm using the MAE, $RMSE$, and the r criteria. We can see that in all metrics, the proposed algorithm achieves better performance than other methods. Fig. 6 also shows the scattergram of different methods. It is readily seen that the points are much better clustered together and closer to the red line for the our method, which indicates the robustness of the proposed algorithm. Moreover, to have a better sense of the performance of the algorithms, the boxplot of different methods is gathered and presented in Fig. 7. This figure can also show that our Fusion algorithm has a closer estimation of the HR to the ground truth than the other approaches.

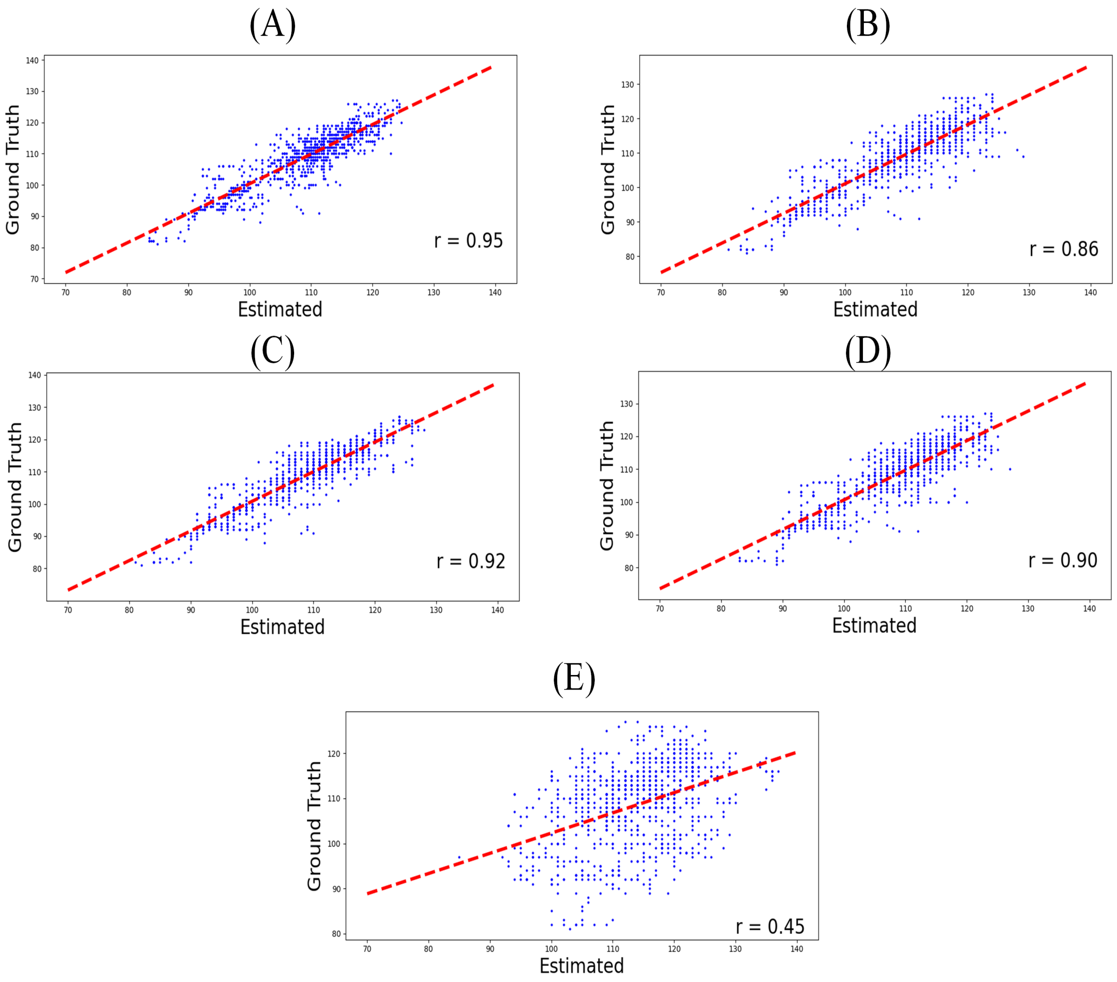

Fig. 6: The scatter plot of different methods applied to the UBFC-Phys dataset. (A) is the proposed method, (B) ICA on RGB, (C) ICA on LAB, (D) ICA on HSV, and (E) the Green method.

