WILEY | Hindawi

Research Article

Video Analytic Based Health Monitoring for Driver in Moving Vehicle by Extracting Effective Heart Rate Inducing Features

Kanghyu Lee,[1] **David K. Han,**[2] **and Hanseok Ko** [1,3]

[1]*Department of Video Information Processing, Korea University, Anam-dong, Sungbuk-gu, 136713 Seoul, Republic of Korea*
[2]*Information Science Division, ARL, Adelphi, MD 20783, USA*
[3]*School of Electrical Engineering, Korea University, Anam-dong, Sungbuk-gu, 136713 Seoul, Republic of Korea*

Correspondence should be addressed to Hanseok Ko; hsko@korea.ac.kr

We propose a novel remote heart rate (HR) estimation method using facial images based on video analytics. Most of previous methods have been demonstrated in well-controlled indoor environments. In contrast, this paper proposes a practical video analytic framework under actual driving conditions by extracting key HR inducing features. In particular, when cars are driven, effective and stable HR estimation becomes challenging as there are many dynamic elements, such as rapid illumination changes, vibrations, and ambient lighting that can exist in the vehicle interior. To overcome those disturbances of HR estimation, the driver face region is first detected and cropped to the region of interest (RoI). Second, the components related to HR are extracted from mixed noisy components using ensemble empirical mode decomposition (EEMD). Finally, the extracted signal is analyzed in frequency domain and smoothed with temporal filtering. To verify our approach, the proposed method is compared with recent prominent methods employing a public HCI dataset. It has been demonstrated that the proposed approach delivers superior performance under driving conditions using Bland-Altman plots.

1. Introduction

Traffic accidents occur due to acute driver heart rate (HR) disease. These accidents can develop into dangerous situations that threaten not only the driver but also the lives of others. If the driver's HR is known in advance, it is possible to prevent the accident by judicially controlling the vehicle. Methods, such as wired contact sensors, have been proposed to measure the driver's HR. However, due to the invasive nature of the in situ sensors, such methods have not gained much interest. For less intrusive and accurate measurements of driver HR, this research proposes a remote estimation method based on a video analytic framework focused on capturing key HR inducing features.

Nowadays, some systems monitor a driver's condition by placing a camera on the vehicle frame or the windshield of the vehicle. Furthermore, since image based remote HR estimation has been shown possible [1], a series of related studies have been subsequently proposed.

Poh et al. demonstrated the HR estimation technique by separating the observed signal into independent source signals [2, 3]. A bandpass filter is applied to each of the signal and the result was analyzed in frequency domain. Zhao et al. proposed an estimation technique for respiration as well as HR using a delay matrix [4]. Another study estimated the pulse rate by amplifying the frequency of the signal using minute movements of the face associated with vibration associated with human pulses [5]. However, these methods can be successful only if the subject is in a static state and any changes in the environment are limited.

In [6], Li et al. proposed a new approach which made slightly different assumptions compared to the previous studies. By assuming that light change to the face is the same as the light change to the background area, HR can be estimated through the difference between these two areas. Wang et al. demonstrated a pruning architecture using CHROM that removes pixels with values that do not correspond to skin tones and pixels distorted by motion [7, 8]. Also based on

CHROM, Tulyakov et al. improved on previous methods by cropping and warping certain facial regions using a self-adaptive matrix [9]. Similar to the assumption of [6], Xu et al. analyzed the background region as the noise reference to the facial region and then applied blind source separation approach. Even though the result was shown quite impressive, the variation of the result was large, making stable detection difficult in a dynamic environment [10]. Cheng et al. also applied an approach to Poh et al. by extracting unique pulse signals through ensemble empirical mode decomposition (EEMD) for the input signal analyzed by joint blind source separation (JBSS) based on the same assumptions [11, 12]. On the other hand, Huan et al. analyzed the input signal using JBSS in a similar way but exploited correlations between them by dividing the face region into several subregions and applied it to a learning based method [13]. However, in the test data, obstruction caused by wires and tapes in skin region was suggested as a challenging point and there was no significant innovation since the authors did not consider rapidly changing environment. In [14], a deep learning based remote photoplethysmography (rPPG) approach that detects skin regions using convolutional neural network (CNN) was proposed. Although it was a unique method of applying deep learning, there is a disadvantage that data must be learned in a new environment every time in terms of machine learning.

These previous studies have steadily improved the technology, but most estimate pulses from a distance in an indoor environment. In each of these papers, experiments have used well-controlled data and been conducted in controlled environments. Only few studies have addressed extreme illumination changes and vibrations associated with automotive environments. Although, Kuo et al. proposed an HR estimation framework under driving conditions, the approach was conventional and suffered very poor performance [15]. In this paper, the proposed method shows stable HR estimation results in indoors as well as in a wide range of outdoor moving environments.

The structure of this paper is as follows. The framework of the proposed method is shown in detail in Section 2. In Section 3, our proposed algorithm is applied to a public human-computer-interface (HCI) dataset to verify its validity and the results compared with those of previous studies. The experimental results of our driving dataset are presented by a Bland-Altman plot. Finally, the conclusions are discussed in Section 3.

