WaveHRV: robust heart rate variability measurement from facial videos

Ismoil Odinaev, Kwan Long Wong, Jing Wei Chin, Raghav Goyal

***Abstract*— Remote Photoplethysmography (rPPG) is a contactless method that aims to capture various physiological signals from facial videos. rPPG utilizes a RGB camera to detect subtle changes in skin color to measure vital signs such as Heart Rate Variability (HRV), an important biomarker related to the autonomous nervous system. This paper presents a novel HRV extraction algorithm, WaveHRV, based on the Wavelet Scattering Transform technique, followed by adaptive band pass filtering and IBI filtering. A novel method is introduced to preprocess noisy contact-based PPG signals. WaveHRV is bench-marked against existing algorithms and public datasets. The results of this paper show that WaveHRV outperforms existing methods on SDNN and RMSSD metrics, with SDNN MAE being 7.3ms and RMSSD MAE 10.6ms.**

***Index Terms*— Heart Rate Variability, Remote Photoplethysmography, Wavelet Scattering Transform, RMSSD, SDNN**

I. Introduction

Heart Rate Variability is a variation in time between consecutive heartbeats. It is closely related to the autonomous nervous system (ANS), actual heart sound, blood pressure, and mental wellness [1]. Traditionally, HRV has been measured by using contact-based electrocardiogram (ECG), which may result in invasive medical procedures such as attaching electrodes to various parts of the body. Recently, non-contact measurement of HRV has gained momentum due to its non-invasive nature and suitability. Contactless HRV can be obtained from an optical technique known as remote Plethysmography (rPPG) by using a portable digital camera.

Within the past decade, much work has focused on HRV using rPPG. A typical pipeline would consist of face detection and tracking, skin segmentation, ROI selection, and rPPG construction [2]–[5]. Further steps were added to clean, filter, or denoise the rPPG signal. Research by Mitsihashi et el., [6] obtained the rPPG signal from facial videos by using 2SR [7] and applied detrending, bandpass filtering [0.75Hz - 3Hz], interpolation, and valley detection to source HRV. A further study [8] used color amplification on a red channel and peak detection to calculate numerous HRV metrics. Qiao et al., [9] applied ICA to calculate the rPPG signal, and then utilized detrending and normalization to further clean the signal. Li et al., [2]btained rPPG with a chrominance-based method [10]and devised an algorithm called SSF that would both enhance the rising trend and decrease the downward trend of an rPPG signal. This paper [3] proposed a post-processing step based on a continuous wavelet transform termed CWT-BP and CWT-MAX. CWT-BP is bandpass filtering [0.75-4 Hz], whereas CWT-MAX is a denoising step based on the scale of the summation of CWT coefficients. He et al., [4] further developed CWT-based denoising methods by introducing CWT-SNR. The algorithm selects coefficients based on signal to noise ratio of the reconstructed rPPG signal. Gudi et al., [5] applied motion noise suppression and narrow fixed bandpass filtering to clean the rPPG signal and subsequently source the HRV. Further research [11], in addition to introducing narrow adaptive bandpass filtering, also made use of a DESA algorithm to calculate instantaneous frequencies, and subsequently HRV. Overall, traditional methods have focused on post-processing steps such as bandpass filtering, detrending and continuous wavelet transform to clean noisy rPPG signals.

A deep learning approach was offered by Song et al., [12]. The most important part of their method was calculating rPPG with traditional algorithms, and then employing GAN to generate a cleaner version of that rPPG signal. This paper [13] proposed an end-to-end deep learning model to obtain an rPPG signal. Their model is based on different 3D-CNN and LSTM networks and benchmarked against frequency domain HRV metrics.

All listed HRV algorithms suffer from relatively poor results when compared to ground truth contact based values. Additionally, deep learning models require a large amount of data to train on, which can be expensive. Since HRV is highly sensitive to noise, improved algorithms should be devised to decrease the gap between contact and camera HRV. Therefore, in this paper, we introduce:

1. a novel HRV algorithm, WaveHRV, that outperforms other algorithms and,

2. an innovative preprocessing step to filter out noisy ground truth data.

This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment. For example, “This work was supported in part by the U.S. Department of Commerce under Grant BS123456.”

The next few paragraphs should contain the authors’ current affiliations, including current address and e-mail. For example, F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@boulder.nist.gov).

S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).

