III. International Conference on Electrical, Computer and Energy Technologies (ICECET 2023)
16-17 November 2023, Cape Town-South Africa

Comparative Study of Face Tracking Algorithms for Remote Photoplethysmography

J.A.S.Y. Jayasinghe[1], Stamos Katsigiannis[2], and Lakmini Malasinghe[3]

[1,3]Department of Electrical and Electronic Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka
[2]Department of Computer Science, Durham University, Stockton Road, Durham, DH1 3LE, UK
Emails: [1]shehanijay@gmail.com, [2]stamos.katsigiannis@durham.ac.uk, [3]lakmini.m@sliit.lk

***Abstract*—Remote Photoplethysmography (rPPG) is a non-invasive approach for monitoring Heart Rate (HR) that can be used in various applications in healthcare and biometrics. rPPG measurements acquired using facial videos have become very popular and one of the main steps of this technique is facial tracking and Region of Interest (ROI) extraction. This research paper investigates four widely used face tracking algorithms, namely MediaPipe Face Mesh (MPFM), Haar Cascade, Multi-task Cascaded Convolutional Network (MTCNN), and Dlib, concerning their ROI extraction capabilities for rPPG HR measurements. Using evaluation metrics such as accuracy, processing time, and ease of extracting ROIs, this work also recommends the most suitable face tracking algorithm from those mentioned above for rPPG measurements, and presents a compilation of a prioritization list of ROIs based on their sensitivities for rPPG measurements. Experimental results showed that the MPFM algorithm and cheek ROIs provided the best measurements of HR.**

***Index Terms*—remote Photoplethysmography, Face Tracking, ROI Extraction, Machine Learning.**

I. Introduction

Heart Rate (HR), which is the number of times that a heart beats per minute, is a vital indicator of many underlying health conditions, and it is easily accessible and contains valuable prognostic information. HR has been used to detect and predict numerous cardiovascular diseases and to get a sense of a person's psychophysiology as well as in behavioral medicine [1]. An approach developed from Photoplethysmography, known by different names including remote Photoplethysmography (rPPG) and imaging Photoplethysmography (iPPG) utilizes a video feed obtained using a camera to analyze the amount of ambient light absorbed by the skin [2]. This absorbed amount of light has been proven to be proportional to the changes in blood volume passing through blood vessels beneath the skin. This approach can be applied in situations where physical contact with a person is undesirable. The most reliable HR measurements have been obtained from the signals extracted from facial videos [3]–[7]. Once videos are acquired, the face of the person should be detected and tracked so that the signal obtained across all of the frames in the video maps to the same Region of Interest (ROI) of the face. It has been found that different regions of the skin have different degrees of sensitivity related to light absorption [8], [9]. This indicates that rPPG requires preserving continuity from frame to frame. This means the tracked region must remain constant across frames, with no additional or reduced information, ensuring that noise is kept at a minimum once spatial averaging for the ROIs is conducted. Therefore, the efficiency and performance of the face detection and tracking algorithm is very vital. There have been many rPPG research studies conducted by utilizing various face tracking algorithms, but due to the unavailability of standardization, the performance of these algorithms needs to be evaluated and compared to find the best approach for rPPG measurements concerning different scenarios and environments.

Rendon et al. [10] have discussed the usage of the Haar Cascade Classifier for face detection through which a signal extracted from the forehead ROI has been utilized to calculate HR. This research is limited to data obtained from the forehead. Nikolaiev et al. [11] and Monsalve et al. [12] use the whole face alongside skin detection to obtain the areas of the skin visible for HR calculations. However, light absorption varies with different parts of the skin [8], [9]. This means that different ROIs, such as cheeks, chin and forehead, have varying degrees of sensitivity related to absorption of light. Therefore, when spatial averaging is applied for the whole face, the area taken into consideration is not precise and the larger area of skin results in more noise in the signal extracted. Hence, the most visible and sensitive ROIs should be reliably selected and extracted.

Nikolaiev et al. [11] used the Haar Cascade Classifier as a means of calculating the stress index using Heart Rate Variability as well as the time duration of a heartbeat, while Chang et al. [13] used it to obtain rPPG measurements for physiological signal feedback control in fitness training. By combining Haar Cascade and Dlib toolkit to detect face bounding boxes and to obtain pre-trained 68 facial landmarks to crop out selected ROI regions respectively, Ma et al. [14] implemented a rPPG system with the addition of an FIR Band-Pass filter. Smelyakov et al. [15] and Zhang et al. [16] compared MTCNN with YOLO and MTCNN with YOLOv3 respectively and found that MTCNN was more accurate. Smelyakov et al. [17] developed their model further by comparing MTCNN, YOLOv3, Dlib, and Retinaface and found MTCNN and Reti-

