

Contactless heart rate estimation from facial video using skin detection and multi-resolution analysis

Khawla BENSALAH
National Engineering
School of Sfax
Research Lab: Math,
Earth Sciences, Modeling
and Intelligent Systems,
Faculty of Sciences of
Gafsa
University of Sfax, BP
1173, Sfax, Tunisia
khawlabensalah8@gmail.com

Mohamed OTHMANI
Faculty of Sciences of
Gafsa, University of
Gafsa
BP 2100, Gafsa, Tunisia
mohamed.othmani@yahoo.fr

Monji KHERALLAH
Faculty of Sciences of
Sfax, University of Sfax
BP 1173, Sfax, Tunisia
monji.kherallah@fss.usf.tn

ABSTRACT

This paper introduces a remote Photoplethysmography (rPPG), which is used to estimate human heart rate without any physical contact, has been extensively applied in multiple fields like medical diagnosis, analysis of human emotions, rehabilitation training programs, biometric, and fitness assessments. The rPPG signals are usually extracted from facial videos. However, it is still a challenging task due to several contributing factors, e.g., variation in skin tone, lighting condition, and subject's motion. Accordingly, in this work, a novel approach based on deep learning skin detection method and the discrete wavelet transform (DWT) is employed to precisely estimate heart rate from facial videos. In the proposed method, by implementing the DWT, the signal is decomposed into approximations and details parts thereby it helps in analyzing it at different frequency bands with different resolutions. The results derived from the experiments show that our proposed method outperforms the state-of-the-art methods on the UBFC-RPPG database.

Keywords

rPPG, Heart rate estimation, facial video, Deep learning, Discrete wavelet transform, skin detection.

1 INTRODUCTION

In the recent years, remote heart rate estimation from facial videos (rPPG) has been increasingly investigated. Remote photoplethysmography (rPPG) traces the blood volume pulse remotely by tracking changes in the skin reflectance during the cardiac cycle by the means of a digital camera, although these color changes are not noticeable to the necked human eye. This color variation helps to determine the required information to measure the heart rate. Heart Rate(HR) estimation is an important component to determine the health state and the well-being of an individual and it is usually evaluated as normal if it is between 60 to 100 beats per minute (bpm) [1]. (HR) analysis has always been an area of interest not only for the medical applications but

also for other applications from a variety of other fields such as include telemetry supervising of elderly people/newborn babies [2], the athletes involved in sports activities [3], emotional recognition [4], human authentication [5], etc. Traditionally, we assure heart rate estimation with equipment, such as an electrocardiogram (ECG) or conventional photoplethysmography (PPG), however, it requires electrodes/sensors attached to the skin surface, gel, and experienced nurses or doctors, and usually causes skin discomfort. Hence, rPPG has gained more and more interest because of its several advantages as comfort, non-complicated, non-invasive and inexpensive characteristics. Additionally, in some cases, such as burns, severe infections, etc the employ of a sensor is not always possible. Despite the diversity of methods, the main challenge consists in performing robustness to variation in skin tone, natural motion and illumination variation that results in undesirable noise and artifacts in the extracted heart rate. In this paper, we present a novel approach to achieve heart rate detection (HR) from facial videos. Our approach consists of 6 following steps: the first step consists of performing face alignment and detecting the face with viola

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

and Jones algorithms. Since the PPG information is more accurate in some part of the face like the cheek and the forehead, we choose to focus the ROI on the forehead. Accordingly, the second step was in detecting the forehead in all the frames. In Step3, we apply our algorithm of skin detection based on convolutional neural networks which is detailed in [6]. In Step 4, the rPPG waveform is obtained from the mean of the sequence green channeled video from the skin-region. In the fifth step, the discrete wavelet transform (DWT) is employed to precisely estimate heart rate from facial videos. A novel spatiotemporal representation is introduced by implementing the DWT, the signal is decomposed into approximations and details parts thereby it helps in analyzing it at different frequency bands with different resolutions. In the final step, we estimate the heart rate by identifying the maximal frequency using spectral representation.

2 RELATED WORK

Remote heart rate measurement based-video was the first attempt by Verkrusse et al.[7]. Poh et al.[8] used Blind Source Separation (BSS) technique to extract PPG signal, which yields a dynamic combination of color channels. Independent component analysis (ICA) was applied to average spatially R, G, and B signals of the facial image sequence. After that, the authors extend their work in [9] to refine the PPG signal by invoking a series of temporal filtering. The extracted PPG signal was more refined further by bandpass filtering. Bandpass filters cover a predefined range of expected HRs at the operational frequency of 0.75 to 4.5 Hz [10]. A number of other methods such as the joint approximate diagonalization of eigenmatrices (JADE)[7] or FastICA [11], and extensions like joint blind source separation (JBSS)[12], Spatio-temporal ICA [13], constrained ICA [14], radical ICA [15], robust ICA [16], Project ICA [17], and Principal Component Analysis (PCA) [11] have been used to (rPPG) signal. Another group of methods proposed a set of model-based methods regarding color vectors such as the chrominance model (CHROM) [18], blood volume pulse signature (PBV) model, spatial subspace rotation (S2R), and a plane orthogonal to the skin (POS) model [19], POS is conceived to extract the heart signal based on a skin reflection model in normalized RGB space. As other researchers [20] have found, the selection of ROI has a major impact on the quality of the (rPPG) signal. Haque and al. [21] have focused on extracting facial landmark points by combining the trajectories of the supervised descent method (SDM) and generating stable trajectories (GFT) method for heart rate estimation. Other recent works have applied deep learning to extract heart rates directly from camera images. They rely on the strength of deep networks to acquire knowledge of which areas in the image correspond to heart