T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

II. Method

The heart rate variability extraction pipeline obtained from a video is presented in Fig. 1. Initially, the subject's face is detected and tracked by Medipipe FaceMesh [14]. This is followed by a process of skin segmentation to remove non-skin regions and improve signal quality. Then, the mean RGB signal is acquired by taking average of each frame spatially and concatenating them temporally. This meanRGB signal is fed to the Plane Orthogonal to Skin (POS) [15] algorithm to get the rPPG signal candidate. Finally, the rPPG signal is interpolated to the nearest power of 2 to make it easier to work with the scattering transform, and make the signal spaced equally in time.

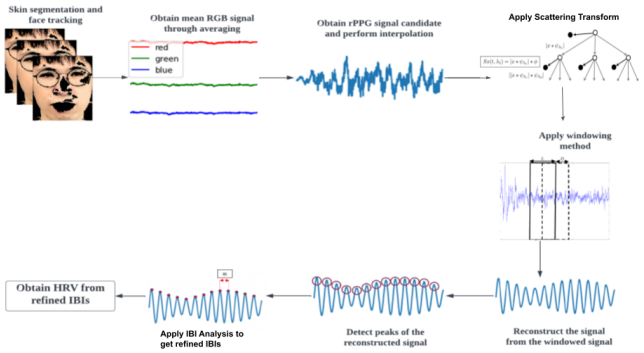


Fig. 1: Pipeline to extract heart rate variability from facial videos.

A. Scattering Transform

The scattering transform (ST) [16] is a complex-valued Convolutional Neural Network (CNN) whose filters are fixed wavelets and has modulus as non-linearity and averaging as pooling. It is invariant to translation, frequency shifting, and change of scale. Coefficients of the ST can be interpreted as energy density, and N^{th} order coefficients are given by:

$$S_N(t, \lambda_1, ...\lambda_N) = |r(t) * \psi_{\lambda_1}| ... * \psi_{\lambda_N}| * \phi \qquad (1)$$

where $r(t)$ is a signal, ψ_λ is wavelet of scale λ, ϕ is average pooling, $|...|$ is complex-valued modulus operation and $*$ is convolutional operation. In this paper, the Kymatio Library [17] was used to implement Scattering Transform, and the Morlet wavelet was chosen to convolve with the signal. An example of coefficients of first-order ST of a signal with pooling size of 16 seconds and filter bank of 20 is given in Fig. 2. Frequencies in the y-axis increase exponentially, while time in the x-axis is given as discreet numbers that are multiples of 16 seconds due to chosen pooling size.

B. Windowing

The interpolated rPPG signal is first cleaned with Butterworth bandpass filtering of order 30 with band size 0.7Hz-5Hz to acquire $rPPG_{clean}$. Then, the first-order scattering transform is applied to the obtained signal as explained in section II-A. Afterwards, a windowing step, shown in Fig. 3, is applied on $rPPG_{clean}$ in the following manner:

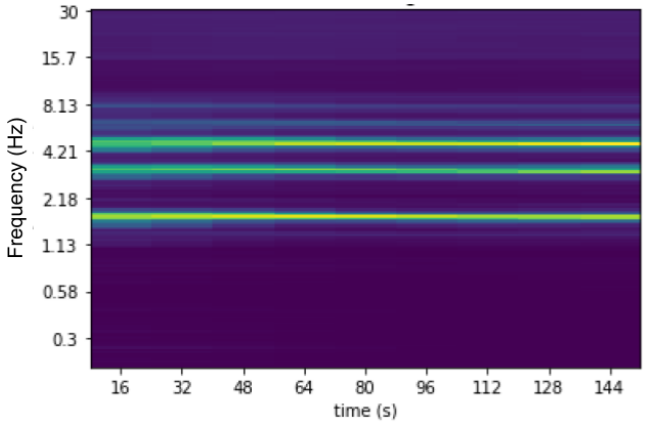


Fig. 2: Example of the coefficients of the first-order scattering transform.

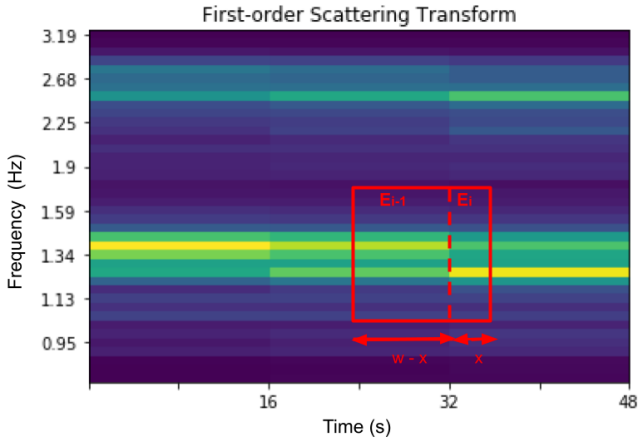


Fig. 3: Example of the energy calculation for a particular window of the first-order scattering transform.