979-8-3503-2781-6/23/$31.00

naface were more accurate than Dlib and YOLOv3. One of the drawbacks of MTCNN and Retinaface is that they only contain 5 facial landmarks as opposed to 68 facial landmarks in Dlib [18]. But it was found that Dlib has a longer processing time than MTCNN [18]. Kaiser et al. [19] compared Dlib and MTCNN and the results showed that Dlib was more accurate. This was in contrast to the outcome obtained by Smelyakov et al. [17] as mentioned earlier. Hence, the outcomes of these past researches are inconsistent and show varying degrees of accuracy. They have also been compared with reference to facial detection and not specifically for rPPG measurements. Therefore, there is a need for better performance comparisons between some of the most utilized facial detection algorithms for HR measurements. Although Haar Cascade Classifiers have been used frequently for rPPG measurements, novel approaches such as MediaPipe Face Mesh, MTCNN, and Dlib toolkit have not been thoroughly examined. Hence, this paper discusses the comparison between the aforementioned facial detection/tracking algorithms for rPPG applications.

ROIs used for signal extraction during the past research studies include Full Face [20]–[22]; facebox detection followed by skin segmentation, Palm [23] ; Palm detection followed by skin detection, Cheeks [24]–[27]; ROI extraction using facial landmarks, Forehead [24], [25], [27]; ROI extraction using facial landmarks and Adaptive ROI [28], [29]; Finding the Signal to Noise ratio and dropping pixels with lower ratios. During real-world scenarios, a specific ROI cannot be predefined assuming there will be no obstructions in this fixed region or this predefined area will always be visible. Hence, in case one ROI is not visible, a different ROI should be utilized to obtain the HR. This has not been implemented in past research and they have not evaluated the accuracy of the signals obtained from different ROIs to find the order of sensitivities in order to find the most suitable ROI in practical scenarios. Therefore, it is essential that for further research in dynamic ROI selection, the different ROIs should be given priorities in order of their sensitivities.

This research paper analyzes four methods of facial tracking and compares different ROIs for HR calculation. The contributions of this paper include the following;

- ROI extraction for MediaPipe Face Mesh, Haar Cascade, MTCNN and Dlib face tracking algorithms.
- Recommendation of the most suitable face tracking algorithm for rPPG HR measurements through the analysis of accuracy, processing time and ease of extracting ROIs.
- Compilation of a prioritization list of different ROIs by evaluating their sensitivities for rPPG measurements.

The rest of this paper is organised in three sections. Section II describes the steps in the methodology. Section III presents and analyses our experimental results, whereas Section IV recounts the main techniques and results of this work.

II. Methodology

This section describes the steps and techniques used for HR calculation. The steps include video acquisition, facial tracking, ROI extraction, spatial averaging, signal processing, spectral analysis, machine learning, and HR calculation. Fig. 1 shows a block diagram of the methodology. The following subsections discuss the above steps in detail.

A. Video Acquisition

This experiment was conducted using a publicly available dataset which included continuous facial recordings and the respective ground truth HR measurements of the subjects. The dataset used was the UBFC-RPPG dataset [20]. UBFC-RPPG contains video recordings and ground truth heart rate measurements of 42 subjects. The videos are in uncompressed 8-bit RGB format and have been acquired using a low-cost Logitech C920 HD pro webcam at 30 frames per second with a resolution of 640×480. The ground truth measurements have been obtained using the Contec Medical CMS50E pulse oximeter finger clip sensor [20]. The data has been obtained from subjects that are stationary under constant lighting conditions.

B. Face tracking and ROI extraction

In this work, four face detection and tracking algorithms have been compared, namely, MediaPipe Face Mesh, Harr Cascade, Dlib, and MTCNN:

1) MediaPipe Face Mesh (MPFM): MPFM utilizes a machine learning algorithm that can estimate the position of 468 3D facial landmarks in real time [30]. Fig. 2 shows the facial landmarks obtained using MPFM, while Fig. 3a shows the extracted ROIs of the subject. As shown in Fig. 3a, the ROIs were extracted by using the coordinates of the landmark points for the forehead, chin, right cheek, and left cheek. These points were used to create a mask to perform bitwise AND operation in order to obtain the area inside the ROIs which were then cropped using a bounding box. Since MPFM provides 468 landmark points, another two ROIs were considered for the two cheeks, as shown in Fig. 3b.

