applied sciences

MDPI

Article

Video-Based Contactless Heart-Rate Detection and Counting via Joint Blind Source Separation with Adaptive Noise Canceller

Kanghyu Lee [1], Junmuk Lee [2], Changwoo Ha [2], Minseok Han [2] and Hanseok Ko [1,*]

[1] Department of Video Information Processing, Korea University, Anam-dong, Sungbuk-gu, Seoul 136713, Korea; khlee@ispl.korea.ac.kr

[2] Hyundai Autron, 113-gil 12, Teheran-ro, Gangnam-gu, Seoul 06171, Korea; JunMuk.Lee@hyundai-autron.com (J.L.); ChangWoo.Ha@hyundai-autron.com (C.H.); MinSeok.Han@hyundai-autron.com (M.H.)

* Correspondence: hsko@korea.ac.kr

Received: 6 September 2019; Accepted: 12 October 2019; Published: 15 October 2019

Abstract: Driver assistance systems are a major focus of the automotive industry. Although technological functions that help drivers are improving, the monitoring of driver state functions receives less attention. In this respect, the human heart rate (HR) is one of the most important bio-signals, and it can be detected remotely using consumer-grade cameras. Based on this, a video-based driver state monitoring system using HR signals is proposed in this paper. In a practical automotive environment, monitoring the HR is very challenging due to changes in illumination, vibrations, and human motion. In order to overcome these problems, source separation strategies were employed using joint blind source separation, and feature combination was adopted to maximize HR variation. Noise-assisted data analysis was then adopted using ensemble empirical mode decomposition to extract the pure HR. Finally, power spectral density analysis was conducted in the frequency domain, and a post-processing smoothing filter was applied. The performance of the proposed approach was tested based on commonly employed metrics using the MAHNOB-HCI public dataset and compared with recently proposed competing methods. The experimental results proved that our method is robust for a variety of driving conditions based on testing using a driving dataset and static indoor environments.

Keywords: heart rate; cardiac signal; blind source separation; remote photoplethysmography

1. Introduction

The heart rate (HR) is one of the most important cardiac signals in the human body. The HR can be used to monitor medical emergencies and to determine general medical health. However, when measuring the HR, it is often difficult to apply measurement sensors in daily life outside of specific situations because of the restriction of human activity caused by attached sensor. Also, patients who have skin irritations in medical institutions such as hospitals can experience difficulty due to direct contact of the sensor with the skin. Overall, people experience a great deal of discomfort if any sensor is attached to their body parts. Non-contact HR measurement can overcome these problems. This approach is based on the fact that HR signals can be detected via the human skin [1], with optical changes on the skin surface visible due to blood perfusion caused by the heartbeat.

The optical variation caused by heartbeats is very subtle, meaning that the accurate detection of cardiac signals was studied in great detail. For example, in Reference [2], the green channel produced the strongest photoplethysmography (PPG) signal among the red/green/blue (RGB) image channels; thus, it was analyzed in the frequency domain as an HR feature [3]. In addition, Li et al. delicately cropped the face region and then applied an adaptive filter to determine the difference between the face

and the background regions based on the assumption that there is a motion correlation between the two regions [4]. However, although the green channel can reflect PPG variation, using this channel only can generate severe noise. To overcome this, blind source separation (BSS) was proposed. Poh et al. reported an HR estimation method using independent component analysis (ICA) with a webcam [5,6]. The webcam took RGB image sequences of the subject's face, and the RGB images were separated into three motion reflecting source components. Each source component was then analyzed by comparing it with the HR ground-truth, and the second component was selected, showing the best similarity fit with the ground-truth, as the desired HR source component. In Reference [7], joint blind source separation (advanced BSS), with the goal of source separation while maintaining consistent sources across multiple datasets, was applied to denoise facial signals.

Wu et al. attempted to detect HR signals by magnifying subtle source components based on changes to skin color due to the pulse [8]. However, applying this in real-world situations is limited because most signals are significantly larger than the color variation caused by the HR, meaning that it can only operate under controlled conditions, e.g., with no motion or changes in illumination.

An approach that linearly combines observation signals was introduced by Haan et al. [9]. It was based on the use of RGB channels under the assumption of a standardized skin tone, and several linear combination features extracted from the RGB signals were compared under equal conditions. Wang et al. then proposed a spatial pruning and temporal filtering method that improved upon previous linear combination features [10]. After the pruning procedure, the HR was extracted using principal component analysis (PCA) based on the assumption that a periodic signal such as the HR should have the highest variance because subject motion is occasional.

Deep-learning-based approaches were also proposed. Niu et al. built a feature map learning framework using synthetic spatiotemporal maps to overcome the problem of limited training data [11]. The estimator model pre-trained using ImageNet was trained on the synthetic data and fine-tuned using limited real facial data. Relying on the same deep-learning method, Tang et al. used a convolutional neural network (CNN) for region-of-interest (RoI) selection divided into three stages: face detection, tracking, and skin-region selection [12]. However, deep-learning-based methods require large datasets and significant computing resources and, if the testing environment changes, a new training dataset is needed.