2. Proposed Method

In this section, the proposed method can be divided into three stages: (1) region of interest (RoI) selection, (2) pulse signal extraction, and (3) power spectral density (PSD) analysis and temporal filtering. The overall flow is illustrated in Figure 1.

2.1. Region of Interest Selection. Kumar et al. demonstrated that the color changes due to pulsation are different for each region of the face, and as a result, the forehead and cheek region represent the strongest PPG signal [16]. Based on

this result, the cheek region is selected as the RoI. While the forehead region depends on hair style, the cheek region provides robust features insensitive to facial expressions. In order to extract the RoI, unnecessary background regions are excluded based on the assumption that the driver's facial position is somewhat fixed. A total of 66 facial landmark points are extracted for the remaining facial regions by using discriminative response map fitting (DRMF) to extract both cheek regions as illustrated in Figure 2 [17].

However, in the case of varying driving situations, not only the rotation and movement of the face but also face detection per video frame slows the processing speed, making the camera-based method ineffective for real-time HR estimation. To mitigate such problems, face tracking is applied using a kernelized correlated filter (KCF) [18]. Therefore, facial landmark point extraction is performed only at the first frame, after which the detected cheek region is tracked.

Nevertheless, the tracked RoI may still be incomplete. If the face is rotated or shaken, a background region may be included within the tracked RoI. Furthermore, as the vehicle runs, numerous illumination changes can cause skin region pixel values saturated such that the HR signal disappears. To prevent this, a skin detection scheme is employed using the hue channel in the HSV color model as in

$$p_{ij} = \begin{cases} skin, & h < \tau \\ non-skin, & h \geq \tau, \end{cases} \tag{1}$$

where p_{ij} denotes the pixel value in ith row and jth column and h denotes the hue channel value. In our method, we set the threshold of 90 for the hue channel as τ and selected pixels less than 90 as skin regions. The value was determined to be the best choice for the set of facial image data collected and used in this study. According to the work by [19], a value of threshold was used for the similar purpose.

2.2. Feature Extraction and Source Separation. Assuming that the ambient light signal has properties such as white noise of uniform magnitude in all frequency bands, the observed signal S from the RoI can be described as

$$S = S_{HR} + S_{motion} + S_{illumination} + S_{ambient}, \tag{2}$$

where S_{motion}, $S_{illumination}$, and $S_{ambient}$ are motion-induced changes, illumination changes, and changes in the ambient light signal, respectively. As shown in Figure 3, the frequency of illumination changes and vibration in the automotive driving environment appears in a fairly low frequency band compared with HR. Thus, the noise signals caused by illumination change and vibration can be significantly excluded using bandpass filtering. However, given the assumption that ambient light is white noise, it cannot be easily filtered out by the bandpass filter, and so may interfere with the HR signal. Therefore, it is necessary to extract the prominent feature signal of the HR and to separate it into each source signal from a feature that contains various components.