1. For each window of size w calculate the energy around the first harmonic by the given equation:

$$E = \frac{w - x}{w} E_{i-1} + \frac{x}{w} E_i \qquad (2)$$

where w is window size, E_i is the energy at time i and x is difference between right-end of the window and time i.

2. Construct K-Means (K = 3) clustering with frequency and energy E as an input and k-mean++ as a centroid initialization to obtain a narrow band as shown in Fig. 4. The centers of the clusters are shown in red in the figure. Then, the band size is [left centroid, right centroid]

3. Apply Butterworth bandpass filtering on the windowing signal with previously obtained bands.

4. Subtract the mean of the windowing signal from the windowing signal itself to retain only the pulsatile part and remove the diffuse part.

5. Slide window over whole signal with window size = w and step size = s that can be optimized for different datasets.

6. Reconstruct the cleaned rPPG signal from the windowing segments by adding the segments.

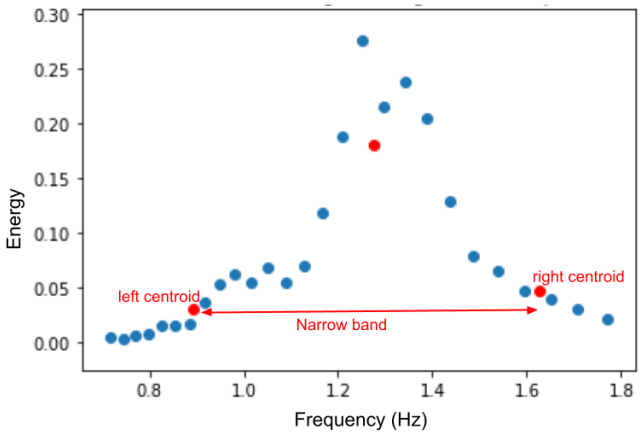


Fig. 4: Narrow band calculation with K-Means Clustering (K = 3)

Due to sliding windows, peaks on the edges will be smaller than the rest of the signal. This may result in peak detection issues which can be solved by multiplying both edges of the signal with coefficients (c), as shown in the pseudo-code below:

Algorithm 1 Peak amplification of the two ends of the signal

$w \leftarrow windowsize$
$s \leftarrow stepsize$
$j \leftarrow 0$
$R \leftarrow signal$
while $j \leq w/s$ **do**
  $c \leftarrow 2\frac{w}{s(j+1)}$
  $R[s \times j : s \times (j+1)] \leftarrow R[s \times j : s \times (j+1)] \times c$
  $j \leftarrow j + 1$
end while

C. IBI Analysis

Peaks of the reconstructed signal are detected with the AMPD [18] algorithm and Inter-beat Intervals (IBIs) are calculated. Then, refined IBIs are calculated by removing physically impossible regions and misplaced peaks, and retaining only those IBIs that satisfy the criteria below:

1. $\forall IBI \in [400ms, 1300ms]$
2. $\forall IBI \in mean(IBI) \pm 0.4mean(IBI)$
3. Slide non-overlapping window over IBIs with window size 10. IBIs in each window should satisfy $\forall IBI_{window} \in mean(IBI_{window}) \pm 0.2mean(IBI_{window})$.

III. HRV Metrics

SDNN is an HRV metric related to the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) and associated with overall mental wellness [1]. Multiple studies show that [1] [19] the range for short-term $(< 5mins)$ SDNN is 32-93ms and it is given by:

$$SDNN = \sqrt{(\frac{1}{N-1})\sum_{i=1}^{N}(IBI_i - \overline{IBI})^2} \quad (3)$$

TABLE I: Summary of datasets used in this paper

Dataset	# videos	fps	compressed	Ground Truth
Stroop	42	60	no	PPG (60Hz)
UBFC rPPG [22]	42	30	no	PPG (30Hz)
VIPL-HR [23]	1968	25/30	yes	PPG (60Hz)
MAHNOB-HCI [24]	1095	60	yes	ECG (256Hz)

RMSSD is an HRV metric related to PNS [1] and strongly related to human productivity and energy levels. Short-term RMSSD lies with 19-75ms [1] [19]. It is given by:

$$RMSSD = \sqrt{(\frac{1}{N-1})\sum_{i=1}^{i=N-1}(IBI(i+1) - IBI(i))^2} \quad (4)$$

where IBI is the inter-beat interval.