In this paper, our goal is to achieve more accurate non-contact HR estimation under more challenging and practical conditions in a driving environment. With the rapid development of advanced driver assistance systems (ADASs) and automotive vehicles, ensuring driver and passenger safety is paramount. In particular, car accidents caused by medical emergencies afflicting the driver, such as a heart attacks, are a major concern. Using non-contact HR measurement, the status of a driver's heart can be monitored, and the driver can be notified when abnormal symptoms are detected. In addition, in automated vehicle systems, the vehicle can automatically be redirected to a nearby hospital. Although Kuo et al. previously attempted to employ the non-contact HR estimation approach in a driving environment, the extraction of pure HR signals was difficult, leading to poor performance [13]. Improving on previous research under driving conditions, a sliding window framework using a linear combination of RGB signals was proposed in Reference [14], achieving excellent performance and a rapid reaction time, making it suitable for driving conditions. In the same test framework as in Reference [14], our proposed method was tested in a public indoor environment, as well as under driving conditions, in order to verify the robustness of our multi-faceted approach.

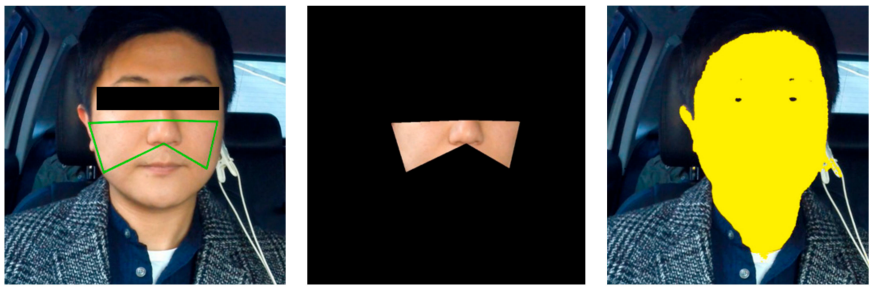
The structure of this paper is as follows: the overall proposed framework is described in detail in Section 2. In Section 3, two experiments are described: testing on a public indoor dataset to verify our method, and testing on a driving dataset containing rapid illumination changes, motion, and vibration. Finally, the experimental results are discussed in Section 4, and conclusions are drawn in Section 5.

2. Materials and Methods

While the signals in a typical video vary greatly, meaning that changes in pixel intensity can be confirmed visually, HR variation is very subtle. Therefore, the HR is likely to be affected by other source components, which need to be separated out. In this section, the framework for our approach to this problem is described. Firstly, the facial skin region was selected as an RoI. Source separation was then used to separate pure HR signals from other noisy signals using joint blind source separation (JBSS) and ensemble empirical mode decomposition (EEMD) [15,16]. Finally, the extracted source component was analyzed in the frequency domain using the power spectral density (PSD) analysis proposed by Welch [17].

2.1. Facial Regions of Interest

There are non-skin regions that cannot be used to observe the HR in the face, such as the eyes, eyebrows, and mustache. However, because typical face detector algorithms (e.g., the Viola–Jones face detector) include non-skin regions, they are not adequate for selecting the RoIs for HR estimation [18]. Previous HR estimation research also showed that the strength of PPG signals differs between different regions of the face, with the cheek and forehead regions tending to produce the strongest PPG signals [19]. Due to forehead occlusion depending on the hair style, the cheek region was selected as the RoI using a discriminative response map fitting (DRMF) face detector that detects 66 facial landmark points to extract only the region needed to detect the PPG signals (Figure 1) [20].



(**a**) (**b**) (**c**)

Figure 1. Region-of-interest (RoI) selection procedure: (**a**) face detection and cheek region cropping. (**b**) Selected skin region and (**c**) background region except for any skin region where the photoplethysmography (PPG) signal can be detected (the skin region is indicated in yellow).

However, face detection in every frame not only requires significant resources, but slight changes to the face or fading caused by light saturation or shadows can also lead to face detection failure. For this reason, our method adopted kernelized correlated filter (KCF) face tracking [21]. Even though face detection and tracking can be used to select the target facial region, regions that are not suitable for the extraction of the HR may be included within the tracked region. In order to obtain the skin region directly associated with HR variation, hue channels were employed as in Reference [14], which demonstrated reasonable performance.

2.2. HR Joint Blind Source Separation

Unlike the environments tested by previously proposed methods, driving conditions have a lot of dynamic signals. In order to extract the HR from these extreme conditions, sophisticated signal extraction techniques are required. In the present study, before the extraction process, each RGB channel was normalized with a mean unit variance of zero.