Based on the property that the signal of PPG is different for each channel, the *RoverG* feature that maximizes HR can
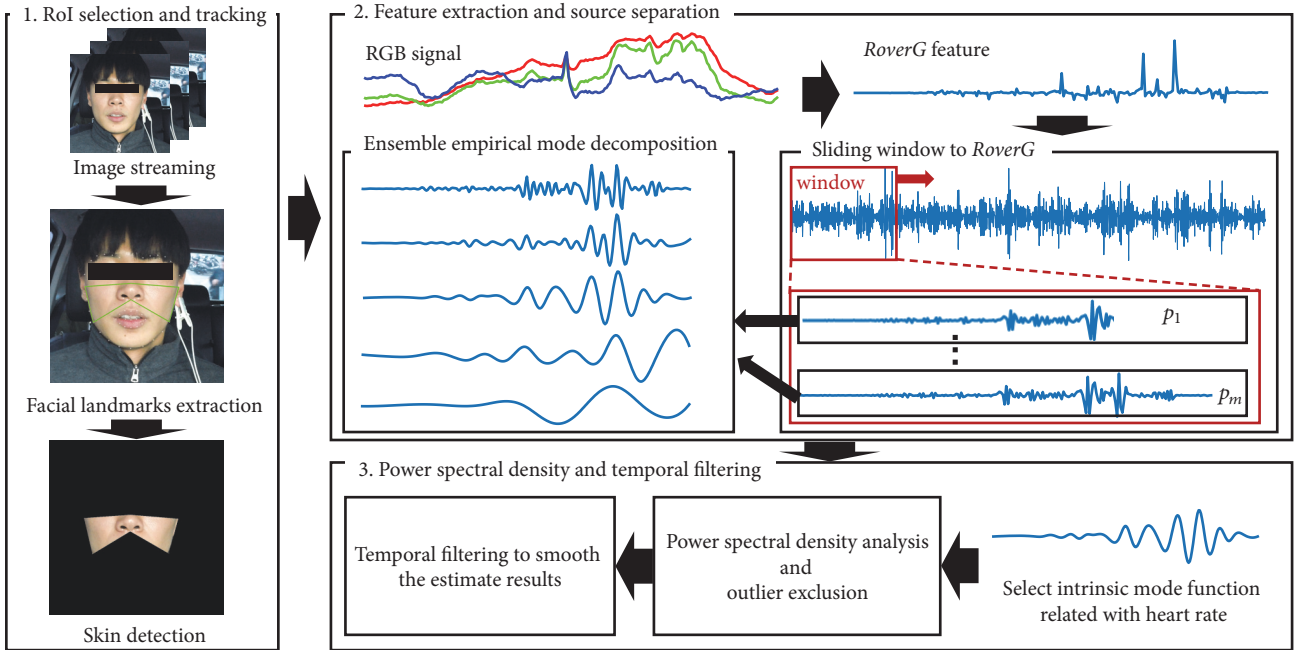
FIGURE 1: Conceptual overview of the proposed heart rate estimation method under driving environment.
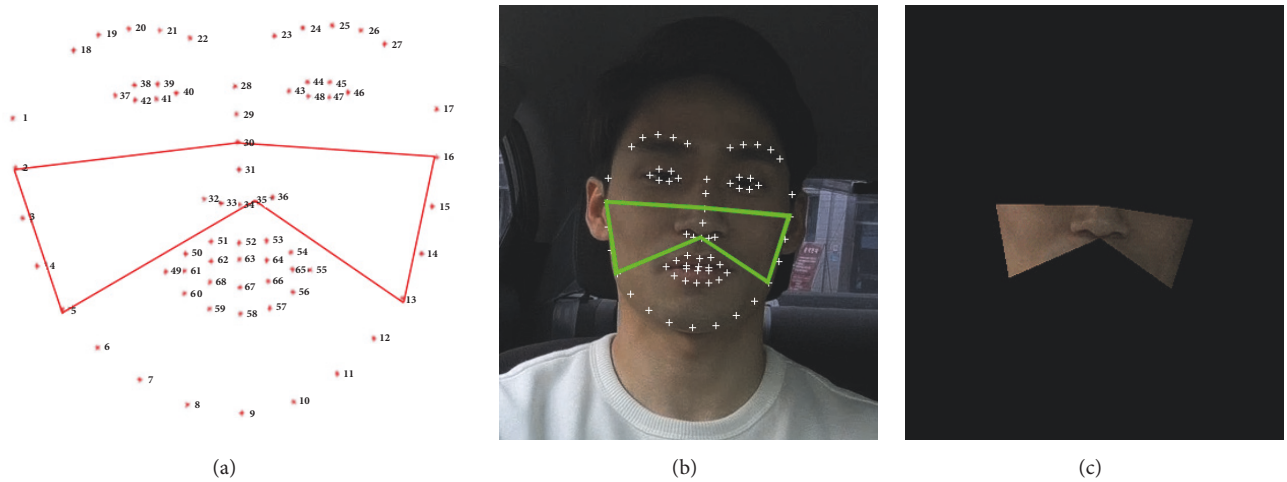

(a) (b) (c)

FIGURE 2: Extraction of facial landmark points. (a) The unique number of each of the 66 facial landmark points and the 6 selected points (4 contour points on the cheek and 2 points on the nose). (b) Detected result of the driving dataset. (c) Result of skin detection.

be obtained by taking a ratio from an RGB signal from the RoI as

$$RoverG = \frac{G_n}{R_n}, \tag{3}$$

where G_n and R_n are the normalized green and red signals [20, 21].

However, $RoverG$ is an unstable HR feature because it takes a fraction of the purely observed signal without any filtering. Therefore, this feature also includes variations due to illumination change and motion and should be separated into pure HR signals.

Before extracting the HR signal, a detrending method was applied to remove the nonstationary component with the smoothing parameter $\lambda = 10$ [22]. Then ensemble empirical mode decomposition (EEMD) is employed to separate the HR source signal from a number of noisy components in $RoverG$ [11]. EEMD is a noise assisted data analysis method that separates the Intrinsic Mode Function (IMF) from the data. The IMF extraction process, called sift, is accomplished by averaging the trials with the signal plus white noise, which is newly generated at every trial. If enough trials are carried out and more white noise is added, the components that make up the observed signal can be separated. In [15], which IMF is close to HR is determined through EEMD, and the fourth IMF is extracted as the HR component.