2) Haar feature-based Cascade Classifier: The Haar-like features serve as the main building block for Haar classifier object detection. These features are extracted using the Viola Jones algorithm [31] which uses the difference in contrast between neighboring rectangular groupings of pixels instead of intensity values of individual pixels. Relatively light and dark areas are identified using the contrast discrepancies between pixel groups [32]. A Haar-like feature is formed by two or three neighboring groups with a relative contrast variance. Using a combination of these features, a face can be detected [33]. After using AdaBoost algorithm to train predictors sequentially, each trying to correct its predecessor, the following stage chooses the features with the lowest error rate because they are the features that most effectively distinguish a face from a non-face, while other features are rejected. OpenCV provides a training method as well as pre-trained models that can be used as Haar Cascade classifiers. The application of this face detection model is shown in Fig. 4 where the image of a subject is given on the left and the cropped out face detected by the Haar Cascade classifier is given on the right. By using proportional values inside the

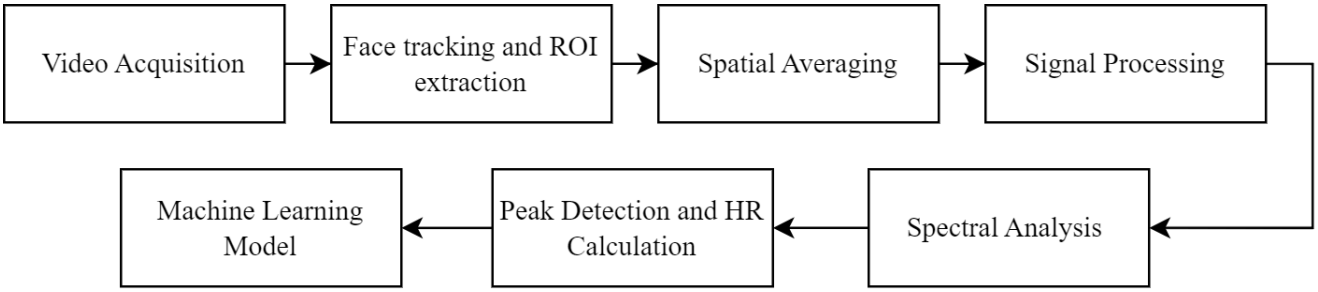


Fig. 1: Block Diagram of the Methodology

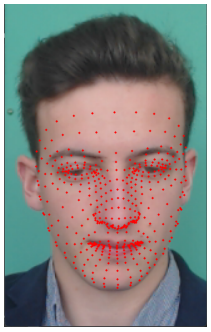
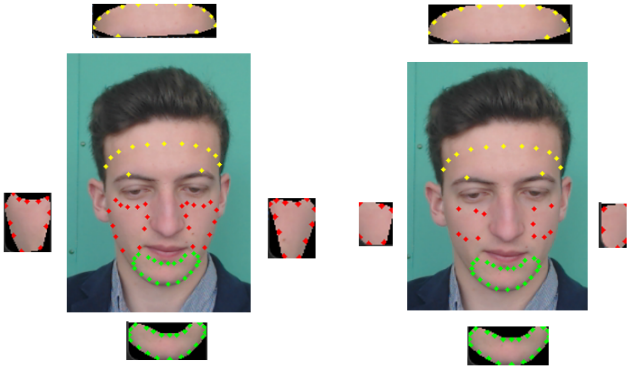


Fig. 2: MediaPipe Face Mesh Facial Landmarks.



(a) (b)

Fig. 3: (a) First set of ROIs extracted using Face Mesh. (b) Second set of ROIs extracted using Face Mesh.

bounding box of the face, 4 coordinates for each ROI were obtained to extract them as shown in Fig. 5.

Fig. 4: Face detected through Haar Cascade Classifier.

3) Multi-task Cascaded Convolution Networks (MTCNN): MTCNN is a network architecture built by cascading multiple tasks of a CNN. The positions of the faces and facial

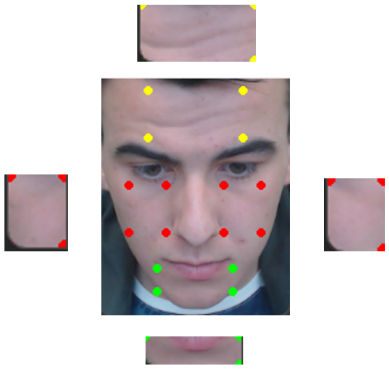
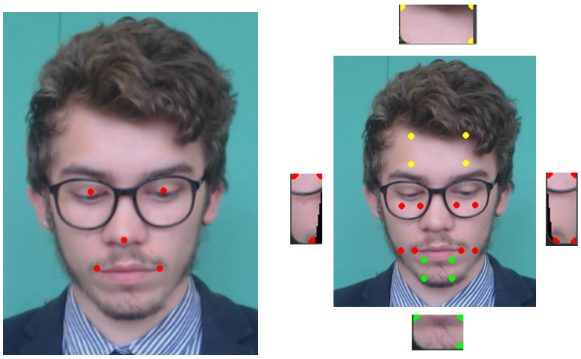


Fig. 5: ROIs extracted using Haar Cascade Classifier.