rate. HR-CNN [22] uses a two-step convolutional neural network. To estimate a heart rate from a sequence of facial images. DeepPhys [23] propose an end-to-end method to extract heart and breathing rate from videos using a deep convolutional network. Rhythm-Net [24] uses a spatial-temporal representation encoding the HR signals to form the CNN, and they use Gated Recurrent Unit (GRU) to consider the relationship of adjacent HR measurements from video sequences. AutoHR [25] employs neural architecture search (NAS) to discover temporal difference convolution (TDC) as a strong backbone to capture intrinsic rPPG signal from frames. Bousefsaf and al.[26] present a convolutional 3D networks, without any special frame processing, and a particular training procedure which involves only synthetic data is proposed. Hsu and al.[27] employed Short-Time Fourier Transform (STFT) to transform the signals to time-frequency representation, which is cascaded with Convolutional Neural Network (CNN). Lee and al.[28] introduce meta-learning framework for remote heart rate estimation which involves three modules: convolutional encoder, rPPG estimator and synthetic gradient generator. The first module was used to infer the rPPG signal but the two later were used during transductive learning.

3 METHODOLOGY

This section explains the overall methodology applied to estimate heart rate remotely using face videos; We follow a six-step process from face detection from video frames to heart rate estimation in this work. Fig.1 shows the flowchart of our proposed method, which is described in the following subsections.

3.1 Face Detection and face alignment

The first step of the proposed method is face alignment. Normally head it is exposed to different motion classified as internal such as (smiling/laughing, talking, screaming, eating, etc), and external related to different voluntary head motion. Thereby, we need to perform face alignment as seen in Fig.2, by using a simple method that focuses only on areas around the eyes. Three basic affine transformations are carried out: rotation, translation, and scaling. For eye detection, we have used OpenCV Haar cascade configurations (eye detection module). After that, we have calculated the center of the detected eyes and then drawing 2 lines the first connects the two central points of the eyes, the second is a horizontal line that forms an angle. Our aim is to rotate the image based on this angle. To calculate the angle, we first need to determine the length of the two legs of a right triangle. Then we can obtain the required angle using the following formulas:

$$\begin{cases} \Delta_x = x_{\{right\ eye\}} - x_{\{left\ eye\}} \\ \Delta_y = y_{\{right\ eye\}} - y_{\{left\ eye\}} \end{cases} \quad (1)$$