As a noisy component separation technique, our method adopted JBSS. Standard BSS is limited by the necessity to maintain the consistency of the separated source components across different datasets. For example, in ICA, which is a form of BSS, if the source components are separated in the frequency domain, there is a permutation problem in that the separated frequency bins have to be

aligned. This is caused by blindness to permutation, meaning ICA cannot automatically perform the alignment. Therefore, JBSS is suitable for datasets consisting of several subjects, and it was employed in this paper.

In the JBSS framework, it is assumed that groups of sources from multiple datasets are uncorrelated, while sources within each dataset are highly correlated. Based on this assumption, independent vector analysis (IVA) was adopted to separate the HR source component from other noise [22]. IVA was devised to solve the permutation problem with Kullback–Leibler divergence between the source and observation joint probability in ICA. The cost function C was described by

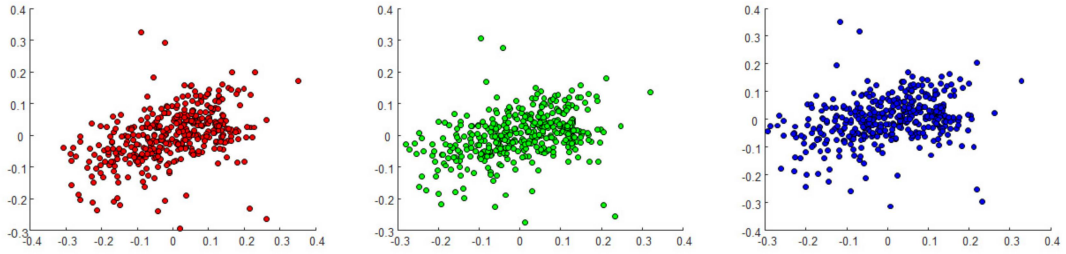
$$C = KL\left(p(s_1, \ldots, s_L) \parallel \prod_i q(s_i)\right), \tag{1}$$

where $p(s_1, \ldots s_L)$ and $\prod_i q(s_i)$ denote the probability density function and the marginal probability distribution function of the source component, respectively. The cost function converged to simultaneously minimize the entropy of all source components, and the mutual information was maximized within each source component. Detrending was then applied to each source component with the smoothing parameter $\lambda = 10$ to remove nonstationary components [23].

However, as mentioned previously, due to the fact that our algorithm was employed under driving conditions, the use of JBSS only was not sufficient to filter out noisy components. We assumed that there would be residual noise components after the JBSS process. The signals that undergo JBSS can be divided into face and background regions. The HR component should be present in the face region, while other noisy components are present in both regions. This assumption means that there should be some correlation between the face and the background regions. Based on this assumption, the noise components could be excluded, and the HR signal could be retained by subtracting the background signals from the face signals as follows:

$$HR_{comp} = S_{Face} - S_{Back}, \tag{2}$$

where S_{Face} and S_{Back} are the face and background signals after the JBSS process, respectively. The background region was defined as the uncorrelated region with skin that can emit a PPG signal. The background region is illustrated in Figure 1c. In order to validate this, we plotted the correlation between the face and background results from JBSS for each channel in Figure 2. Because of the presence of significant noise, the correlation coefficients were not strong, but they were positive because a large portion of the noise was present in both the face and background regions.



(**a**) $\rho = 0.4514$ (**b**) $\rho = 0.3762$ (**c**) $\rho = 0.3151$

Figure 2. Plots of the correlation and correlation coefficients (ρ) between the face (horizontal axis) and background (vertical axis) signals in each channel. Each dot denotes samples of the joint blind source separation (JBSS) results at every frame: (**a**) red channel; (**b**) green channel; (**c**) blue channel.

2.3. Normalized Least Squares Filter and Ensemble Empirical Mode Decomposition

Because the driving environment is dynamic, HR_{comp} cannot be used directly as a measure of the HR. In order to refine the extracted features via post-processing, an adaptive filter was applied.

The normalized least mean squares (NLMS) filter is one of the most useful adaptive filters used in noise cancellation [24]. The reasonable performance of an NLMS filter in a remote PPG experiment was previously reported [4]. Although Li et al. used NLMS to subtract the noisy background signal from the face signal, our method employed it to refine each HR_{comp} channel. NLMS can be represented as follows:

$$d_i = x_i^t w_i + n_{i,}, \tag{3}$$

where x_i^t and n_i are the input vector and white Gaussian noise with $\mathcal{N}\left(0, \sigma_v^2\right)$, which is independent from x_i^t, at the i_{th} iteration, respectively, and d_i is the signal to be refined. For each iteration, the optimal weight coefficient is updated and d_i is estimated as

$$w_i = w_{i-1} - \mu\left[\nabla^2 J[w_{i-1}]\right]^{-1} [\nabla J[w_{i-1}]]^*, \tag{4}$$

where * and ∇ denote the conjugate operation and derivation, respectively, and $J(w_{i-1})$ is the cost function with respect to the weight coefficient. The NLMS algorithm operated iteratively to minimize the error n_i, which was the difference between the desired signal d_i and $x_i^t w_i$. The weight coefficient w was initialized as zero at the beginning of the iterations.