landmarks are obtained using a cascaded three-layer network structure that consists of three stages, namely Proposal Network (P-Net), Refine Network (R-Net), and Output Network (O-Net) [34]. P-Net utilizes an image pyramid with the original image being resized and is used to extract the preliminary features [16] as well as the corresponding vectors which contain bounding box regression information [35]. In R-Net layer, calibration based on bounding box regression and NMS takes place while in order to choose numerous groups of locally optimal face candidate frames, elimination of the face candidate frames with poor scores also occurs [16]. O-Net layer is similar to R-Net layer but it selects the best candidate frames and outputs five facial landmark points [16]. Fig. 6a shows the landmark points obtained for a sample taken from the UBFC-RPPG dataset. These points were then used to obtain approximate coordinate points for the forehead, chin, left cheek, and right cheek, and these ROIs were extracted as shown in Fig. 6b.

4) Dlib: Dlib is a cutting-edge C++ toolkit that includes machine learning techniques and tools for developing sophisticated software to address real-world issues. The Dlib library consist of four components namely Bayesian Nets, Linear Algebra, Optimization and Machine Learning Tools [36]. Dlib contains a facial landmark detector that utilizes pre-trained models for face detection using 68 facial landmarks corresponding to the outer edge of the chin, cheeks, eyebrows, eyes, the outline of the nose, and the inner and outer edges of the mouth. Fig. 7a shows the landmarks detected on a sample video from the dataset, whereas Fig. 7b depicts the landmark points considered for each of the forehead, chin, left cheek, and right cheek and the extracted ROIs.



(a) (b)

Fig. 6: (a) MTCNN Facial Landmarks. (b) ROIs extracted using MTCNN.

(a) (b)

Fig. 7: (a) Dlib Facial Landmarks. (b) ROIs extracted using Dlib.

C. Spatial Averaging

Once the ROIs were extracted, the average pixel values of the ROIs in each frame were recorded. For the videos in the UBFC-RPPG dataset, a 30-second window corresponds to 900 frames. This calculation is shown in Eq. (1) where WS = window size, FPS = Frames Per Second, and WT = window duration in seconds.

$$WS = FPS \times WT \tag{1}$$

By considering all of the frames corresponding to the last 30-second window, the mean and the standard deviation of these average pixel values were calculated in order to obtain the normalized pixel values of each frame. Equation (2) shows the calculation of the normalized pixel values.

$$normalized = \frac{window - mean}{std} \tag{2}$$

D. Signal Processing

After obtaining the normalized pixel values of the ROIs, Independent Component Analysis was applied to reduce the Signal to Noise Ratio (SNR) by separating independent signals from a linear mixture of underlying sources. For this experiment, ICA was implemented using the FastICA algorithm from the scikit-learn machine learning library. Here it was assumed that the signal is formed of a mixture of three source signals (RGB) and the 2nd component is said to contain the most information regarding the HR of a person [37]. Malasinghe et al. [38] has shown that when using a three-channel video, green channel offers the best HR measurements. Therefore, the green channel source signal was used for the HR calculation. The range of heart rates considered in this research are in between 50bpm to 240 bpm. These corresponds to frequencies 0.833Hz and 4Hz. Hence, the source signal obtained through ICA is then filtered using a Butterworth band-pass filter having a lower cutoff frequency of 0.8Hz and an upper cutoff frequency of 4Hz to remove any unnecessary frequencies that falls out of range of the heart rates considered.

E. Spectral Analysis and HR Calculation

Once the filtered signal is obtained, Fast Fourier Transform (FFT) is applied to generate the power spectrum. The heart rate range considered in this research is 50bpm to 240bpm, and earlier research has shown that the frequency corresponding to the maximum power of the spectrum corresponds to the heart rate in beats per second (bps) at a given time [37]. Therefore by multiplying this frequency value (in Hz) by 60 s, the HR in beats per minute (bpm) was calculated for each second.

F. Machine Learning Model

The raw HR values obtained from the spectral analysis and the corresponding ground truth values were used to train a machine learning algorithm to determine the most accurate and precise HR value. The ML algorithm used for this experiment is the decision tree. The raw HR values of all ROIs considered, for each of the MPFM, Haar Cascade, MTCNN, and Dlib methods, as well as the ground truth HR values, were then split into training data and test data according to a 4:1 ratio. The training data were used to train the decision tree regression model after which the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) of each face tracking algorithm were used as the evaluation metrics for comparison.

Equation (3) shows the calculation of the Root Mean Squared Error (RMSE), where N is the sample size, $\hat{y}_i$ is the predicted value for the i^{th} observation, and y_i is the actual value for the i^{th} observation.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{3}$$

Equation (4) shows the calculation of the Mean Absolute Error where N is the sample size, x_i is the actual value for the i^{th} observation, and y_i is the predicted value for the i^{th} observation.

$$MAE = \sum_{i=1}^{N}|x_i - y_i| \tag{4}$$

The HR results obtained through the four face tracking algorithms have been analysed in the following section.