Baevsky Stress Index (BaevskySI) is a stress metric that represents the mental or physical stress one is experiencing. It is very sensitive to SNS and has a range of 80 to 1000-1500 depending on stress level and stress-related illnesses [20]. It is derived using HRV metrics as follows:

$$BaevskySI = \frac{AMo(IBI)}{2 * Mo(IBI) * MxDMn(IBI)} \quad (5)$$

Where $AMo(IBI)$ is mode amplitude of IBIs, $Mo(IBI)$ is the mode of the IBI, and $MxDMn(IBI)$ is the difference between the maximum and minimum IBI.

Finally, **LF/HF** is a frequency domain metric that represents the balance between the PNS and the SNS [1]. The LF [0.04 - 0.15 Hz] represents the SNS and the HF (0.15 - 0.4 Hz] represents the PNS. This is considered a metric that provides insight into the equilibrium of the Autonomic Nervous System and the resilience of the body to changes and stress [1]. LF/HF values range between 1.1 and 1.6 [1] [19].

IV. Dataset

To validate the algorithm we used our dataset and three publicly available datasets. The summary of these datasets is shown in Table I.

The Stroop dataset in Table I is our private dataset and it is described below.

A. Stroop Dataset

Fourteen adults of ages ranging from 18 to 33 and with varying skin tones took part in our experiment. They were seated one meter in front of a Logitech Brio camera which recorded video at 60 fps in ambient room lighting. A CONTEC CMS-60C pulse oximeter set at a frequency of 60 Hz was used to record the ground truth PPG signal. The Stroop test [21] was used to induce cognitive stress and allow for HRV measurement under different experimental stages. The test consisted of 4 parts: A Ground Truth Rest Stage (1 min), the Stroop test with sound stimulus (3 mins), a Rest Stage (2 mins), and the Stroop test without sound stimulus (3 mins). During the Stroop Test with sound stimulus, participants heard a positive or negative audio sound based on whether they gave the correct answer.

B. Publicly Available Datasets

UBFC rPPG [22] consists of 42 subjects and 42 videos. Each video is approximately a minute long, 30 fps, and uncompressed. Videos are recorded in uniform, ambient lighting and subjects play math tests to induce stress and increase heart rate.

VIPL-HR [23] consists of 107 subjects and 2378 videos. Videos length ranges from 10 seconds to 1 minute. Videos are compressed and recorded by 3 different devices. Subjects are recorded under 7 different scenarios: stable scenario, talking scenario, large head movements, dark lighting, bright lighting, long distance scenario, and after-exercise. In this paper, we only used videos that are longer than 16 seconds because it is difficult to obtain meaningful HRV results based on measurements that are less than 15 seconds.

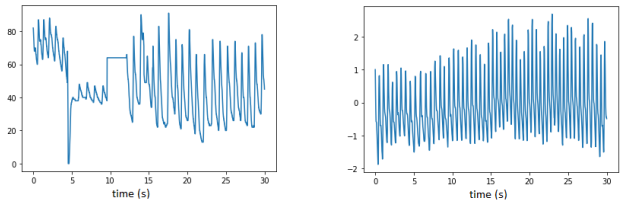
MAHNOB-HCI [24] consists of 27 subjects and 3465 videos. To induce different emotions and feelings of stress, subjects watch different videos while sitting in front of the camera. Videos are compressed and range from 5 seconds to 3.5 minutes. In this dataset as well, only videos that are longer than 16s are selected.

C. Ground Truth Preprocessing

HRV is a sensitive biomarker and even a slight disturbance during the data collection process can alter the outcome dramatically. This paper [25] shows the impact of false peaks on HRV measurement and points out that if a small percentage of peaks are dislocated, HRV results will be significantly different. Therefore, ground truth data must be preprocessed before comparing it with camera HRV. There are several reasons why ground truth data is noisy: disconnection of the ground truth device, covering the face with hands, excessive body motion, slight motion of the fingers inside a pulse oximeter, etc. Examples of a noisy and clean PPG signal are shown in Fig. 5.

To filter out these noisy ground truth data, we came up with criteria based on biological restrictions and data analysis. First, since we are calculating HRV from the face, any obstacle between the face and the camera leads to discontinuity in the signal. Therefore, this type of data is discarded. Second, if the measured heart rate is beyond physiological and biological limits at any point, then the subject is disconnected from the ground truth data-collecting device. This kind of ground truth data cannot be used as a reference. Third, van Gent et al. [25] demonstrate that false peaks change HRV results significantly and removing them is an optimal solution. They suggest removing $IBIs$ that are off by 30% from $meanIBI$ of the chosen segment. Finally, this meta-analysis [1] reports results of more than 20 studies concluding that short-term SDNN and RMSSD should be less than 92ms and 75ms respectively. Contact-based PPG and ECG HRV results that are beyond the physiologically possible region should be removed. These criteria can be summarised as follows:

1. Remove data with a covered face at any instant in time
2. Remove data that have $HR \notin [45, 200]$
3. Remove $\forall IBI \notin mean(IBI_{segment}) \pm 0.3mean(IBIsegment)$, where $segment$ is 20-30 IBIs
4. Remove data that have $SDNN \ \& \ RMSSD > 100ms$



(a) noisy PPG example (b) clean PPG example

Fig. 5: Example of noisy and clean PPG signals.