Haan et al. proposed several HR feature signals using a linear combination of RGB channels, and Lee et al. analyzed the optimal feature signals in a dataset for various environments [8,15]. Based on previous results, the signal for each channel refined using NLMS was transformed into a *RoverG* signal, which was the ratio of the red and blue channels.

Ensemble empirical mode decomposition (EEMD) is a noise-assisted method for extracting the intrinsic mode function (IMF) from observed data. EEMD operates by iteratively averaging a series of trials in which white noise is added to the observed signal. As demonstrated in Reference [25], IMFs representing the source components of complicated observations can be decomposed after several iterations. Therefore, EEMD was employed on the *RoverG* signals to extract the IMF corresponding to the HR. In Reference [25], Chen et al. discovered that the fourth IMF of facial observation signals is the closest to the HR; thus, our approach also used the fourth IMF from *RoverG* via EEMD.

An example of these steps is presented in Figure 3. While the other approaches fluctuated over time, HR estimation using the NLMS filter followed the ground-truth in a relatively stable manner.

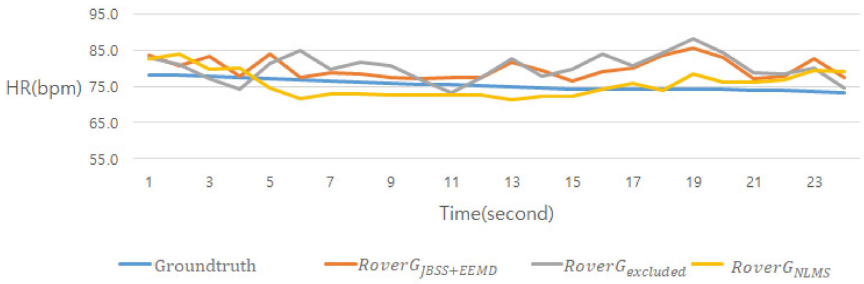


Figure 3. Heart rate (HR) estimation over time. $RoverG_{JBSS+EEMD}$ adds JBSS to Lee et al.'s method [15]. $RoverG_{excluded}$ employs the consistent correlation assumption by subtracting the noisy components in the background from the face signal. $RoverG_{NLMS}$ is our proposed method. EEMD—ensemble empirical mode decomposition; NLMS—normalized least mean squares.

2.4. HR Determination Using Sliding Window and Temporal Filtering

In the previous steps, we extracted the HR signal from an input image. However, even after considerable signal processing, detecting a subtle HR signal in a driving environment, which is characterized by severe interference, is challenging. In order to produce stable estimation results, our method used sliding windows of different lengths to derive several candidate estimation results at a single time point. The number of candidate estimates was set at seven, with a window length ranging from 8 to 14 s, and the Mahalanobis distance measure was employed for the final estimated HR to

exclude those results that differed the most from the others, as in the same framework as Reference [14]. The final estimation result was, thus, the average of the remaining candidate HRs. To transform the HR into a time-series signal, the Welch method was employed [13]. Finally, temporal filtering was applied to smooth the intermittently fluctuating final HR as follows:

$$HR^t = \frac{1}{s}\sum_{r=t-s}^{t-1} HR^r \ when \ HR^t - HR^{t-1} \geq \alpha, \tag{5}$$

where HR^t denotes the HR at time t. HR^t is determined by whether the threshold α of the difference between the current final HR at time t and the previous HR at time $t-1$ is exceeded.

3. Results

In this step, we evaluate our framework by firstly applying it to a public human–computer interaction (HCI) dataset in an indoor environment, and then by applying it to a driving dataset. The indoor dataset was used to validate our framework, with our method implemented in MATLAB 2019a with an Intel i7-4790 central processing unit (CPU) and 24 GB of memory.

3.1. Comparison of Features Using a Public Dataset

In order to determine the optimal algorithm structure, the performance of each step was analyzed. Our framework could be divided into three stages. Firstly, JBSS was applied to the detected RoIs and transformed into a *RoverG* feature signal, followed by HR IMF decomposition using EEMD. Secondly, the correlation between the background region and the face region was analyzed for each channel to exclude noise that was found in both regions before transformation into *RoverG*. Finally, NLMS was applied to each channel after the noise was excluded in the previous stage.

For stable analysis, the MAHNOB-HCI public indoor dataset was used [26]. The dataset contains videos of the frontal facial images of 27 subjects (15 females and 12 males) captured at a resolution of 780×580 and a frame rate of 61. It is the result of two experiments—emotion elicitation and implicit tagging—but we only used emotion elicitation because it was recorded in color and captured by the subject in various genres of motion situations reflecting every possible pulse change. The HRs of the subjects were recorded as electrocardiography (ECG) signals at 256 Hz while the subjects watched several video clips, and they were synchronized with the video.