III. Results and Discussion

The results obtained through the processing steps are depicted in Fig. 8, Fig. 9, and Fig. 10), by using outcomes from the MPFM algorithm applied for the left cheek ROI of one

subject from the UBFC-RPPG dataset. The plot of the mean green pixel values calculated using the MPFM algorithm for the left cheek ROI is shown in Fig. 8. The power spectra obtained for the unfiltered signal as well as the band-pass filtered signal are shown in Fig 9. Here, it can be seen that the frequencies lower than 0.8Hz and higher than 4Hz have been filtered out while the frequencies between these two thresholds have been allowed to pass. The maximum power (shown in the legend of the graph) of the unfiltered signal can be seen as 85.647 dB and corresponds to a frequency of 0.033 Hz. This would indicate that at the moment, the subject has a HR 1.98 bpm, which is incorrect. But, once the Butter-worth band-pass filter is applied, the maximum power equals 40.245 dB and corresponds to a frequency of 1.833 Hz. This shows that the HR is around 109.98 bpm which was approximately equal to the ground truth value of 110 bpm. This conveys that noise from external illumination sources has been filtered out through the band-pass filter. Fig. 10 shows the plot of heart rates of the subject obtained from the left cheek using the MPFM algorithm.

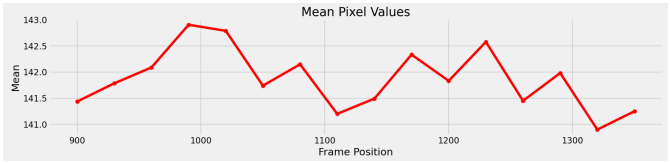


Fig. 8: Plot of Mean pixel values and Calculated Heart Rates

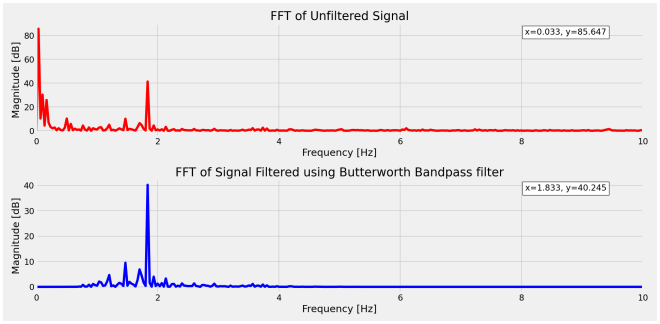


Fig. 9: Power spectra of unfiltered and filtered signal

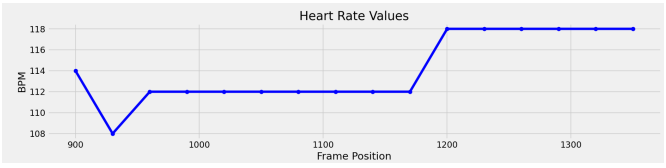


Fig. 10: Plot of calculated heart rates

A. Performance Comparison of Face tracking Algorithms

The performance of the four face tracking algorithms have been discussed in terms of accuracy, processing time, and ease of ROI selection in the following subsections.

1) Accuracy: The RMSE and MAE values of the four algorithms are illustrated in TABLE I. For MPFM, the RMSE and MAE for both set of ROIs are given in the table. These metrics were calculated using the training and test data used for the ML model which included data from the whole dataset. The last two columns of the table have been added to include the RMSE and MAE averages between the two cheeks. Here it can be seen that the lowest RMSE and MAE corresponds to the MPFM algorithm followed by Dlib, MTCNN and Haar Casade. This conveys that for rPPG HR measurements, among the considered face tracking algorithms, MediaPipe Face Mesh offers the best accuracy. The table also illustrates how the bandpass filter (BPF) has improved the HR measurements.

2) Processing time: TABLE II provides the processing times (on an Intel Core i7-10750H CPU @ 2.60GHz with 16 GB RAM) of each of the four face tracking algorithms for HR calculations. These were calculated using a video containing 1547 frames taken from the dataset. Here it can be seen that MPFM took the lowest time to process the video, followed by Haar Cascade, while Dlib took the longest, followed by MTCNN. Hence, it can be concluded that Dlib and MTCNN are not suitable for real-time applications of rPPG HR calculations.

3) Ease of ROI selection: By referring to Fig. 2, Fig. 6a, and Fig. 7a, it can be seen that MPFM offers the highest number of options for ROI selection followed by Dlib and MTCNN. This is due to the higher number of landmark points available for drawing ROIs. The higher number of facial landmark points recognized by MPFM allows more ROI options and precise ROI locations even when the face is angled. This can be seen in Fig. 11. Even in situations where the face is obstructed, the higher number of landmark points enables dynamic selection of points for ROI selection.