(a)   (b)
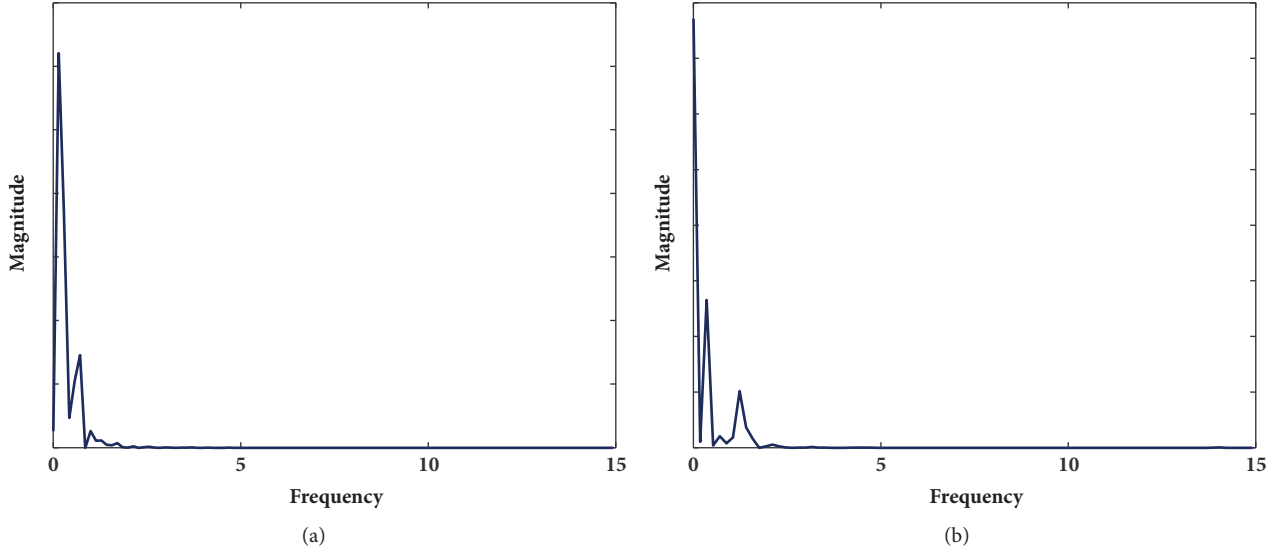
FIGURE 3: Frequency (in Hz) analysis of (a) illumination change and (b) vibration under automotive driving conditions.

However, since the automotive driving environment is very dynamic, several estimated HRs are derived as candidates for one estimation window for a stable HR estimation. Thus, the *RoverG* feature signal conversion and EEMD IMF extraction is iteratively performed in a window. The kth window, denoted as I_k, is divided into m periods by accumulating one second intervals from the first starting point to $p_1, p_2, \ldots, p_m (= I_k)$. Then, the HR for each period is estimated, and m estimated HRs are derived from the window. However, since all of the m estimated HRs have different inconsistent results, Mahalanobis distance is employed to exclude the result that is the furthest from most of the m results as

$$d\left(HR_{cand}, \mu_{cand}\right)$$
$$= \left[\left(HR_{cand} - \mu_{cand}\right)^T S^{-1} \left(HR_{cand} - \mu_{cand}\right)\right]^{1/2}, \quad (4)$$

where HR_{cand} and μ_{cand} are $m \times 1$ vectors consisting of m estimated candidate results and the mean of HR_{cand}, respectively, and S^{-1} is the covariance matrix. The candidate estimated HRs left after this exclusion are averaged and adopted as a result at the k second.

2.3. Power Spectral Density Analysis and Temporal Filtering. In order to calculate the final HR per minute, PSD is analyzed using the Welch method [23]. The cutoff frequency is set as (0.7, 4) HZ, corresponding to (42, 240) beats/min (bpm) and 128-order hamming window is used as the bandpass filter. However, the ambient light of the external noise in the cutoff frequency band may still cause intermittent peaking of the estimate. In order to cope with this problem, temporal filtering is applied to smooth the estimate trend as

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

where HR^t denotes the HR at time t. Threshold α denotes the allowable maximum value for the difference between the previous HR estimate and the current estimate. The parameter s determines the number of frames used for smoothing. These parameters (α and s) were chosen for optimal performance from the data set collected based on the assumption that HR does not change substantially in one second. The overall algorithm flow is shown in Algorithm 1.

3. Experiments and Results

In this section, we compare the performance of the proposed features against those presented in recent studies with the public HCI dataset.

3.1. Comparative Analysis of Features. As mentioned in Section 2, the green channel has the strongest PPG signal [6, 20]. On the other hand, Haan et al. proposed *XminY* with *RoverG* and proved that *XminY* has the highest performance in terms of experimental results [7]. Thus, it is necessary to determine which of the various feature signals produces the best HR signal.

For stable analysis, the MAHNOB-HCI dataset [24], a public indoor environment dataset, was used to compare the results of the five features, and the results are shown in Table 1.

Several commonly used performance indicators are employed to compare the performance of each feature [6]. M_e and SD_e are the mean and standard deviation, respectively, of the difference between ground truth and the obtained estimate, $HR_{dif} = HR_{est} - HR_{gt}$. Additionally, the root mean square error ($RMSE$) and M_{eRate}, which is the percentage of $\sum_{n=1}^{N}(|HR_{dif}(n)|/HR_{gt}(n))$, are employed to measure precision. Finally, r is the Pearson correlation coefficient that can evaluate the correlation between the two values.

```
Input: Image frame consist of RGB channel
Output: Estimated heart rate
Initialization: A video sequence within sliding window
For frame = 1, 2, ..., N
    If frame == 1
        Detect a facial landmark points
        Select 6 facial landmark points for cheek and nose
    End
    Track the detected region of interest
    Detect skin region within region of interest
    If mod(frame, frame rate) == 0 and frame >= length of window
        For period = p₁, p₂, ..., pₘ
            RGB normalization
            Calculate feature signal, RoverG = Gₙ/Rₙ
            Extract intrinsic mode function for heart rate from RoverG
            Power spectral density analysis
        End
        Filtering outlier using Mahalanobis distance, d(HRₑₛₜ, μₑₛₜ)
        Obtain heart rate result HRᵗₐᵥᵣ by averaging remaining estimates
        If HRᵗₐᵥᵣ − HRᵗ⁻¹ > α
            Temporal filtering with estimated result
        End
    End
End
```

ALGORITHM 1: Heart rate estimation algorithm.

