Journal of Industrial Information Technology and Application Vol.5 No.1 *ISSN(Online): 2586-0852*

Remote Heart Rate Estimation for Driver Drowsiness Detection

Hyunduk Kim*, Sang-Heon Lee, Myoung-Kyu Sohn
Division of Automotive Technology, DGIST, Daegu, Republic of Korea

Abstract. Remote heart rate (rHR) estimation, which aims to measure heart activities without any physical contact with the subject, is performed using remote photoplethys- mography (rPPG) and has great potential in many applications. In this paper, we introduce a remote heart rate estimation algorithm to detect driver drowsiness. The proposed method consists of two parts: an rPPGNet for PPG signal prediction from input video frames and an rHRNet for heart rate estimation from predicted PPG signal. Moreover, we apply a skin-based attention module and partition constraint to estimate more accurate rHR. To evaluate the performance of the proposed network, we train and test the proposed network on three public datasets.

Keywords; remote heart rate; remote photoplethysmography; skin-based attention module; spatio-temporal convolution network; driver drowsiness detection

1. Introduction

Heart rate is an important physiological signal that reflects the physical state of a person and widely applied to medicine, sports, and healthcare applications. Especially, the acquisition of accurate heart rate is useful for the driver drowsiness detection. Heart rate is usually obtained by electrocardiogram (ECG) and photoplethysmography (PPG) that require commonly contact with a subject's skin which may be inconvenient. Hence, numerous remote photoplethysmography (rPPG) estimation algorithms from face video have been introduced. In early study, most rPPG algorithms used handcrafted features, which achieved low signal-to-noise (SNR) for the dark skin and the age changes of skin. Recently, deep learning based rPPG estimation algorithms have been proposed to overcome these problem [1-3].

* Corresponding author: hyunduk00@dgist.ac.kr
Received: Jan. 28, 2021; Revised: Feb. 22, 2021; Accepted: Mar. 31, 2021

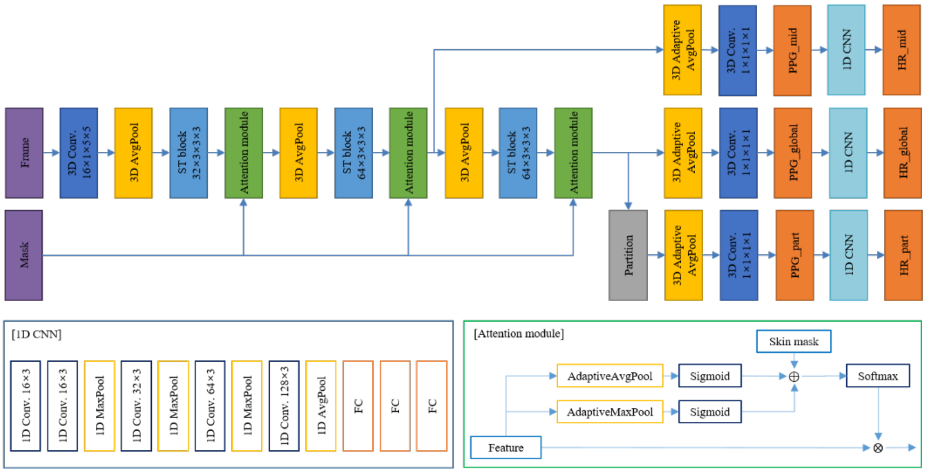

Fig 1. Architecture of the proposed network

2. Methodology

In this section, we describe the proposed remote heart rate estimation network, which is composed of an rPPGNet for PPG signal recovery from input video frames and an rHRNet for heart rate estimation from recovered PPG signal. First our algorithm detects face bounding boxes and landmarks from given an input video frame. Subsequently, affine face alignment based on detected face landmark is performed for each face. Moreover, we define a facial mask where the value of each pixel is equal to 1 if the corresponding pixel belong to the facial area, except eyes and mouth, otherwise, its value is considered equal to 0. The rPPGNet is composed of a spatio-temporal block (ST block) and skin-based attention module. It takes T-frame face images with RGB channel and masks as the inputs. Usually, spatial pooled RGB is projected into another color space for better representation of the PPG signal information. Moreover, temporal context based normalization was used to get rid of irrelevant information, such as noise caused by illumination or motion. Hence, similar to [1, 2], we use a spatio-temporal, which performs spatial convolution with 1×3×3 kernel followed by temporal convolution with t×1×1 kernel instead of convention 3D convolution with t×3×3 kernel. Additionally, we perform a skin-based attention module after the ST block to remove the influence of non-skin regions and enhance dominant PPG features. On the other hand, to estimate the heart rate from predicted PPG signal, we use 1-dimensional convolutional neural network (1D CNN), which is composed of several 1D convolutional layers and FC layers. Moreover, we also predict PPG signal using the mid-level and part-level features for stable convergence in training. Fig. 1 shows the architecture of the proposed network.