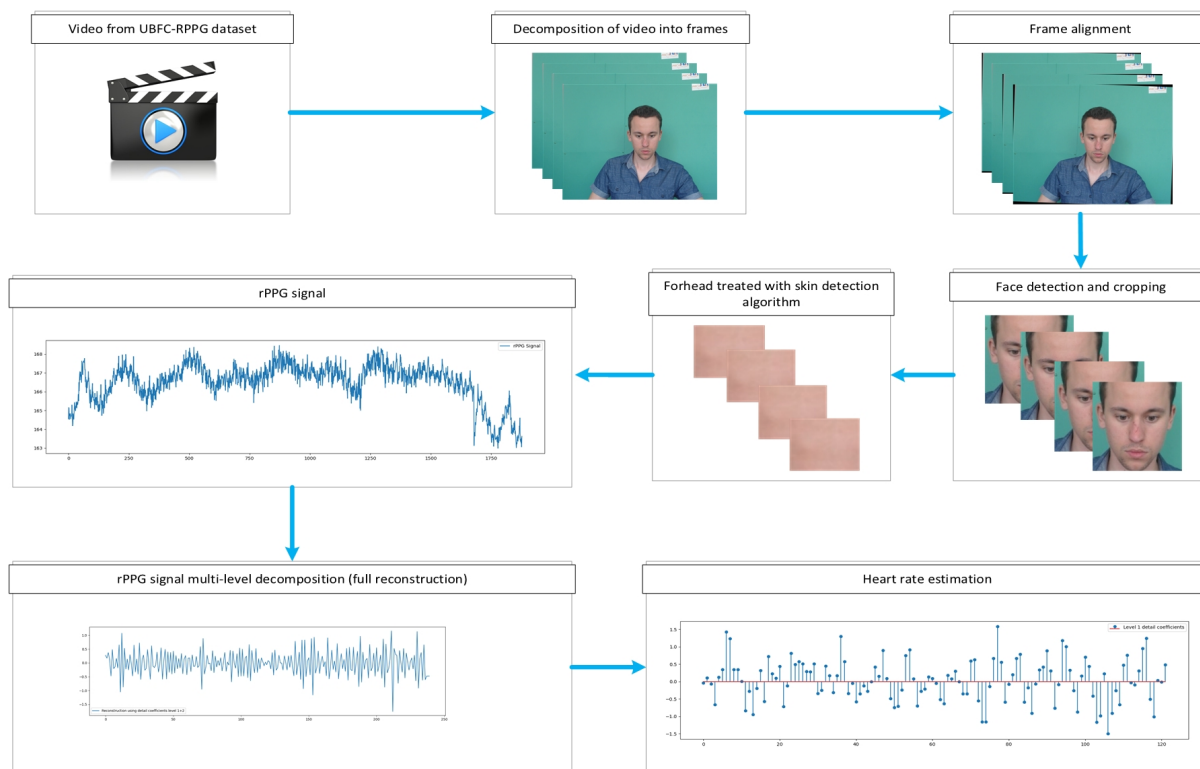


Figure 1: The flowchart of the proposed method.

$$\theta = \arctan \left(\frac{\Delta_x}{\Delta_y} \right) \quad (2)$$

we have specified in which direction our image will rotate. If the ordinate Y the left eye is bigger than the ordinate Y of the right eye, rotation is in the clockwise direction. Else, the rotation will be in the counter-clockwise direction. After rotation, we did the same thing with all the other images that we are processing. Finally, we scale our images based on that ratio. The second step is face detection and cropping which is carried out for each frame. The detection and tracking have been implemented by using the Viola-Jones face detection method [31] ; This algorithm is extremely rapid and achieves high detection rates.

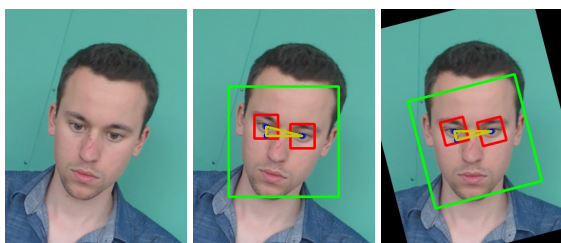


Figure 2: An example from UBFC-RPPG dataset illustrating face alignment step

3.2 Heart Rate Estimation by Forehead location

The chosen region for estimating heart rate is the forehead, the position of the forehead is determined by the coordinates of the two eyes. We have used OpenCV Haar cascade configurations (eye detection module) for eye detection. After determining the coordinate of the two eyes, we have searched for the center of both of them, the orthogonal on the center of the left eye represents the left limit of the forehead rectangle, and the orthogonal on the center of the right eye represents the right side of the forehead rectangle. The outer limit of the forehead is represented by the line of y-axis where $y = 0$. The lower limit of the forehead rectangle is determined by the line $y = (\text{height of the image}/4)$.

3.3 Skin detection

Cardiac activity causes a subtle color variation only on the facial skin during the cardiac cycle. It is, therefore, crucial to observe only the pixels that belong to the skin while ignoring the others like background, clothes, teeth, hair, and other unrelated parts. Consequently, it is important to consider that skin detection is a very crucial step in remote heart rate estimation. It improves the quality of the PPG signal by increasing the Signal-to-Noise Ratio SNR. In a previous work, we suggested a novel approach based on a convolutional neural network (CNN) for skin detection[6]. To reduce false positives, our training strategy consists of patch-based train-

ing which is robust to background clutter and detects skin pixels precisely. We proposed an efficient algorithm to detect and extract skin pixels from the whole RGB image after integrating the trained model. Our purpose was in achieving a very precise human skin detection. Despite the fact that skin detection represents a challenging problem because it can vary dramatically in its appearance due to many factors such as illumination conditions, pose variations, race, aging, and complex background. The proposed approach overcomes most of the difficulties in skin detection. Table 1 illustrate the architecture of the proposed skin detection CNN model. The Processing algorithm overview is shown in Fig.3.

3.4 rPPG signal

The green channel the RGB space transcribes better the PPG signal rather than the red and blue channels, and it provides a good signal noise ratio SNR [7] because the green light penetrates human skin better than blue light and it is absorbed by hemoglobin better than red light [29]. After performing the skin detection algorithm, we were able to extract rPPG signal by calculating the mean green values of pixels inside each frame, which concludes a temporal raw signal T_s :

$$T_s = [m_1(g), m_2(g), m_3(g), \dots, m_n(g)] \quad (3)$$

Where n denotes the index of frame and $m(g)$ correspond to the mean green value of pixels at each frame of the video.

$$m_i(g) = \left(\frac{\sum_{(x,y) \in ROI} I_g(x,y)}{|(x,y)|} \right) \quad (4)$$

Where $I_g(x,y)$ denotes the green channel intensity of each pixel, and $|(x,y)|$ represents the number of skin pixels for each frame. The source signals were passed by band-pass filtered (Hamming window filter with low- and high-frequency in the range (0.75Hz-4.5Hz) which corresponds to 45 bpm and 270 bpm respectively).