TABLE II: Performance of SDNN measurement for the UBFC rPPG, VPIL-HR and MAHNOB-HCI datasets. Best results are presented in bold

Dataset	UBFC rPPG	VIPL-HR	MAHNOB-HCI
Metrics	MAE±STD (ms)	MAE±STD (ms)	MAE±STD (ms)
WaveHRV	**7.72±6.15**	**29±45**	**69±234**
FaceRPPG* [5]	19±14.5	49±45	107±51
SSF [2]	25	-	-
PulseGAN [12]	24.3	-	-

*The results of this method are given against a cleaned version of the data

V. Results and Analysis

A. Benchmarking WaveHRV

We reported the results of WaveHRV on publicly available datasets in Table II for SDNN and Table III for RMSSD. All other algorithms except FaceRPPG reported their results on the UBFC rPPG dataset only. FaceRPPG recorded their results on many datasets, however, their results were benchmarked against cleaned versions of datasets. It can be seen from Table II and Table III that WaveHRV outperformed all other methods by a significant margin. We observed that VIPL-HR and MAHNOB-HCI have large MAEs and even larger standard deviations. The primary reason for this phenomenon is caused by disconnected and noisy ground truth data mentioned in IV-C.

B. WaveHRV on the Preprocessed Datasets

After filtering out noisy ground truth data according to criteria mentioned in Section IV-C, we get results presented in Table IV. Comparing the results of Table IV against Tables III and II we see that the proposed ground truth data cleaning method performed well. MAE of RMSSD of UBFC rPPG decreased from 12.6ms to 10.6ms. The effect of the proposed criteria is very noticeable on datasets like MAHNOB-HCI and VIPL-HR where the ground truth device is under motion or poorly connected. By looking at the tables, we can see that the SDNN MAE of VIPL-HR decreased from 29ms to 13.3ms, RMSSD MAE of VIPL-HR decreased from 41ms to 15.1ms.

TABLE III: Performance of RMSSD measurement for the UBFC rPPG, VPIL-HR and MAHNOB-HCI datasets. Best results are presented in bold

Dataset	UBFC rPPG	VIPL-HR	MAHNOB-HCI
Metrics	MAE±STD (ms)	MAE±STD (ms)	MAE±STD (ms)
WaveHRV	**12.6±12.8**	**41±70**	**93±317**
FaceRPPG* [5]	16±22.5	73±57.8	108±51
SSF [2]	47	-	-

*The results of this method are given against a cleaned version of the data

TABLE IV: SDNN, RMSSD, Baevsky Stress Index and LF/HF performance of WaveHRV on Prepreocessed Datasets

Dataset	SDNN (ms) MAE±STD	RMSSD (ms) MAE±STD	BaevskySI MAE±STD	LF/HF MAE±STD
Stroop	7.0 ± 4.80	11.35± 9.13	38 ± 45	0.67 ± 0.76
UBFC rPPG	7.30 ± 5.98	10.60±8.86	57 ± 54	0.26 ± 0.32
VIPL-HR	13.3 ± 11.1	15.1 ± 13.1	98±122*	0.43± 0.63*
MAHNOB-HCI	17.5 ± 12.5	21.5 ± 14.5	55±65*	0.33±0.39*

* Videos that are longer than 30s are considered

As for MAHNOB-HCI, SDNN MAE decreased from 69ms to 17.5ms, while RMSSD MAE decreased from 93ms to 21.5ms. When we look at the standard deviations (STD) of VIPL-HR and MAHNOB-HCI, we see that STD of VIPL-HR decreased from 45ms to 11.1ms for SDNN and from 70ms to 13.1ms for RMSSD. STD of MAHNOB-HCI decreased from 234ms to 12.5ms for SDNN and from 317ms to 14.5ms for RMSSD.

From Table IV we note that WaveHRV has lower MAEs on UBFC rPPG and Stroop datasets than on challenging datasets like VIPL-HR and MAHNOB-HCI. UBFC rPPg and Stroop are not compressed and have uniform ambient light whereas VIPL-HR and MAHNOB-HCI are compressed and recorded under non-uniform or dim lighting. Furthermore, in some scenarios of the VIPL-HR, subjects perform large head movements, talk or sited further away from the camera.