TABLE 1: Comparison of heart rate estimation using different features (best performance in bold).

Feature	$M_e(SD_e)$ (bpm)	RMSE (bpm)	M_{eRate}	r
Green	-10.6(4.19)	11.3	14.22%	-0.35
Green_mah	-10.33(10.17)	14.45	13.71%	-0.20
Green_mah_TF	-6.63(7.21)	13.68	15.84%	-0.50
XminY	-20.1(6.54)	21.0	27.2%	-0.32
XminY_mah	-11.93(9.91)	15.3	16.11%	0.07
XminY_mah_TF	-12.07(5.15)	13.22	15.11%	0.39
RoverG	-2.43(7.27)	7.27	4.93%	0.59
RoverG_mah	**-0.57**(5.94)	**3.26**	5.58%	0.59
RoverG_mah_TF	0.80(**3.35**)	**3.26**	**3.68**%	**0.75**

Of the features, Green and *RoverG* are the signal from the pure green channel value in the RGB image and the feature from (2), respectively. *XminY* is the difference between *X* and *Y*, which is a linear combination feature of the RGB signal as described in (6)

$$X = 3R_n - 2G_n$$
$$Y = 1.5R_n + G_n - 1.5B_n. \tag{6}$$

RoverG_mah is a method of removing the peak candidate estimation value by applying the Mahalanobis distance to the estimated values of *RoverG*, and *RoverG_mah_TF* is the result of smoothing the outlier through temporal filtering.

As shown in Table 1, of the five metrics, *RoverG_mah_TF* shows the best performance. Although *RoverG* without any postprocessing shows a considerable fluctuation in its the result, the *RoverG_mah* with the statistical exclusion method of candidates has a relatively stable result. On the other hand, *XminY*, which showed the highest performance in [7], shows a lower performance than the other features with the MAHNOB-HCI dataset.

3.2. Validation Using Public Indoor Dataset. To validate the proposed method, its performance was compared with the recently proposed methods using a public dataset. The MAHNOB-HCI dataset is a public HCI dataset captured with several vital signals in the indoor environment. The dataset consists of two experiments containing emotion elicitation and implicit tagging. The subjects consist of 12 males and 15 females, each of whom was synchronized with the image by attaching an electrocardiography (ECG) sensor to their body. The ECG and image are recorded at 256 Hz and a frame rate of 61, respectively, and the resolution of the image is 780 by 580. Since it is of interest to estimate HR change over time, emotion elicitation data is adopted in the experiment. Emotion elicitation data is a data recording the vital signal and the facial image according to the stimulus by showing some videos (e.g., nature documentary or horror movie) to the subject. A comparison of the performance of the related methods on the MAHNOB-HCI dataset is shown in Table 2. For the previous methods, while the MAHNOB-HCI dataset was quite a challenging dataset, Li2014 and Tulyakov2016 achieved substantial accuracy with marginal improvement thereafter. Nevertheless, our algorithm, which is proposed to

TABLE 2: Comparison of the performance of related methods with the MAHNOB-HCI dataset (best performance in bold).

Feature	$M_e(SD_e)$ (bpm)	$RMSE$ (bpm)	M_{eRate}	r
Poh2010	-8.95(24.3)	25.9	25.0%	0.08
Poh2011	2.04(13.5)	13.6	13.2%	0.36
De Haan2013	4.62(6.50)	6.52	6.39%	0.82
Balakrishman2013	-14.4(15.2)	21.0	20.7%	0.11
Li2014	-3.30(6.88)	7.62	6.87%	0.81
Tulyakov2016	3.19(5.81)	6.23	5.93%	**0.83**
Ours	**0.80(3.35)**	**3.26**	**3.68**%	0.75

target a dynamic environment (e.g., the automobile driving environment), shows very high accuracy performance in this indoor environment. In terms of the Pearson correlation coefficient, its performance is comparable to the best performing previous method (e.g., Tulyakov2016). Except for this indicator, given the residual performance results related to the error, the estimate result of the proposed method is shown to outperform over all previous methods.

3.3. Demonstration on Dynamic Driving Dataset.

To demonstrate the proposed method under a driving scenario, a real driving dataset was collected under driving condition with 19 subjects in their 20s and 30s. The subjects included men and women of different ethnic backgrounds from countries such as Korea, China, and the Middle East. The driving dataset was captured by an action camera, Go-pro HERO 3+, fixed on a windscreen recording at a 30 frames per second rate and a resolution of 1920-by-1080. The ground truth was obtained by attaching a contact based pulse sensor to the earlobe of the subjects and synchronized with the captured dataset (the MP507 model of MEK was used as the earlobe pulse sensor). In order to securely obtain the dataset, the subject in the passenger seat was recorded instead of the actual driver, and they were asked to move their head up and down sometimes during the course of the driving. The subjects were also asked to rush up a hill before boarding the vehicle to check for pulse rate changes. It was recorded as naturally as possible without any additional constraints on the experiment. The driving course included a variety of actual driving road elements such as shade, curved sections, hills, and speed bumps. The ground truth is recorded in synchronization with the dataset using an earlobe attached sensor.