3.5 rPPG Signal Multilevel decomposition and heart rate estimation

Even though Short-Time Fourier Transform (STFT) is the most used tool for frequency analysis, but it is an inappropriate choice for non-stationary signals. Naturally, the physiological signals are non-stationary such as rPPG, STFT which provide a constant resolution at all frequencies and it loses time resolution in the spectral domain. Additionally, the fixed-length window is an inconvenient factor since the frequency behavior is not known, to overcome this shortcoming and to perform different resolutions analysis, we adopt the wavelet technique.

The basic formula of wavelet transform can be written as:

$$Wt(a,b) = \frac{1}{\sqrt{a}} \int_{(-\infty)}^{(+\infty)} X(t) \psi^*(a,b) dt \quad (5)$$

and the complex conjugate of the mother wavelet is defined as:

$$\psi^*(a,b) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) \quad (6)$$

where X(t) is the processed signal, a is the scaling parameter and b is the translation parameter. ψ^* is the mother wavelet, which plays the role of both a window and basis function, and is then scaled or shifted by the parameter a to regulate window length. Narrow windows ease detecting the presence of high-frequency components and wide windows facilitate extracting low frequencies components. In this work, we employed Discrete wavelet transform (DWT) because it offers a highly efficient wavelet representation. In Discrete wavelet transform, scale is only modified in the power of base 2 (a = 1, 2, 4, 8,...), and translation is also done in comparative relation to the scale. Furthermore, DWT is not shift-invariant. The DWT can be defined as:

$$Wt(j,k) = \frac{1}{\sqrt{2^j}} \int_{(-\infty)}^{(+\infty)} X(t) \psi^*(j,k) dt \quad (7)$$

For the discretized mother wavelet eq.6 Can be rewritten as:

$$\psi^*(j,k) = \frac{1}{\sqrt{2^j}} \psi \left(\frac{t-k2^j}{2^j} \right) \quad (8)$$

Where the scale and the shift parameters are commonly used as 2^j and $k2^j$. The signal is decomposed into approximation and detail coefficients in the first level. The decomposition for the remaining levels is realized using only the approximation coefficients. At each level, the decomposition is performed by using low-pass and high-pass filters. The output coefficients are named approximation coefficient and detail coefficient. These coefficients assure reconstruction of the original signal with a process called inverse discrete wavelet transform (IDWT). In this paper, the Coiflet wavelet [32] was chosen as the mother wavelet with a four levels decomposition is performed. As illustrated in the wavelet decomposition tree Fig.4. Coiflets were originally derived from the Daubechies wavelet. Coiflets wavelet family from 1st to 5th order, It uses many overlapping windows. It is based on six scaling and wavelet function coefficients, The Coiflet wavelet also follows the mirror technique. The signal decomposition was made with (coif1) and the reconstruction was made with the same wavelet (coif1) but we choose only the third and the fourth detailed coefficient (D3 and D4). The result is shown in Fig.5. The multi-resolution representation with the discrete wavelet transforms is illustrated

Table 1: The architecture of the proposed skin detection CNN model.

Layers No.	Type	Kernel size	No. Kernels	Stride	Outupt shape
Layer 1	Conv 1	1×1	16	1	$1 \times 1 \times 16$
Layer 2	Conv 2	2×2	32	1	$2 \times 2 \times 32$
Layer 3	Conv 3	1×1	64	1	$2 \times 2 \times 64$
Layer 4	Max pooling 1	2×2		0	$1 \times 1 \times 64$
Layer 5	Flatten 1				64
Layer 6	Dense 1		128		128
Layer 7	Dense 2		64		64
Layer 8	Dense 3		1		1