Bland-Altman plots of SDNN and RMSSD of three preprocessed datasets namely Stroop, UBFC rPPG, and VIPL-HR are shown in Fig.7, Fig.8 and Fig.6 respectively. We also show correlation plots of SDNN and RMSSD for the most cited dataset, UBFC rPPG. Looking deeper into Fig. 7, it can be observed that SDNN mean error is -0.29ms, whereas RMSSD is 4.03ms for Stroop Dataset. Hypothesis testing between contact HRV and camera HRV means shows p-values < 0.83 and 0.09 for SDNN and RMSSD respectively. Fig. 8 reveals mean error of 4.04ms for SDNN and 2.82ms for RMSSD. P-values are $<$ 0.06 and 0.21 for SDNN and RMSSD respectively. For VIPL-HR we have SDNN mean error of 1.44 ms (p-value < 0.03) and RMSSD -1.58ms (p-value < 0.06). Finally, MAHNOB-HCI has SDNN -2.72ms mean error and RMSSD -8.5ms mean error corresponding to p-values $<$ 0.02 and 0.0001 respectively. Statistical Analysis at a 95% Confidence Interval implies that WaveHRV is slightly prone to either overestimate or underestimate HRV results. This is especially obvious for large datasets like VIPL-HR and MAHNOB-HCI. A deeper look into Bland-Altman plots like Fig.7 and Fig.8b, Fig.8d reveal that waveHRV is overestimating for small HRV values and underestimating for larger values.

C. Stress Measurement

The performance of WaveHRV on physiological stress-related metrics is given in Table IV. Baevsky Stress Index (BaevskySI), also known as the Strain Index, characterizes a person's sympathetic nervous system activity (SNS) and is good indicator of physical and mental load. MAE of BaevskySI from the contact-based device and WaveHRV is 40-60 for UBFC rPPG, Stroop, and MAHNOB-HCI datasets, while VIPL-HR has BaevskySI MAE 100. As mentioned above in Section II-C, BaevskySI has values between 50-1500

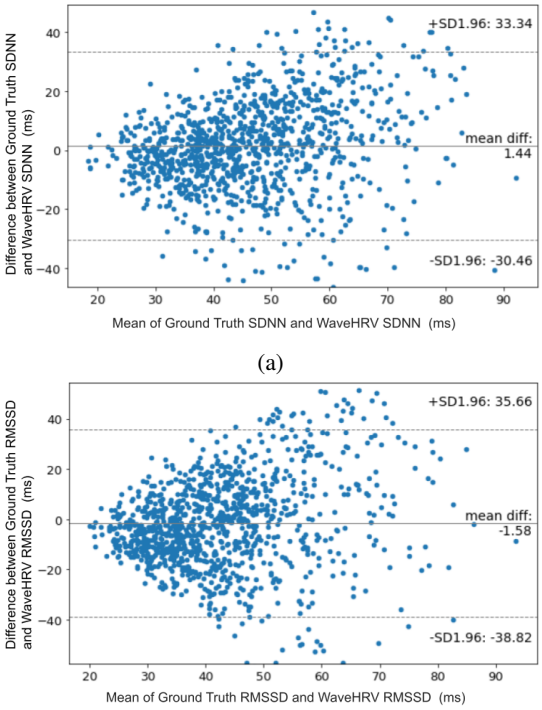


Fig. 6: Bland-Altman plots for WaveHRV compared to the ground truth PPG device on the preprocessed VIPL-HR dataset: (a) SDNN and (b) RMSSD. Error bars and 95% confidence intervals are marked (in ms).