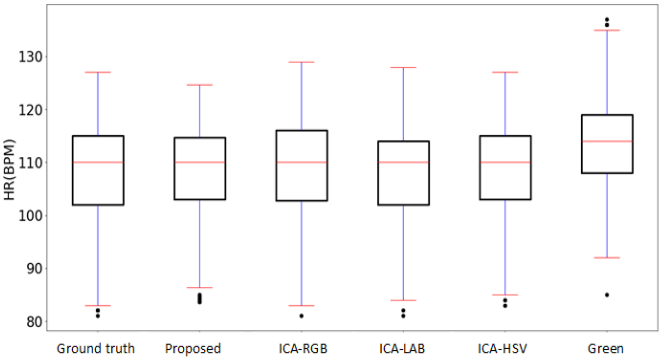

Fig. 7: The boxplot of the proposed framework and the baseline methods.

IV. CONCLUSION AND REMARKS

In this paper, we presented a robust framework for estimating heart rate (HR) from a recorded video. To this end, at first, face and eye detection and tracking algorithms were employed to detect and track our regions of interest (ROIs). Then our tracked ROIs converted to different color spaces representations (RGB, Lab, HSV) and different techniques such as detrending, normalization, filtering and ICA was exploited in order to extract preliminary PPG signal for further processes. For better approximation of HR, our framework was equipped with a novel PPG amplitude variation elimination technique to equalize the unwanted fluctuation in the PPG signal. lastly, HRV is reported by peak detection and fusion technique that discussed earlier. The evaluation of our framework and other baseline methods indicated that the proposed approach could achieve superior performance. It should be noted that the proposed scheme can be further extended to using other PPG extraction techniques to acquire more numbers of initial estimated HR. Consequently, the fusion technique can estimate a more robust HR. Another point that is worth mentioning is that the dataset on which we conducted our experiment has a lot of freedom compared to some other datasets. As a result, the Mean-Absolute-Error acquired was in the order of three which indicates that there are still rooms for improvement. Currently, we are working on efficient non-contact PPG extraction from other parts of the body such as palms or the exposed skin in general. Our future work is to extend the algorithm to capture HR from even less restricted recorded videos such as surveillance cameras where the subject is on the move. We intend to further estimate SpO2 from recorded video frames to fully exploit videos and to measure physiological metrics in a contactless manner.

Acknowledgment

This research is supported by Iran's National Elites Foundation (INEF).

References

[1] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2396–2404.

[2] W. Wang, A. C. den Brinker, S. Stuijk, and G. De Haan, "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2016.

[3] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 4264–4271.

[4] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.

[5] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3430–3437.

[6] R. Irani, K. Nasrollahi, and T. B. Moeslund, "Improved pulse detection from head motions using dct," in *2014 international conference on computer vision theory and applications (VISAPP)*, vol. 3. IEEE, 2014, pp. 118–124.

[7] W. Wang, A. C. Den Brinker, and G. De Haan, "Single-element remote-ppg," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 7, pp. 2032–2043, 2018.

[8] Q. Zhan, W. Wang, and G. de Haan, "Analysis of cnn-based remote-ppg to understand limitations and sensitivities," *Biomedical Optics Express*, vol. 11, no. 3, pp. 1268–1283, 2020.

[9] X. Niu, S. Shan, H. Han, and X. Chen, "Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation," *IEEE Transactions on Image Processing*, vol. 29, pp. 2409–2423, 2019.

[10] C.-k. Park and H.-j. Choi, "Effective methods to extract ppg signals from face using stochastic state space modeling approach," in *2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. IEEE, 2015, pp. 172–173.

[11] H. Ghanadian and H. Al Osman, "Non-contact heart rate monitoring using multiple rgb cameras," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2019, pp. 85–95.

[12] C. Park and H.-j. Choi, "Motion artifact reduction in ppg signals from face: Face tracking & stochastic state space modeling approach," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 3280–3283.

[13] A. Lam and Y. Kuno, "Robust heart rate measurement from video using select random patches," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3640–3648.

[14] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation." *Optics Express*, vol. 18, no. 10, pp. 10 762–10 774, 2010.

[15] Z. Yang, X. Yang, J. Jin, and X. Wu, "Motion-resistant heart rate measurement from face videos using patch-based fusion," *Signal, Image and Video Processing*, vol. 13, no. 3, pp. 423–430, 2019.

[16] Y. Maki, Y. Monno, K. Yoshizaki, M. Tanaka, and M. Okutomi, "Inter-beat interval estimation from facial video based on reliability of bvp signals," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 6525–6528.

[17] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light." *Optics express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

[19] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *British Machine Vision Conference, Nottingham, September 1-5, 2014*. Bmva Press, 2014.

[20] M. P. Tarvainen, P. O. Ranta-Aho, and P. A. Karjalainen, "An advanced detrending method with application to hrv analysis," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 2, pp. 172–175, 2002.

[21] R. Meziatisabour, Y. Benezeth, P. De Oliveira, J. Chappe, and F. Yang, "Ubfc-phys: A multimodal database for psychophysiological studies of social stress," *IEEE Transactions on Affective Computing*, 2021.