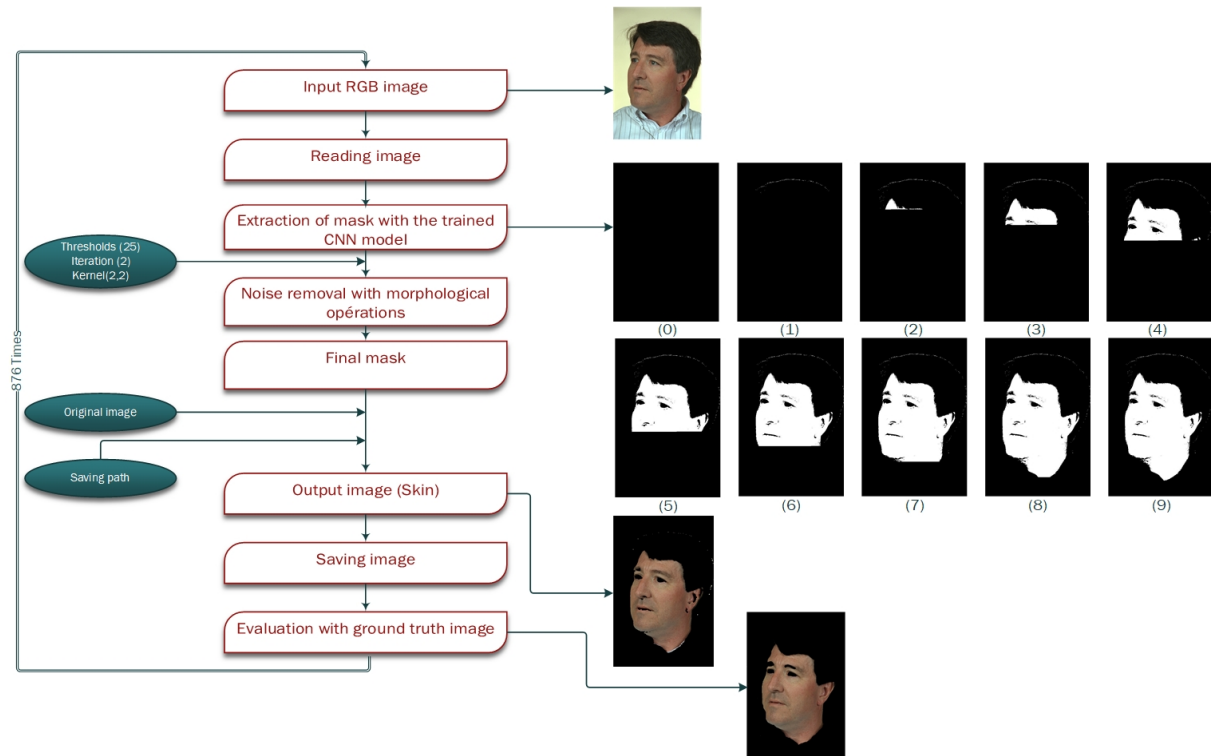


Figure 3: Framework of the proposed skin detection algorithm. The skin map is constructed progressively and displayed in steps 0-9. Each pixel of the image is predicted with the trained model.

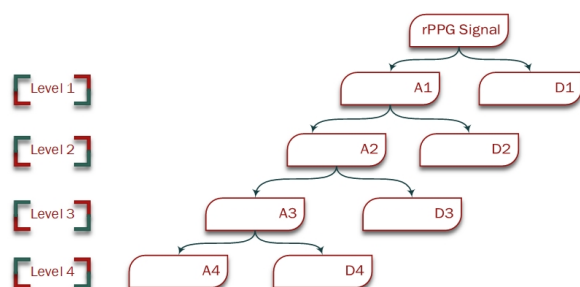


Figure 4: Discrete Wavelet Transform Decomposition Tree of four levels.

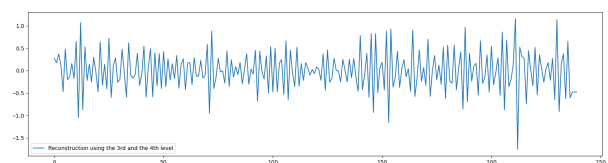


Figure 5: rPPG signal reconstruction with the third and the fourth detailed coefficient using discrete wavelet transform.

in spectrogram as seen in Fig.6. We can estimate

heart rate with peak detection. Peaks can then be simply identified using a diversity of peak detection algorithms but most of them require the selection of an appropriate threshold value to avoid incorrect peak detection. Once the peaks are detected the distance between the peaks

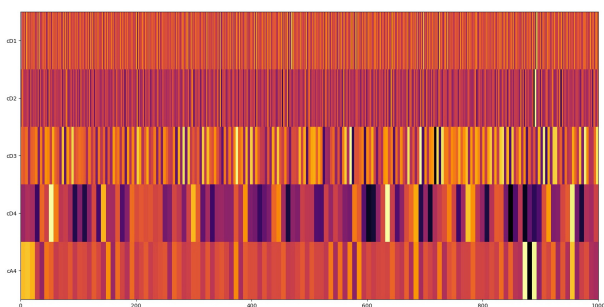


Figure 6: Spectrogram illustrating the four level decompositions.

can be used for the calculation of heart rate from the PPG signals using eq.9:

$$Heartbeatrate = 60 \times \frac{Fs}{intsys} \quad (9)$$

where F_s is the sampling rate and $intsys$ corresponds to the distance between the systolic peaks. Otherwise, for this aim, using frequency analysis, the frequency equivalent to the index with the highest spectral power is selected as an estimate for the HR. In the Fig.7 we see that the maximum power of the first detailed level coefficient is at frequency bin 1.602. This, in turn, gives HR $1.602 \times 60 = 96.12 \text{ bpm}$.