Fig. 11: Angled face recognized by Face Mesh

Even though Haar Cascade has a lower processing time, a considerable flaw of Haar Cascade is the unreliability of the algorithm in the presence of patterns that may seem similar to Haar features. This can be seen in Fig. 12 where the pattern in the subject's clothes have been recognized as Haar features and hence, the facebox extracted is inaccurate and has led to incorrect ROI extraction. Another drawback is that Haar Cascade does not provide landmark points. The process of ROI selection in the Haar Cascade algorithm depends solely on a calculation using the dimensions of the facebox. Hence, if the face is angled as shown in Fig. 13a, the ROIs extracted would be incorrect. However as can be seen in Fig. 13b, MPFM is much better for such instances.

By taking into consideration the accuracy, the processing time and the number of landmark pints of the four face tracking algorithms, it can be concluded that MediaPipe MPFM is the most suitable approach for accurate and fast real-time HR monitoring.

B. Prioritization list of ROIs

By analyzing the results given in TABLE I, when considering the different ROIs, it can be concluded that the most

TABLE I: RMSE and MAE of different ROIs corresponding to each face tracking algorithm

Face Tracking Algorithm	ROI	RMSE		MAE		Average RMSE With BPF	Average MAE With BPF
		Without BPF	With BPF	Without BPF	With BPF		
MediaPipe Face Mesh	Forehead	11.4436	9.1415	7.9628	6.3681	9.1415	6.3681
	Chin	13.9826	11.6099	10.2766	8.2552	11.6099	8.2552
	Left Cheek ROI 1	11.3001	9.3603	7.8576	6.4347	10.5319	7.4145
	Right Cheek ROI 1	14.0226	11.7035	10.3537	8.3943		
	Left Cheek ROI 2	8.4206	6.488	5.5881	4.617	6.14565	4.3931
	Right Cheek ROI 2	6.789	5.8033	4.8667	4.1692		
Haar Cascade	Forehead	12.1799	11.4241	8.8769	8.974	11.4241	8.974
	Chin	14.3916	13.3531	10.9268	10.1685	13.3531	10.1685
	Left Cheek	12.3403	10.9737	8.8667	7.9054	10.46215	7.41745
	Right Cheek	11.1409	9.9506	7.9267	6.9295		
MTCNN	Forehead	14.0401	12.5782	10.7017	9.0973	12.5782	9.0973
	Chin	14.901	11.9909	11.2184	8.8844	11.9909	8.8844
	Left Cheek	8.5034	6.2706	5.7688	4.3256	7.86175	5.41645
	Right Cheek	11.5042	9.4529	8.0674	6.5073		
Dlib	Forehead	9.9025	8.7857	6.4888	6.4493	8.7857	6.4493
	Chin	15.5086	12.4773	11.7919	9.0823	12.4773	9.0823
	Left Cheek	11.1143	8.5824	7.3184	5.9809	7.4954	5.1639
	Right Cheek	9.0459	6.4084	5.7593	4.3469		

TABLE II: Processing times of each face tracking algorithm

Face Tracking Algorithm	Processing time	Frames processed per second
Face Mesh with ROI 1	00:22.8924	67.5770125
Face Mesh with ROI 2	00:19.6821	78.5993365
Haar Cascade	00:54.7827	28.2388418
MTCNN	14:06.8707	1.82672514
Dlib	36:51.9116	0.69939504

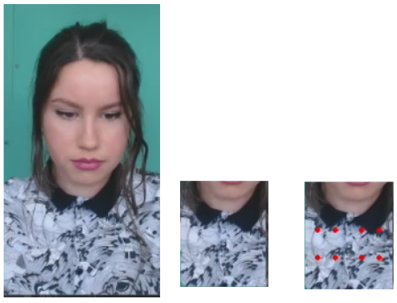
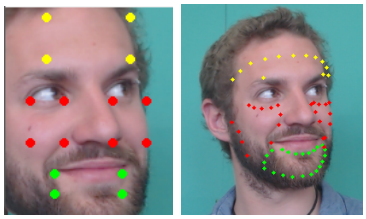


Fig. 12: Pattern in subject's clothes detected as Haar Features

sensitive ROIs are the cheeks followed by the forehead and the chin. In the case of the Face Mesh algorithm, the smaller ROIs performed better than the larger areas. This may have occurred owing to the lower amount of noise from motion artifacts and external illumination sources in the smaller area. It can be seen that except for MTCNN, the right cheek offered better HR measurements than the left cheek. This could also be due to external illumination conditions that affect the two sides of the face to different degrees. This order of sensitivity can be used to give precedence to the ROIs with higher sensitivity when only a part of the face is visible or when obstructions are present.