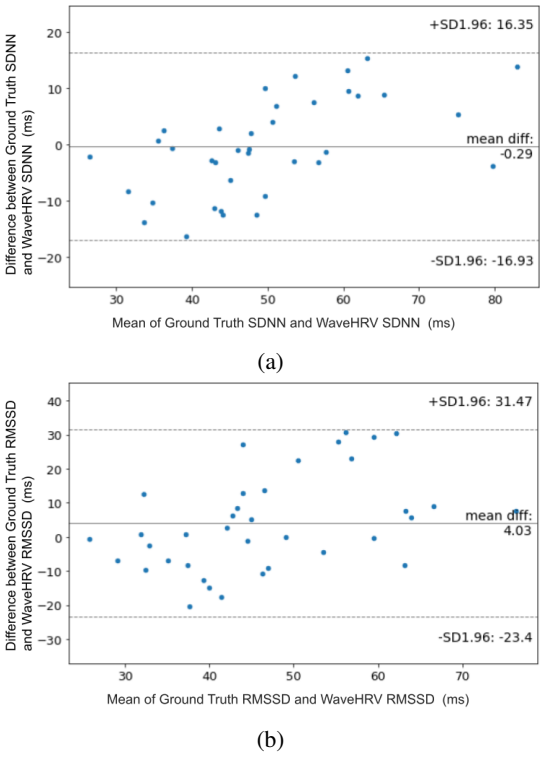
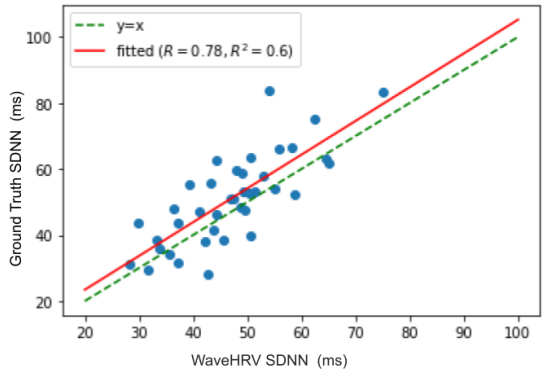
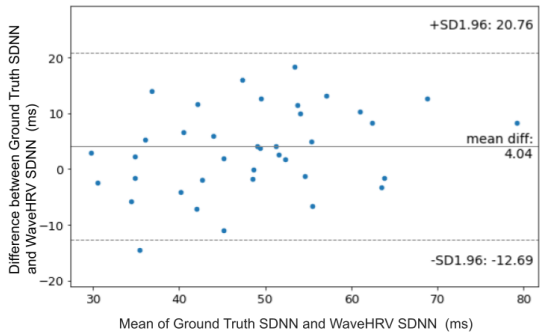


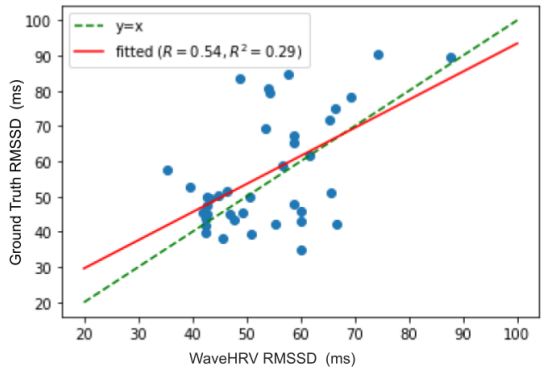
Fig. 7: Bland-Altman plots for WaveHRV compared to the ground truth PPG device on the preprocessed Stroop dataset: (a) SDNN and (b) RMSSD. Error bars and 95% confidence intervals are marked (in ms).



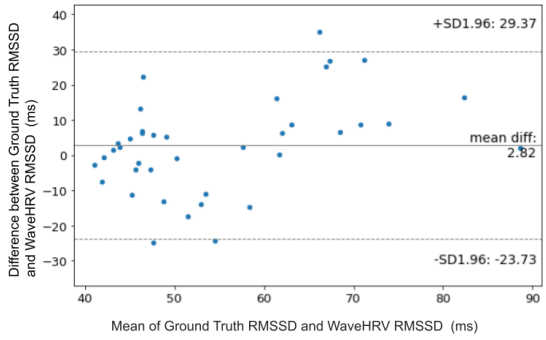
(a)



(b)



(c)



(d)

Fig. 8: Bland-Altman and Correlation plots for WaveHRV compared to the ground truth PPG device on the preprocessed UBFC-rPPG dataset: (a) SDNN Correlation, (b) SDNN Bland-Altman, (c) RMSSD Correlation and (d) RMSSD Bland-Altman plots. Error bars and 95% confidence intervals are marked (in ms) in Blant-Altman plots. Pearson Correlation Coefficient (R) and Coefficient of Determination (R^2) are given in Correlation plots

and looking at the results of WaveHRV, it can be inferred that our algorithm can be utilized to distinguish between different stress levels.

LF/HF is a metric of homeostasis and resilience of the ANS to the stress and anxiety. LF/HF values range between 1-11.5 and TableIV illustrates that LF/HF MAE lies between 0.26-0.67. Therefore, WaveHRV could be used to obtain LF/HF and has potential to give insights into the balance and equilibrium of ANS.