In order to address the stable performance of the proposed method, a Bland-Altman plot is employed. A Bland-Altman plot is a statistical plotting method that represents the agreement between two measurements. Each coordinate of the plot is denoted as in

$$BA(x, y) = \left(\frac{HR_{est} + HR_{gt}}{2}, HR_{est} - HR_{gt} \right). \quad (7)$$

The agreement A at the 95% confidence interval is shown in

$$A = \frac{1}{N} \sum_{i=1}^{n} a_i \times 100,$$

$$with \ a_i = \begin{cases} 1, & if \ \left| HR_{est} - HR_{gt} \right| > 1.96 \times \sigma \\ 0, & if \ \left| HR_{est} - HR_{gt} \right| < 1.96 \times \sigma, \end{cases} \quad (8)$$

where N is the total number of measurements and σ denotes the standard deviation between the two data sample sets. Figure 4 shows the Bland-Altman plot results of our proposed method with four randomly selected subjects from the driving dataset. The red and green line denotes the mean and standard deviation of the measurements, respectively. Each measurement is a combination of the estimated HR and ground truth per second. Figure 4 shows that although the results are applied to all four driving data sets, the mean of the errors is substantially small and a high agreement is obtained.

In order to visualize the tendency of the estimated HR and ground truth over time, the result is shown in Figure 5. Although the estimated value is slightly fluctuated compared with the ground truth, the difference is maintained within a maximum of 3 beats per minute. Moreover, it maintains similar stability to the normal interval even in the interval of fluctuation caused by speed bump and the rapid illumination change.

3.4. Performance Analysis Based on Execution Speed.

Our proposed method is applied to vehicle environment. Therefore, fast performance is required even if some performance degradation occurs using constrained resources. By Huang *et al.* [11], the true IMF can be defined as an ensemble of many trials as shown in

$$EEMD \ Components = \frac{std(x)}{N} \sum_{i=1}^{N} \{EMD(x + n_i)\}. \quad (9)$$

N is the number of trials and x, n denote the observation signal and noise, respectively. However, this approach requires a very large N resulting a large number of EMD calculations. Our proposed approach here limits the number of EMD calculations by exploiting independent identically distributed (iid) property of the white noise. Self-cancellation of the white noise can be accomplished by

$$EEMD \ Components = \frac{std(x)}{M} \sum_{i=1}^{M} [EMD\{x$$
$$+ (\text{mod}(i, 2)) \bullet n_i - (\text{mod}(i, 2) - 1) \bullet n_i\}]. \quad (10)$$

mod is a function to obtain the remainder and M denotes the number of limited trials. However, based on the characteristic that noise n is iid like in theoretical EEMD, the process of adding noise in (10) was performed only in $M/2$ trials ($M \approx M/2 \ll N$). This method and (9) are called EEMD_n1 and EEMD, respectively, and 10 and 100 trials are performed, respectively, to compare with EEMD which is commonly used as [12].
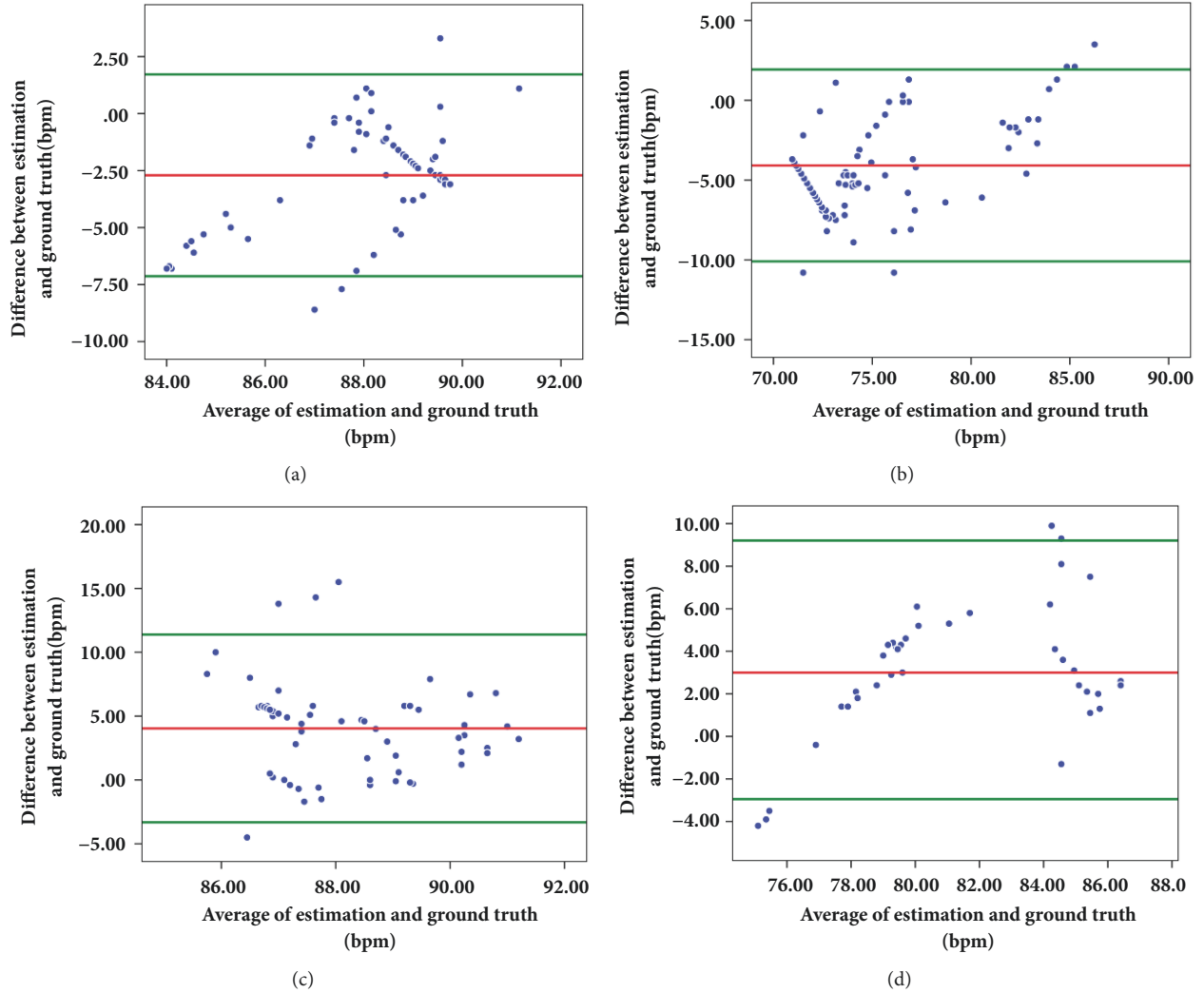
(a)