In order to allow an accurate comparison with previously proposed methods, the metrics used in previous methods were employed in this paper. M_e and SD_e are the mean and standard deviation of Equation (6), respectively.

$$HR_{dif} = HR_{est} - HR_{gt}, \tag{6}$$

where HR_{est} and HR_{gt} denote the estimated result and ground-truth, respectively.

In addition, root-mean-square error (*RMSE*) and M_{eRate} were employed as shown below.

$$RMSE = \left[\frac{1}{N}\sum_{n=1}^{N}\left[HR_{est}[n] - HR_{gt}[n]\right]^2\right]^{1/2}, \tag{7}$$

$$M_{eRate} = \sum_{n=1}^{N}\left(\left|\frac{HR_{dif}|n|}{HR_{gt}|n|}\right|\right) \times 100, \tag{8}$$

where N denotes the length of the entire video in seconds, and n is time index for each second. Finally, the Pearson correlation coefficient r between the estimated HR and ground-truth was evaluated. Table 1 displays the performance of each stage of our framework using the MAHNOB-HCI dataset in terms of these metrics.

Table 1. Comparison of the performance for each stage of the proposed framework using the MAHNOB-HCI dataset (best performance in bold). RMSE—root-mean-square error; JBBS—joint blind source separation; EEMD—ensemble empirical mode decomposition; NLMS—normalized least mean squares.

Feature	$M_e(SD_e)$	*RMSE*	M_{eRate}	*r*
$RoverG_{JBSS+EEMD}$	−3.36 (5.10)	7.39	9.29%	0.10
$RoverG_{excluded}$	**−1.22** (4.16)	7.49	9.58%	0.06
$RoverG_{NLMS}$	−3.36 (**3.33**)	**5.31**	**6.57%**	**0.72**

$RoverG_{JBSS+EEMD}$, $RoverG_{excluded}$, and $RoverG_{NLMS}$ represented stages 1, 2, and 3, respectively. Their performance was somewhat consistent with the overall metrics. However, the performances of $RoverG_{JBSS+EEMD}$ and $RoverG_{excluded}$ were poor in terms of Pearson correlation ***r*** except for the other metrics (e.g., M_e, SD_e, *RMSE*, M_{eRate}). This is because the estimated HR fluctuated inconsistently, leading to a difference between the ground-truth and the estimated HR. Based on the experimental results, $RoverG_{NLMS}$ was verified to be the superior framework.

3.2. Validation Using the MAHNOB-HCI Indoor Dataset

In this section, our proposed method is compared with previous approaches using the MAHNOB-HCI public dataset. The evaluation metrics used in the previous section were also employed (Table 2).

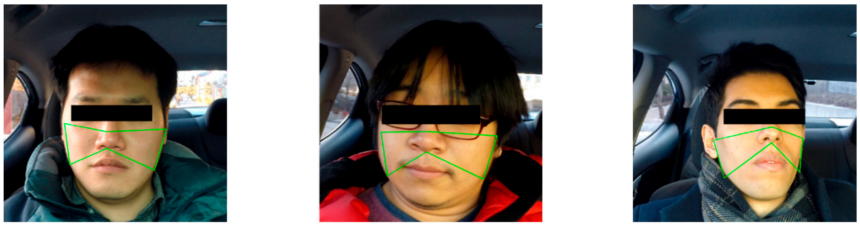
Table 2. Comparison of the performance of our proposed framework with other approaches using the MAHNOB-HCI dataset (best performance in bold).

Method	$M_e(SD_e)$	*RMSE*	M_{eRate}	*r*
Poh (2010)	−8.95 (24.3)	25.9	25.0%	0.08
Poh (2011)	2.04 (13.5)	13.6	13.2%	0.36
Li (2014)	−3.30 (6.88)	7.62	6.87%	0.81
Tulyakov (2016)	3.19 (5.81)	6.23	5.93%	**0.83**
Lee (2018)	**0.80** (3.35)	**3.26**	**3.68%**	0.75
Ours	−3.36 (**3.33**)	5.31	6.57%	0.72

The algorithm that produced the best performance differed for each evaluation metric. Even though our method only outperformed the others in terms of SD_e, it did not demonstrate a notably worse performance compared to the best algorithms based on the other metrics. In contrast to earlier studies, recent frameworks were steadily refined and improved; given that the performance of our proposed method falls within the top three based on the metrics overall, it can be concluded that our method is fairly stable and that it was validated by the public dataset.

3.3. Demonstration Using a Driving Dataset

Because our framework was designed to be employed under driving conditions, a real driving dataset was collected. This dataset included 19 subjects, both male and female, in their 20s to 30s. The test subjects were from the Middle East and Asia (Korean, Chinese, and Taiwanese), with some wearing glasses (Figure 4).