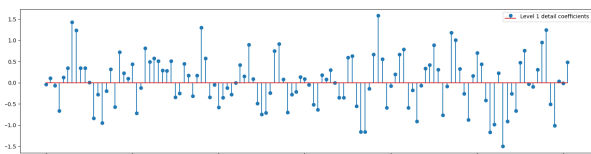


Figure 7: Spectral representation of detailed coefficient level 1

4 EXPERIMENTAL RESULTS

We conducted experiments to demonstrate the effectiveness of the proposed method on UBFC-rPPG signal datasets [30], which is publicly available, and is dedicated to rPPG algorithm evaluation. It is composed of 43 uncompressed videos, where each video is joined with Ground truth heart rate for evaluation. Each video have a duration of 60 seconds and corresponds to 30 frames per second (fps) with a resolution of 640×480 in 8-bits RGB format. Subjects recorded while playing a stressful game. It is worth noting that the rPPG signal is much noisier if we don't apply human skin detection and face alignment algorithm to fix the subject's voluntary movements. At the next stage, the decomposition-reconstruction operation is performed by using the wavelet transformation approach where *coiflet* wavelets gave a satisfying results. We generate different spectral representation at different level as shown in Fig. from 8 to 11. The

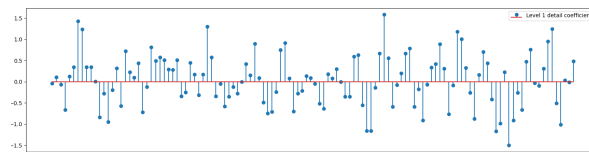


Figure 8: Spectral representation of detailed coefficient at level 1

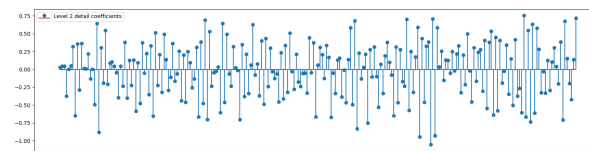


Figure 9: Spectral representation of detailed coefficient at level 2

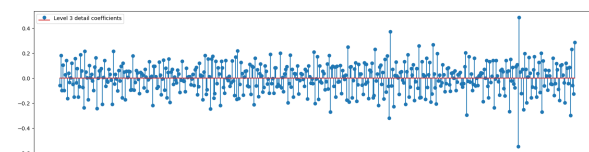


Figure 10: Spectral representation of detailed coefficient at level 3

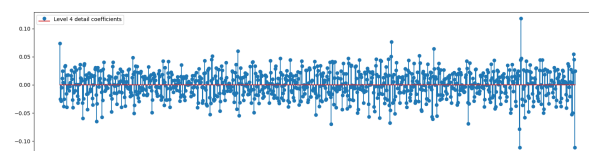


Figure 11: Spectral representation of detailed coefficient at level 4

performance of the proposed approach is compared to the state-of-the-art methods on the experimental dataset listed in Table 2. We have measured the accuracy in terms of Mean Absolute Error (MAE) and mean squared error (RMSE), our method generated the lower error, which relavate a high accuracy. We report results with two other metrics standard deviation of error (SD) and Pearson correlation coefficient (R), From the results we can determine that the proposed method showed consistent performance although the data acquisition scenarios (subject movement, closed eyes, luminosity variation). Lastly, the proposed Skindetection-DWT method was not tested yet on the most challenging use case scenarios that imply a lot of motion. This is mostly due to the fact that to the best of our knowledge there is no datasets for the motion robustness publicly available.