(b)

(c)

(d)

FIGURE 4: Bland-Altman plot analyzed at a 95% confidence interval. Each agreement of plot: (a) 95.9%; (b) 93.2%; (c) 93.5%; (d) 90%.
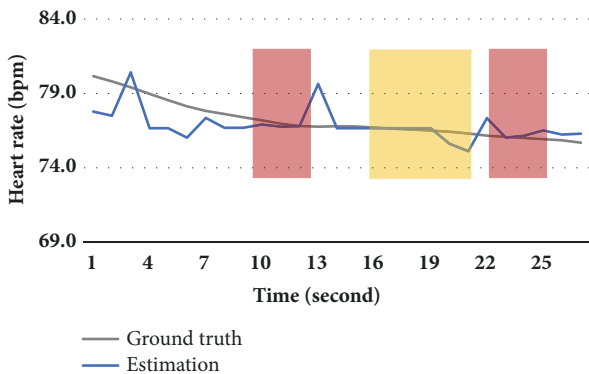

FIGURE 5: Heart rate trend between estimation and ground truth on challenging driving course. Red box: speed bump. Yellow box: rapid illumination change.

TABLE 3: The time it takes to operate once for each method.

Method	Operation time (second)
DRMF detection	0.86
KCF tracking	0.27
EEMD_n1	0.33
EEMD	4.52

tracking takes a considerable amount of time to process. It also presents a challenge when facial motion takes place. The time taken to operate each module is analyzed and shown in Table 3. While DRMF detection and KCF tracking are performed at every frame, EEMD_n1 and EEMD are performed as many as the number of candidate occurrences when an image frame is presented as input by the sliding window length.

Based on the result, four approaches are constructed as shown in Table 4, and their performance is compared to

On the other hand, in case of RoI selection, the previously proposed method that detects face per frame instead of face

TABLE 4: Comparison of heart rate estimations using different features (best performance in bold).

Approach	Absolute mean	Standard deviation	RMSE
DRMF+EEMD	4.35	2.29	4.89
DRMF+EEMD_n1	4.75	**1.82**	5.08
KCF+EEMD	**3.66**	2.42	**4.37**
KCF+EEMD_n1(ours)	3.71	3.07	4.74

determine the most efficient algorithm. Overall, the performance is better when using KCF than when using DRMF. This is because DRMF has difficulty in detecting the correct RoI corresponding to the cheek region when a part of the face is occluded due to shaking or facial motion. In the case of EEMD_n1, although the operation time is greatly reduced, the performance decline is very small.