(**a**) (**b**) (**c**)

Figure 4. Examples of the appearance of the subjects used in the driving dataset: (**a**) Asian; (**b**) Asian with glasses; (**c**) Middle Eastern.

The driving dataset was captured using a GoPro HERO3+, with the camera fixed on the windscreen looking at the subject's face. The video was recorded at a resolution of 1920 × 1080 and a frame rate of 30. The ground-truth was obtained using a contact-based pulse sensor (the MP507 model from MEK) attached to each subject's earlobe and synchronized with the captured video data. The subjects sat in the passenger seat to ensure their safety. In order to simulate the most common movements of real drivers, the subjects were asked to rotate their head and to look up, down, left, and right. In addition, in order to determine whether our method was able to accurately detect a change in the pulse while driving, the subjects were asked to run on a steep hill before entering the vehicle. In all driving situations, subjects were asked to freely make motions with no special restrictions, to set the environment as close as possible to the actual driving conditions. The driving course was subject to a lot of shadows, rapid illumination changes, and vibrations caused by the unevenness of the road surface. As stated in Appendix A, we collected driving dataset on research ethics.

Table 3 presents the experimental results compared with recent methods for the driving dataset. The compared methods were selected based on the latest results and major algorithms, and these methods were re-implemented to obtain the results on the same driving dataset. Poh (2011) and Zhao (2013) were some of the backbones of related research, but they did not show robustness in the driving dataset in dynamic conditions. Unlike the previous methods, Cheng (2017) showed significantly higher results in M_e but did not show strength in other metrics. On the other hand, Lee (2018) showed an overall high performance compared with previous methods in several metrics. However, in terms of r, there was a slight increase compared to the previous results, but still no reliable performance. On the other hand, our method demonstrated a slightly lower performance using the public dataset compared to recently published approaches. Our method not only showed the best performance in most metrics, but experimenting on this driving dataset (e.g., condition) also illustrated that our method is robust in various driving and environmental situations.

Table 3. Experimental results for the driving dataset (best performance in bold).

Method	$M_e(SD_e)$	*RMSE*	M_{eRate}	r
Poh (2011)	20.73(35.45)	40.10	29.56%	−0.21
Zhao (2013)	14.85(16.43)	21.76	20.12%	−0.01
Cheng (2017)	4.48(17.93)	17.93	20.43%	−0.06
Lee (2018)	2.66(**2.42**)	4.74	4.15%	0.26
Ours	**−1.64** (3.70)	**3.94**	**3.97%**	**0.66**

Figure 5 displays the relationship between the ground-truth and the estimated HR results. The estimated HR followed the ground-truth closely and did not fluctuate dramatically due to external factors.

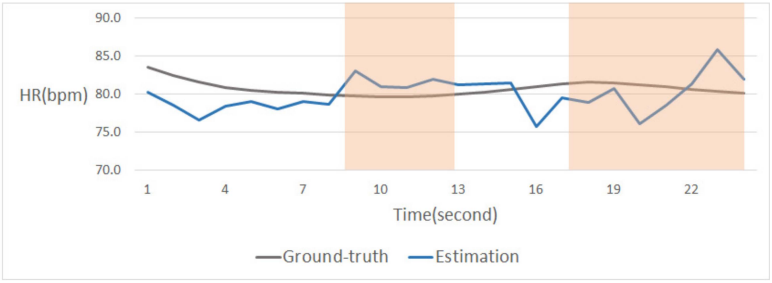


Figure 5. Relationship between the ground-truth HR and the estimated HR results. Vibrations caused by irregularities on the road surface occurred consistently over the whole period. The red zones represent periods during which there was a rapid illumination change.

In order to assess the precision of our method, we used Bland–Altman plots on driving subject samples, which represent a statistical method for comparing two kinds of data, for the results from the driving dataset (Figure 6). The Bland–Altman plot was proposed to quantify the agreement

between two measurements by Altman and Bland. Not only is it a simple way to compare the bias between the mean differences, but it also analyzes the deviation within a 95% confidence interval. The Bland–Altman agreement is calculated as

$$A = \frac{1}{N}\sum_{i=1}^{n} a_i \times 100, \; with \; a_i = \begin{cases} 1, & if \left|HR_{est} - HR_{gt}\right| > 1.96 \times \sigma \\ 0, & if \left|HR_{est} - HR_{gt}\right| < 1.96 \times \sigma \end{cases}, \tag{9}$$

where N and σ denote total number of measurements and the standard deviation between HR_{est} and HR_{gt}, respectively. The vertical axis denotes the difference between the ground-truth and estimated results, and the horizontal axis denotes the average of both at each second. The plots had a low bias and standard deviation, and illustrated high agreement.