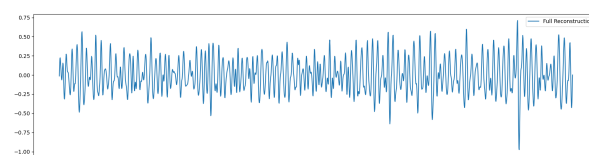


Figure 12: rPPG signal full reconstruction using all coefficients

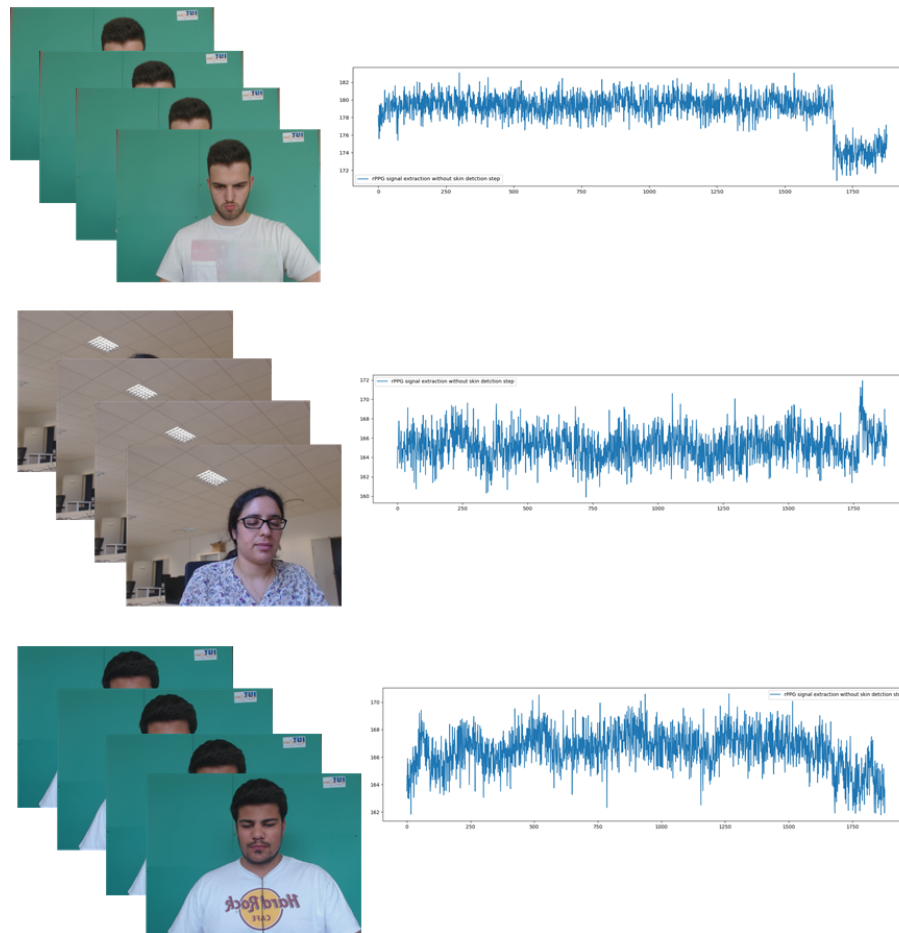


Figure 13: rPPG signals extraction without skin detection step

Table 2: Performance comparison between the proposed method and the state of the art methods of heart rate estimation on UBFC-rPPG dataset.

Method	HR(bpm)			
	SD	MAE	RMSE	R
GREEN[7]	20.2	10.2	20.6	-
ICA[8]	18.6	8.43	18.8	-
CHROM[18]	19.1	10.6	20.3	-
POS[19]	10.4	4.12	10.5	-
3D[26]	8.55	5.45	8.64	-
End-of-end(baseline)[28]	13.70	12.78	13.30	0.27
Meta-rPPG(inductive)[28]	14.17	13.23	14.63	0.35
Meta-rPPG(proto only)[28]	9.17	7.82	9.37	0.48
Meta-rPPG(synth only)[28]	11.92	9.11	11.55	0.42
Meta-rPPG(proto + synth)[28]	7.12	5.79	7.42	0.53
Ours	5.32	3.24	5.64	0.46

5 CONCLUSION

This paper discusses a novel approach for heart rate estimation using facial videos. The green channel yields the most prominent cardiac cycles and it was extracted from the forehead region. Since the most important problem associated with rPPG signal was variation in

skin tone, lighting condition, and subject's motion, the presence of these three factors represents many interferences and noise on the resulting signal. Accordingly, a previously proposed approach to skin detection is applied which relies on the strength of the convolutional neural network, also the face alignment al-

gorithm improves the efficiency of the proposed signal, then we adopted the multi-resolution representation with the discrete wavelet transform using spectrogram, so it keeps equally the time and the frequency information, and its facilities signal analysis. The signal decomposition and reconstruction were made with coiflet wavelet, the reconstruction of the signal was made only with the two final detailed coefficients.

The choice was relying on the visualization of the relation between the detailed coefficients of the wavelet transform.

In the final step we evaluate our heart rate estimation with the state of art methods on a public dataset UBFC-RPPG.