(a) HaarCascade (b) Face Mesh

Fig. 13: ROIs of an angled face recognized by Haar Cascade and Face Mesh

IV. Conclusion

In this research, we investigated and compared different face tracking algorithms, namely MediaPipe Face Mesh, Haar Cascade, MTCNN, and Dlib, for rPPG heart rate measurements with reference to their ROI extraction methods. We also analyzed their accuracy, processing time and ease of ROI selection along with the sensitivities of ROIs for HR measurements.

Through the results obtained, it can be concluded that the highest accuracy and lowest processing time are provided by MPFM. When compared to Haar Cascade, Dlib and MTCNN, MPFM provides the most flexibility when choosing a ROI because it is possible to select ROIs using a larger choice of landmark points. Since MPFM can identify a greater number of landmarks on the face, it can pinpoint the location of a ROI regardless of the angle at which the face is seen. The increased number of landmark points permits dynamic selection of points for ROI selection even when the face is obscured. We conclude that MediaPipe Face Mesh is the optimal method for precise and quick real-time HR monitoring based on its superior performance across all three metrics compared to the other three face tracking algorithms. Based on the data, the cheeks are the most effective ROI for extracting rPPG readings, followed by the forehead and the chin. This discovery can be utilized to prioritize these ROIs when only a portion of the face is visible or when there are impediments in the way. With these contributions, we believe that this research gave deep insight to methods and techniques that can further improve rPPG HR monitoring systems.

Acknowledgment

This research was supported by the Special Funding for Research leading to Mphil and PhD Degrees offered by Faculty of Graduate Studies and Research, Sri Lanka Institute of Information Technology.

REFERENCES

[1] G. G. Berntson, J. Thomas Bigger Jr, D. L. Eckberg, P. Grossman, P. G. Kaufmann, M. Malik, H. N. Nagaraja, S. W. Porges, J. P. Saul, P. H. Stone *et al.*, "Heart rate variability: origins, methods, and interpretive caveats," *Psychophysiology*, vol. 34, no. 6, pp. 623–648, 1997.

[2] A. Al-Naji, K. Gibson, S.-H. Lee, and J. Chahl, "Monitoring of cardiorespiratory signal: Principles of remote measurements and review of methods," *IEEE Access*, vol. 5, pp. 15 776–15 790, 2017.

[3] B. Lokendra and G. Puneet, "And-rppg: A novel denoising-rppg network for improving remote heart rate estimation," *Computers in biology and medicine*, vol. 141, p. 105146, 2022.

[4] X. Niu, H. Han, S. Shan, and X. Chen, "Continuous heart rate measurement from face: A robust rppg approach with distribution learning," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 642–650.

[5] R. Song, S. Zhang, C. Li, Y. Zhang, J. Cheng, and X. Chen, "Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7411–7421, 2020.

[6] K. M. van der Kooij and M. Naber, "An open-source remote heart rate imaging method with practical apparatus and algorithms," *Behavior research methods*, vol. 51, pp. 2106–2119, 2019.

[7] Z. Yu, W. Peng, X. Li, X. Hong, and G. Zhao, "Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 151–160.

[8] 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.

[9] M. Huelsbusch and V. Blazek, "Contactless mapping of rhythmical phenomena in tissue perfusion using ppgi," in *Medical Imaging 2002: Physiology and Function from Multidimensional Images*, vol. 4683. SPIE, 2002, pp. 110–117.

[10] J. R. Maestre-Rendon, T. A. Rivera-Roman, A. A. Fernandez-Jaramillo, N. E. Guerron Paredes, and J. J. Serrano Olmedo, "A non-contact photoplethysmography technique for the estimation of heart rate via smartphone," *Applied Sciences*, vol. 10, no. 1, p. 154, 2019.

[11] S. Nikolaiev, S. Telenyk, and Y. Tymoshenko, "Non-contact video-based remote photoplethysmography for human stress detection," *Journal of Automation, Mobile Robotics and Intelligent Systems*, pp. 63–73, 2020.

[12] D. Botina-Monsalve, Y. Benezeth, R. Macwan, P. Pierrart, F. Parra, K. Nakamura, R. Gomez, and J. Miteran, "Long short-term memory deep-filter in remote photoplethysmography," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 306–307.

[13] C.-M. Chang, C.-C. Hung, C. Zhao, C.-L. Lin, and B.-Y. Hsu, "Learning-based remote photoplethysmography for physiological signal feedback control in fitness training," in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2020, pp. 1663–1668.

[14] X. Ma, D. P. Tobón, and A. El Saddik, "Remote photoplethysmography (rppg) for contactless heart rate monitoring using a single monochrome and color camera," in *International Conference on Smart Multimedia*. Springer, 2019, pp. 248–262.

[15] K. Smelyakov, A. Chupryna, O. Bohomolov, and I. Ruban, "The neural network technologies effectiveness for face detection," in *2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*. IEEE, 2020, pp. 201–205.