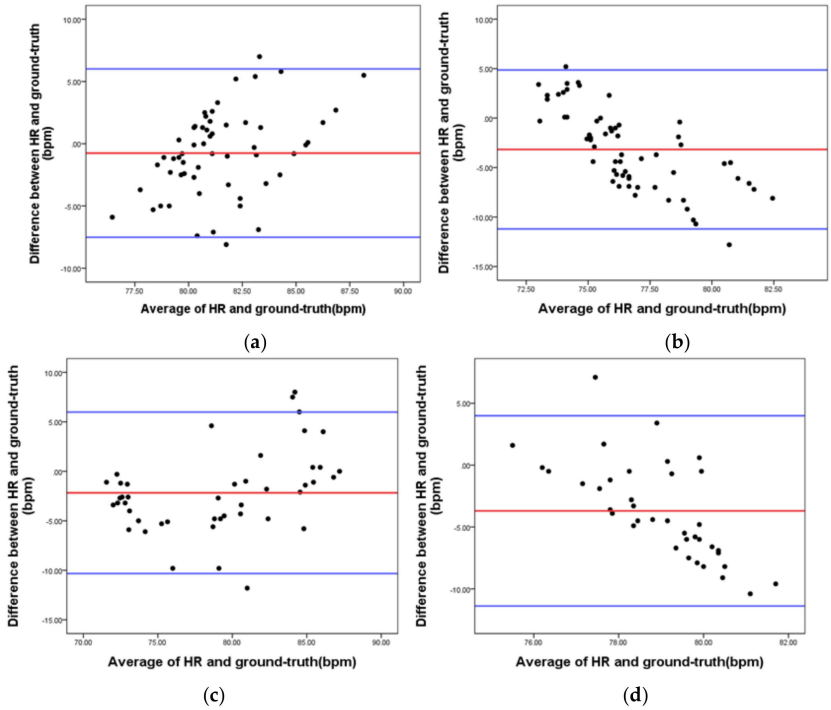


Figure 6. Bland–Altman plots for the driving dataset analyzed at a 95% confidence level. The above four plots were selected to show the stable results for various human heart-rate cases. The red and blue lines denote the means and standard deviations of the data points, respectively. Agreement levels were as follows: (**a**) 96.49%, (**b**) 96.72%, (**c**) 93.33%, and (**d**) 97.37%.

It may not be as apparent that the experimental results in Figure 6 followed the tendency of the HR, because the subjects included in Figure 6 ranged from the relatively short interval of a fast HR to the long interval of the normal state. However, even though it was not sharply revealed by the Bland–Altman plot, not only were the differences between the estimated values and the ground-truth less than 5 bpm, but the highest performance was also obtained based on the quantitative numerical values shown in Table 3. Moreover, as shown in Table 3, the Pearson correlation (r), which is related to tendency, yielded the highest performance using the proposed algorithm compared to other algorithms.

4. Conclusions

In this paper, we proposed a heart-rate estimation framework that operates under driving conditions to prevent car accidents caused by acute heart disease. Because the driving environment has a lot of noise that needs to be excluded in order to extract a driver's HR, various signal processing techniques are employed in our method. Firstly, JBSS is applied to an input image to separate the source components. Secondly, the noise components found in both the face and background regions are

excluded by analyzing the correlation between the two regions. Thirdly, an adaptive noise cancelling filter is employed on the results of the previous step to remove the remaining noise using NLMS. Fourthly, the refined signal is transformed into a *RoverG* feature, and the IMF corresponding to the HR is extracted using EEMD. Finally, the results are analyzed in the frequency domain, and temporal filtering is applied to smooth intermittently fluctuating values.

In previous methods, the environment within which the HR was measured was specifically constructed so that the proposed algorithm could work well, and most of these methods were tested indoors. However, tests conducted in such environments do not accurately reflect real-world suitability and robustness. To demonstrate the robustness of our method, we tested it on an unrestricted driving dataset containing illumination changes and motion. Prior to testing on the driving dataset, we divided our framework into several stages and analyzed the performance of each stage. We then tested our framework on the publicly available dataset NAHNOB-HCI to verify that our method can be generalized and that it is not specific to our driving dataset. We then demonstrated our proposed method using the driving dataset, which included various challenging disturbances such as illumination change, vibration, and head rotation. Our approach produced a stable performance in this environment, and there was no significant degradation in performance compared to the indoor environment dataset.

5. Patents

The authors have a patent entitled "method for estimating the condition of a driver" (Korea application number 10-1988581).

Author Contributions: Conceptualization, K.L.; methodology, K.L.; software, K.L.; investigation, K.L.; writing—original draft preparation, K.L.; writing—review and editing, H.K.; supervision, H.K.; funding acquisition, J.L., C.H., and M.H.

Funding: This work was funded by Hyundai Autron Co., Ltd. and was technically supported by the Application SW Development Team and the R&D Innovation Team.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Authors complied with the WMA Declaration of Helsinki when collecting the driving dataset.