3. Experiments

We train and test the proposed remote heart rate estimation network on three public datasets, UBFC_DATASET, PURE, and LGI-PPGI. The UBFC-RPPG database [4] consists of two datasets which are focused for an rPPG estimation. The videos were captured using a Logitech C920 HD Pro webcam at 30fps with a resolution of 640x480 in uncompressed 8-bit RGB format. PURE dataset [5] consists of 10 persons performing different, controlled head motions in front of a camera. The videos were captured using an eco274CVGE camera at a 30 fps with a cropped resolution of 640x480 pixels. LGI-PPGI dataset [6] was collected by four scenarios, resting, head rotation, exercising in a gym, talking. The data collection consists of 20 male and 5 female. The camera device is selected as Logitech HD C270 webcam at 25fps. The reference pulse rate data in all datasets were captured using a finger clip pulse oximeter, CMS50E device, with a sampling rate of 60 Hz.

The proposed remote heart rate estimation system detects face bounding boxes and landmarks from given an input video frame using RetinaFace [7]. Subsequently, affine face alignment based on detected face landmark is performed for each face. Moreover, we perform face segmentation using BiSegNet [8] to extract a facial mask. The face image and mask are resized to 128×128 and 64×64 respectively. Finally, these are fed to the proposed network as inputs. The length of frame is set to 64. The proposed network is trained in NVIDIA Titan RTX using PyTorch. We pre-train rPPGNet and rHRNet individually for 20 epochs. After that we fine-tuning the full network for extra 10 epochs. The proposed network is optimized using TensorRT to improve processing time. To evaluate the performance of the proposed network on three public datasets we use three statistical metrics, such as the mean absolute error (MAE), the root mean square error (RMSE), the Pearson correlation coefficient (R). Table I shows the results of the proposed remote heart rate estimation network on three public datasets. Because the PURE dataset was collected controlled environment and the LGI-PPGI dataset was collected in various dynamic scenario, the proposed network achieved the best performance on PURE dataset and the worst performance on LGI-PPGI dataset.

TABLE I. RESULTS OF THE PROPOSED NETWORK ON THREE PUBLIC DATASETS

Dataset	MAE	RMSE	R
UBFC-DATASET	3.99	5.0885	0.9824
PURE	2.19	2.7042	0.9944
LGI-PPGI	5.94	6.6291	0.9902

4. Conclusions

In this paper, we proposed remote heart rate estimation system for driver drowsiness detection. To estimate heart rate from face video frames, we designed an rPPGNet based on spatio-temporal block and skin-based attention module to predict PPG signal, and an rHRNet based on 1-dimensional convolution to estimate heart rate value. Moreover we applied some strategies to improve convergence in training. To evaluate the performance of the proposed network, we trained and tested it on three public datasets. The experimental results show that the proposed network is accurate and can be applicable in practical application. In the future, we will optimize the proposed network to make it more accurate and faster and apply to driver drowsiness detection.

Acknowledgment

This work was supported by the DGIST R&D Program of Ministry of Science and ICT (21-IT-02).

References

[1] 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, pp. 151-160, 2019.

[2] O. Perepelkina, M. Artemyev, M. Churikova, and M. Grinenko, M, “HeartTrack: Convolutional neural network for remote video-based heart rate monitoring,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 288-289, 2020.

[3] E. Lee, E.Chen, and C. Y. Lee, “Meta-rppg: Remote heart rate estimation using a transductive meta-learner,” In European Conference on Computer Vision, pp. 392-409, 2020.

[4] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," Pattern Recognition Letters, vol. 124, pp. 82-90, 2017.

[5] R. Stricker, S. Müller, and H. M. Gross, “Non-contact video-based pulse rate measurement on a mobile service robot,” In The 23rd IEEE International Symposium on Robot and Human Interactive Communication, pp. 1056-1062, 2014.

[6] C. S. Pilz, S. Zaunseder, J. Krajewski, and V. Blazek, V, “Local group invariance for heart rate estimation from face videos in the wild,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1254-1262. 2018.

[7] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, “Retinaface: Single-stage dense face localisation in the wild,” arXiv preprint arXiv:1905.00641, 2019.

[8] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, “Bisenet: Bilateral segmentation network for real-time semantic segmentation,” In Proceedings of the European conference on computer vision, pp. 325-341, 2018.