[16] N. Zhang, J. Luo, and W. Gao, "Research on face detection technology based on mtcnn," in *2020 international conference on computer network, electronic and automation (ICCNEA)*. IEEE, 2020, pp. 154–158.

[17] K. Smelyakov, A. Chupryna, O. Bohomolov, and N. Hunko, "The neural network models effectiveness for face detection and face recognition," in *2021 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*. IEEE, 2021, pp. 1–7.

[18] H. Kim, H. Kim, and E. Hwang, "Real-time facial feature extraction scheme using cascaded networks," in *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2019, pp. 1–7.

[19] A. N. Zereen, S. Corraya, M. N. Dailey, and M. Ekpanyapong, "Two-stage facial mask detection model for indoor environments," in *Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020*. Springer, 2021, pp. 591–601.

[20] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recognition Letters*, vol. 124, pp. 82–90, 2019.

[21] M. Artemyev, M. Churikova, M. Grinenko, and O. Perepelkina, "Robust algorithm for remote photoplethysmography in realistic conditions," *Digital Signal Processing*, vol. 104, p. 102737, 2020.

[22] R. Macwan, Y. Benezeth, and A. Mansouri, "Remote photoplethysmography with constrained ica using periodicity and chrominance constraints," *Biomedical engineering online*, vol. 17, no. 1, pp. 1–22, 2018.

[23] L. Feng, L.-M. Po, X. Xu, Y. Li, C.-H. Cheung, K.-W. Cheung, and F. Yuan, "Dynamic roi based on k-means for remote photoplethysmography," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 1310–1314.

[24] Y.-Y. Tsou, Y.-A. Lee, C.-T. Hsu, and S.-H. Chang, "Siamese-rppg network: Remote photoplethysmography signal estimation from face videos," in *Proceedings of the 35th annual ACM symposium on applied computing*, 2020, pp. 2066–2073.

[25] R. Song, J. Li, M. Wang, J. Cheng, C. Li, and X. Chen, "Remote photoplethysmography with an eemd-mcca method robust against spatially uneven illuminations," *IEEE Sensors Journal*, vol. 21, no. 12, pp. 13 484–13 494, 2021.

[26] L. Feng, L.-M. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 5, pp. 879–891, 2014.

[27] H. Takeuchi, M. Ohsuga, and Y. Kamakura, "A study on region of interest in remote ppg and an attempt to eliminate false positive results using svm classification," in *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*. IEEE, 2021, pp. 1–5.

[28] L.-M. Po, L. Feng, Y. Li, X. Xu, T. C.-H. Cheung, and K.-W. Cheung, "Block-based adaptive roi for remote photoplethysmography," *Multimedia Tools and Applications*, vol. 77, pp. 6503–6529, 2018.

[29] Y. Yang, C. Liu, H. Yu, D. Shao, F. Tsow, and N. Tao, "Motion robust remote photoplethysmography in cielab color space," *Journal of biomedical optics*, vol. 21, no. 11, pp. 117 001–117 001, 2016.

[30] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[31] T. Paul, U. A. Shammi, M. U. Ahmed, R. Rahman, S. Kobashi, and M. A. R. Ahad, "A study on face detection using viola-jones algorithm in various backgrounds, angles and distances," *International Journal of Biomedical Soft Computing and Human Sciences: the official journal of the Biomedical Fuzzy Systems Association*, vol. 23, no. 1, pp. 27–36, 2018.

[32] K. Wanjale, A. Bhoomkar, A. Kulkarni, S. Gosavi, and V. Pune, "Use of haar cascade classifier for face tracking system in real time video," *International Journal of Engineering Research and Technology*, vol. 2, no. 4, 2013.

[33] V. T., "Face detection using opencv with haar cascade classifiers," Apr 2022. [Online]. Available: https://becominghuman.ai/face-detection-using-opencv-with-haar-cascade-classifiers-941dbb25177

[34] H. Ge, Y. Dai, Z. Zhu, and B. Wang, "Robust face recognition based on multi-task convolutional neural network," *Math Biosci Eng*, vol. 18, no. 5, pp. 6638–6651, 2021.

[35] Y. Zhang, P. Lv, and X. Lu, "A deep learning approach for face detection and location on highway," in *IOP Conference Series: Materials Science and Engineering*, vol. 435, no. 1. IOP Publishing, 2018, p. 012004.

[36] D. E. King, "Dlib-ml: A machine learning toolkit," *The Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[37] 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.

[38] L. Malasinghe, S. Katsigiannis, N. Ramzan, and K. Dahal, "Remote heart rate extraction using microsoft Kinect™ v2. 0," in *Proceedings of the 2018 10th International Conference on Bioinformatics and Biomedical Technology*, 2018, pp. 1–6.