References

1. Huelsbusch, M.; Blazek, V. Contactless mapping of rhythmical phenomena in tissue perfusion using PPGI. In *Medical Imaging 2002: Physiology and Function from Multidimensional Images*; International Society for Optics and Photonics: San Diego, CA, USA, 2002; Volume 4683.
2. Verkruysse, W.; Svaasand, L.O.; Nelson, J.S. Remote plethysmographic imaging using ambient light. *Opt. Express* **2008**, *16*, 21434–21445. [CrossRef] [PubMed]
3. Zhao, F.; Li, M.; Qian, Y.; Tsien, J.Z. Remote measurements of heart and respiration rates for telemedicine. *PLoS ONE* **2013**, *8*, e71384. [CrossRef] [PubMed]
4. Li, X.; Chen, J.; Zhao, G.; Pietikainen, M. Remote heart rate measurement from face videos under realistic situations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
5. Poh, M.-Z.; McDuff, D.J.; Picard, R.W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express* **2010**, *18*, 10762–10774. [CrossRef] [PubMed]
6. Poh, M.-Z.; McDuff, D.J.; Picard, R.W. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans. Biomed. Eng.* **2010**, *58*, 7–11. [CrossRef] [PubMed]
7. Cheng, J.; Chen, X.; Xu, L.; Wang, Z.J. Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 1422–1433. [CrossRef] [PubMed]

8. Wu, H.-Y.; Rubinstein, M.; Shih, E.; Guttag, J.; Durand, F.; Freeman, W. *Eulerian Video Magnification for Revealing Subtle Changes in the World*; Association for Computing Machinery: Los Angeles, CA, USA, 2012.
9. De Haan, G.; Jeanne, V. Robust pulse rate from chrominance-based rPPG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2878–2886. [CrossRef] [PubMed]
10. Wang, W.; Stuijk, S.; de Haan, G. Exploiting spatial redundancy of image sensor for motion robust rPPG. *IEEE Trans. Biomed. Eng.* **2014**, *62*, 415–425. [CrossRef] [PubMed]
11. Niu, X.; Han, H.; Shan, S.; Chen, X. Synrhythm: Learning a deep heart rate estimator from general to specific. In Proceedings of the IEEE 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018.
12. Tang, C.; Lu, J.; Liu, J. Non-contact heart rate monitoring by combining convolutional neural network skin detection and remote photoplethysmography via a low-cost camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018.
13. Kuo, J.; Koppel, S.; Charlton, J.L.; Rudin-Brown, C.M. Evaluation of a video-based measure of driver heart rate. *J. Saf. Res.* **2015**, *54*, 55-e29. [CrossRef] [PubMed]
14. Lee, K.; Han, D.K.; Ko, H. Video Analytic Based Health Monitoring for Driver in Moving Vehicle by Extracting Effective Heart Rate Inducing Features. *J. Adv. Transp.* **2018**, *2018*, 8513487. [CrossRef]
15. Anderson, M.; Adali, T.; Li, X. Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis. *IEEE Trans. Signal Process.* **2011**, *60*, 1672–1683. [CrossRef]
16. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [CrossRef]
17. Welch, P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **1967**, *15*, 70–73. [CrossRef]
18. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), Kauai, HI, USA, 8–14 December 2001; Volume 1, pp. 511–518.
19. Kumar, M.; Veeraraghavan, A.; Sabharwal, A. Distance PPG: Robust non-contact vital signs monitoring using a camera. *Biomed. Opt. Express* **2015**, *6*, 1565–1588. [CrossRef] [PubMed]
20. Asthana, A.; Zafeiriou, S.; Cheng, S.; Pantic, M. Robust discriminative response map fitting with constrained local models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013.
21. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 583–596. [CrossRef] [PubMed]
22. Kim, T.; Eltoft, T.; Lee, T.-W. Independent vector analysis: An extension of ICA to multivariate components. In Proceedings of the International Conference on Independent Component Analysis and Signal Separation, Charleston, SC, USA, 5–8 March 2006; Springer: Berlin/Heidelberg, Germany, 2006.
23. Tarvainen, M.P.; Ranta-Aho, P.O.; Karjalainen, P.A. An advanced detrending method with application to HRV analysis. *IEEE Trans. Biomed. Eng.* **2002**, *49*, 172–175. [CrossRef] [PubMed]
24. Song, I.; Park, P. A normalized least-mean-square algorithm based on variable-step-size recursion with innovative input data. *IEEE Signal Process. Lett.* **2012**, *19*, 817–820. [CrossRef]
25. Chen, D.-Y.; Wang, J.J.; Lin, K.Y.; Chang, H.H.; Wu, H.K.; Chen, Y.S.; Lee, S.Y. Image sensor-based heart rate evaluation from face reflectance using Hilbert–Huang transform. *IEEE Sens. J.* **2014**, *15*, 618–627. [CrossRef]
26. Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* **2011**, *3*, 42–55. [CrossRef]