4. Conclusions

This paper proposed a novel approach to estimating HR remotely in actual driving environments. Most previous studies have been proposed under indoor environments, which often lead to high implied levels of performance based on a well-controlled practical application context. On the other hand, the proposed method showed attaining the highest practical applicability by demonstrating its ability under the most challenging environment, the automotive driving environment. Before testing the proposed method under the automotive driving environment with various obstacles, it was compared to other methods using the same indoor public dataset as previous studies and using the same performance index to validate its effectiveness. The proposed method was then applied to data from an actual driving situation and a fairly stable result was obtained. For automotive driver HR estimation, estimating the HR instantaneously is necessary to prevent accidents. Focusing on this issue, an appropriate approach was sought to maximize performance while reducing operation time. Hence, the performance was also analyzed in terms of processing time by comparing the proposed method with the conventional algorithms and the modified algorithm. The proposed method demonstrated a considerably superior performance and yet had a short processing time.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.
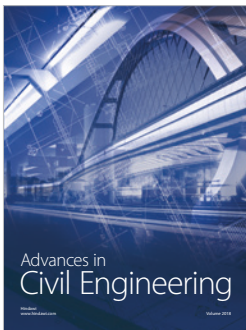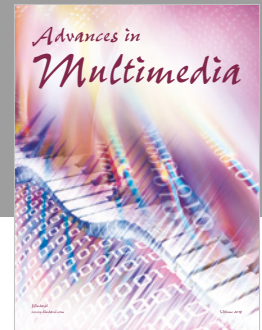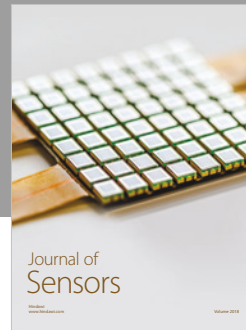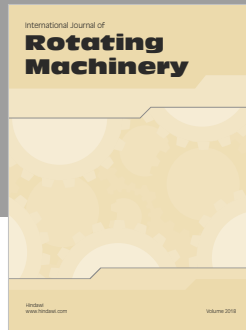
References

[1] M. Huelsbusch, A. V. Clough, C. Chen, and V. Blazek, "Contactless mapping of rhythmical phenomena in tissue perfusion using PPGI," in *Proceedings of the Medical Imaging 2002*, vol. 4683, pp. 110–118, San Diego, CA, USA, 2002.

[2] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, no. 10, pp. 10762–10774, 2010.

[3] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2011.

[4] F. Zhao, M. Li, Y. Qian, J. Z. Tsien, and I. P. Androulakis, "Remote measurements of heart and respiration rates for telemedicine," *PLoS ONE*, vol. 8, no. 10, p. e71384, 2013.

[5] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Transactions on Graphics*, vol. 31, no. 4, article 65, 2012.

[6] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4264–4271, Columbus, OH, USA, June 2014.

[7] 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.

[8] W. Wang, S. Stuijk, and G. De Haan, "Exploiting spatial redundancy of image sensor for motion robust rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 415–425, 2015.

[9] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pp. 2396–2404, Las Vegas, USA, July 2016.

[10] L. Xu, J. Cheng, and X. Chen, "Illumination variation interference suppression in remote PPG using PLS and MEMD," *IEEE Electronics Letters*, vol. 53, no. 4, pp. 216–218, 2017.

[11] Z. H. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis (AADA)*, vol. 1, no. 1, pp. 1–41, 2009.

[12] J. Cheng, X. Chen, L. Xu, and 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*, vol. 21, no. 5, pp. 1422–1433, 2017.

[13] H. Qi, Z. Guo, X. Chen, Z. Shen, and Z. Jane Wang, "Video-based human heart rate measurement using joint blind source separation," *Biomedical Signal Processing and Control*, vol. 31, pp. 309–320, 2017.

[14] C. Tang, J. Lu, and J. Liu, "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*, pp. 1309–1315, Salt Lake City, USA, June 2018.

[15] D.-Y. Chen, J.-J. Wang, K.-Y. Lin et al., "Image sensor-based heart rate evaluation from face reflectance using Hilbert-Huang transform," *IEEE Sensors Journal*, vol. 15, no. 1, pp. 618–627, 2015.

[16] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomedical Optics Express*, vol. 6, no. 5, pp. 1565–1588, 2015.

[17] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2013*, pp. 3444–3451, USA, June 2013.

[18] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[19] T. Sawangsri, V. Patanavijit, and S. Jitapunkul, "Face segmentation based on hue-cr components and morphological technique," in *Proceedings of the IEEE International Symposium on Circuits and Systems 2005, ISCAS 2005*, pp. 5401–5404, Japan, May 2005.

[20] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[21] J. A. Crowe and D. Damianou, "The Wavelength Dependence Of The Photoplethysmogram And Its Implication To Pulse Oximetry," in *Proceedings of the 1992 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2423-2424, Paris, France, Oct 1992.

[22] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, "An advanced detrending method with application to HRV analysis," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 2, pp. 172–175, 2002.

[23] G. Valakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3430–3437, Portland, OR, USA, June 2013.

[24] J. Kuo, S. Koppel, J. L. Charlton, and C. M. Rudin-Brown, "Evaluation of a video-based measure of driver heart rate," *Journal of Safety Research*, vol. 54, pp. 55–59, 2015.

Journal of
Engineering

The Scientific
World Journal

International Journal of
Rotating
Machinery

Journal of
Sensors

Advances in
Multimedia

Advances in
Civil Engineering

Journal of
Control Science
and Engineering

Journal of
Robotics

Journal of
Electrical and Computer
Engineering

Hindawi

Submit your manuscripts at
www.hindawi.com

Advances in
OptoElectronics

VLSI Design

International Journal of
Navigation and
Observation

Modelling &
Simulation
in Engineering

International Journal of
Aerospace
Engineering

International Journal of
Chemical Engineering

International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components

Shock and Vibration

Advances in
Acoustics and Vibration