VI. Conclusion

In this paper, we have presented WaveHRV, a novel algorithm for HRV extraction from a portable camera. We benchmarked our algorithm against other methods and demonstrated that WaveHRV outperforms other methods on publicly available datasets. Furthermore, we presented a straightforward yet powerful technique to clean ground truth data and highlighted its performance. We also demonstrated the potential for an off-shelf camera to measure stress and mental wellbeing. A further direction of this research would include causes for bias in WaveHRV at different HRV values. Further research can be carried out on improving HRV algorithms under challenging scenarios like large head movements and dim lighting and reducing discrepancy between camera HRV and contact HRV.

References

[1] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, p. 258, 2017.

[2] P. Li, Y. Benezeth, K. Nakamura, R. Gomez, C. Li, and F. Yang, "An improvement for video-based heart rate variability measurement," in *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, pp. 435–439, IEEE, 2019.

[3] R.-Y. Huang and L.-R. Dung, "Measurement of heart rate variability using off-the-shelf smart phones," *Biomedical engineering online*, vol. 15, no. 1, pp. 1–16, 2016.

[4] L. He, K. S. Alam, J. Ma, E. Burkholder, W. C. C. Chu, A. Iqbal, and S. I. Ahamed, "Remote photoplethysmography heart rate variability detection using signal to noise ratio bandpass filtering," in *2021 IEEE International Conference on Digital Health (ICDH)*, pp. 133–141, IEEE, 2021.

[5] A. Gudi, M. Bittner, and J. van Gemert, "Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation," *Applied Sciences*, vol. 10, no. 23, p. 8630, 2020.

[6] R. Mitsuhashi, K. Iuchi, T. Goto, A. Matsubara, T. Hirayama, H. Hashizume, and N. Tsumura, "Video-based stress level measurement using imaging photoplethysmography," in *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 90–95, IEEE, 2019.

[7] W. Wang, S. Stuijk, and G. De Haan, "A novel algorithm for remote photoplethysmography: Spatial subspace rotation," *IEEE transactions on biomedical engineering*, vol. 63, no. 9, pp. 1974–1984, 2015.

[8] G. H. Martinez-Delgado, A. J. Correa-Balan, J. A. May-Chan, C. E. Parra-Elizondo, L. A. Guzman-Rangel, and A. Martinez-Torteya, "Measuring heart rate variability using facial video," *Sensors*, vol. 22, no. 13, p. 4690, 2022.

[9] D. Qiao, F. Zulkernine, R. Masroor, R. Rasool, and N. Jaffar, "Measuring heart rate and heart rate variability with smartphone camera," in *2021 22nd IEEE International Conference on Mobile Data Management (MDM)*, pp. 248–249, IEEE, 2021.

[10] 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.

[11] A. Pai, A. Veeraraghavan, and A. Sabharwal, "Hrvcam: robust camera-based measurement of heart rate variability," *Journal of Biomedical Optics*, vol. 26, no. 2, p. 022707, 2021.

[12] R. Song, H. Chen, J. Cheng, C. Li, Y. Liu, and X. Chen, "Pulsegan: Learning to generate realistic pulse waveforms in remote photoplethysmography," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1373–1384, 2021.

[13] Z. Yu, X. Li, and G. Zhao, "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," *arXiv preprint arXiv:1905.02419*, 2019.

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

[15] 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.

[16] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.

[17] M. Andreux, T. Angles, G. Exarchakis, R. Leonarduzzi, G. Rochette, L. Thiry, J. Zarka, S. Mallat, J. Andén, E. Belilovsky, *et al.*, "Kymatio: Scattering transforms in python.," *J. Mach. Learn. Res.*, vol. 21, no. 60, pp. 1–6, 2020.

[18] F. Scholkmann, J. Boss, and M. Wolf, "An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals," *Algorithms*, vol. 5, no. 4, pp. 588–603, 2012.

[19] A. Voss, R. Schroeder, A. Heitmann, A. Peters, and S. Perz, "Short-term heart rate variability—influence of gender and age in healthy subjects," *PloS one*, vol. 10, no. 3, p. e0118308, 2015.

[20] R. M. Baevsky and A. G. Chernikova, "Heart rate variability analysis: physiological foundations and main methods," *Cardiometry*, pp. 66–76, 2017.

[21] J. R. Stroop, "Studies of interference in serial verbal reactions.," *Journal of experimental psychology*, vol. 18, no. 6, p. 643, 1935.

[22] 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.

[23] X. Niu, H. Han, S. Shan, and X. Chen, "Vipl-hr: A multi-modal database for pulse estimation from less-constrained face video," in *Asian conference on computer vision*, pp. 562–576, Springer, 2018.

[24] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 42–55, 2011.

[25] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the taking the fast lane project," *Journal of Open Research Software*, vol. 7, no. 1, 2019.