6 REFERENCES

- [1] G. Casalino, G. Castellano, V. Pasquadibisceglie, and G. Zaza: Contactless real-time monitoring of cardiovascular risk using video imaging and fuzzy inference rules. *Information*, vol. 10, no. 1, p. 9, 2019. 1.
- [2] Chaichulee, S., Villarroel, M., Jorge, J. o., Arteta, C., McCormick, K., Zisserman, A., Tarassenko, L. (2019): Cardio-respiratory signal extraction from video camera data for continuous non-contact vital sign monitoring using deep learning. *Physiological Measurement*, 40(11). <https://doi.org/10.1088/1361-6579/ab525c>
- [3] Kwon, S., Kwon, Y. T., Kim, Y. S., Lim, H. R., Mahmood, M., Yeo, W. H. (2020): Skin-conformal, soft material enabled bio-electronic system with minimized motion artifacts for reliable health and performance monitoring of athletes. *Biosensors and Bioelectronics*, 151(December 2019), 111981. <https://doi.org/10.1016/j.bios.2019.111981>
- [4] Z. Tong, X. Chen, Z. He, K. Tong, Z. Fang and X. Wang :Emotion Recognition Based on Photoplethysmogram and Electroencephalogram. *IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, Tokyo, Japan, 2018, pp. 402-407, doi: 10.1109/COMPSAC.2018.10266.
- [5] Sancho J, Alesanco a, Garcia J: Biometric Authentication Using the PPG: A Long-Term Feasibility Study. *Sensors (Basel)*. 2018
- [6] Salah, K.B., Othmani, M. Kherallah, M. A novel approach for human skin detection using convolutional neural network. *Vis Comput* (2021). <https://doi.org/10.1007/s00371-021-02108-3>
- [7] Verkruysse W, Svaasand LO, Nelson JS: Remote plethysmographic imaging using ambient light. *Optics Express*, 2008, 16(26): 21434-21445
- [8] Poh M-Z, McDuff DJ, Picard RW: Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 2010, 18(10): 10762-10774
- [9] M. Z. Poh, D. J. McDuff, and R.W. Picard: Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7-11, 2011
- [10] McDuff D, Gontarek S, Picard RW: Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera. *IEEE Trans Biomed Eng* 2014;61:2948-54.
- [11] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak: Measuring pulse rate with a webcam a non-contact method for evaluating cardiac activity. *federated conference on computer science and information systems (FedCSIS)*, pp. 405-410, IEEE, 2011.
- [12] J. Cheng, X. Chen, L. Xu, Z.J. Wang :Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition. *IEEE Journal of Biomedical and Health Informatics*, 2017, pp. 1422-1433
- [13] Sun Y, Hu S, Azorin-Peris V, Greenwald S, Chambers J, Zhu Y: Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *J Biomed Opt* 2011;16:77010
- [14] Tsouri GR, Kyal S, Dianat S, Mestha LK: Constrained independent component analysis approach to nonobtrusive pulse rate measurements. *J Biomed Opt* 2012;17:077011.
- [15] Holton BD, Mannapperuma K, Lesniewski PJ, Thomas JC: Signal recovery in imaging photoplethysmography. *Physiol Meas* 2013;34:1499-511
- [16] Christinaki E, Giannakakis G, Chiarugi F, Pediaditis M, Iatraki G, Manousos D, et al: Comparison of blind source separation algorithms for optical heart rate monitoring., pp. 339-42, *ICST*, 2014.
- [17] Qi, L., Yu, H., Xu, L., Mpanda, R. S., Greenwald, S. E. (2019): Robust heart-rate estimation from facial videos using Project ICA. *Physiological Measurement*, 40(8). <https://doi.org/10.1088/1361-6579/ab2c9f>.
- [18] De Haan, G., Jeanne, V.: Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering* 60(10),

- 2878-2886 (2013).
- [19] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang: Video-based heart rate measurement: Recent advances and future prospects, *IEEE Transactions on Instrumentation and Measurement*, 2018.
- [20] S. Ren, X. Cao, Y. Wei, and J. Sun :Face alignment at 3000 fps via regressing local binary features. In the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1685-1692, 2014.
- [21] Haque, M. A., Irani, R., Nasrollahi, K., Moeslund, T. B. (2016):Heartbeat Rate Measurement from Facial Video. *IEEE Intelligent Systems*, 31(3), 40-48.
- [22] R. Spetlik, J. Cech, V. Franc, and J. Matas: Visual heart rate estimation with convolutional neural network.08 2018
- [23] W. Chen and D. McDuff: Deepphys: Video-based physiological measurement using convolutional attention networks, in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 349-365, 2018
- [24] Niu, X., Shan, S., Han, H., Chen, X.: Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. *IEEE Transactions on Image Processing* (2019)
- [25] Yu, Z.; Li, X.; Niu, X.; Shi, J.; Zhao, G. AutoHR: A Strong End-to-end Baseline for Remote Heart Rate Measurement with Neural Searching. *arXiv 2020*, arXiv:2004.12292
- [26] Bousefsaf, F., Pruski, A., Maaoui, C. (2019): 3D convolutional neural networks for remote pulse rate measurement and mapping from facial video. *Applied Sciences (Switzerland)*, 9(20), 1-21. <https://doi.org/10.3390/app9204364>
- [27] Hsu, G. S. J., Xie, R. C., Ambikapathi, A. M., Chou, K. J. (2020): A deep learning framework for heart rate estimation from facial videos. *Neurocomputing*, 417, 155-166. <https://doi.org/10.1016/j.neucom.2020.07.012>
- [28] Lee, E., Chen, E., Lee, C. Y. (2020): Meta-rPPG: Remote Heart Rate Estimation Using a Transductive Meta-learner. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12372 LNCS, 392-409. <https://doi.org/10.1007/978-3-030-58583-9-24>
- [29] Lister, T., Wright, P. A., Chappell, P. H. (2012): Optical properties of human skin. *Journal of biomedical optics*, 17(9), 090901.
- [30] Bobbia S, Macwan R, Benezeth Y, Mansouri A, Dubois J.: Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognit Lett* 2017.
- [31] P. Viola, Ms. Jones : Rapid object detection using a boosted cascade of simple features, *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001, pp. I-I, doi: 10.1109/CVPR.2001.990517.
- [32] Shyh-Jier Huang and Cheng-Tao Hsieh :Coiflet wavelet transform applied to inspect power system disturbance-generated signals,in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 1, pp. 204-210, Jan. 2002, doi: 10.1109/7.993240.

