

Heart Rate Measurement Combining Motion and Color Information

Jean-Pierre Lomaliza[†], Hanhoon Park^{††}, Kwang-Seok Moon^{†††}

ABSTRACT

Daily monitoring of the heart rate can facilitate detection of heart-related diseases in their early stages. Therefore, providing an easy-to-use and noninvasive heart rate monitoring system has been a very popular research topic in the field of healthcare. One of good candidate methods is to use commonly available cameras and extract information that can help to estimate heart rate from a human face. Generally, such information can be retrieved using two different approaches: photoplethysmography (PPG) and ballistocardiography (BCG). PPG exploits slight color changes caused by blood volume variations during heartbeats; thus, it tends to be vulnerable to unstable lighting conditions. BCG exploits subtle head motions caused by pumped blood travelling through the carotid artery during heartbeats; thus, it is vulnerable to the voluntary head movements that are not related to heartbeats. Nevertheless, most related works use either to estimate the heart rate. In this paper, we propose to combine two approaches to be robust to challenging conditions. Specifically, we explore possible ways to combine raw signals obtained from two approaches and verify that the proposed combination shows better accuracies under challenging conditions, such as voluntary head movements and ambient lighting changes.

Key words: Heart Rate Measurement, Photoplethysmography, Ballistocardiography, Hybrid Approach, Face Video

1. INTRODUCTION

According to the American Health Association, heart-related diseases are the leading causes of death in the United States of America and around the world [1]. Thus, daily heart rate monitoring is a very important task to detect heart-related diseases in their early stages and provide a preventive treatment. The estimation of heart rate can be achieved by using omnipresent devices on modern human living space, such as cameras. Moreover, the best way to ease such monitoring is to provide an easy-to-use and noninvasive way in a touchless

manner. A practical way to estimate the heart rate is exploiting physiological changes on the human face that are caused by heartbeats. Such physiological changes are expressed by two different approaches: photoplethysmography (PPG) and ballistocardiography (BCG). PPG, first explained at [2], measures the changes of colors in the human skin due to changes of blood volumes during heartbeats. On the other hand, BCG measures the head movements caused by the blood influx through the carotid artery during heartbeats. The color changes or head movements are extremely tiny and invisible by a naked human eye, but can be captured

※ Corresponding Author : Kwang-Seok Moon, Address: (608-737) Yongso-ro 45, Nam-gu, Busan, Korea, TEL : +82-51-629-6218, FAX : +82-51-629-6210, E-mail : ksmoon@pknu.ac.kr

Receipt date : Oct. 15, 2020, Revision date : Nov. 5, 2020
Approval date : Nov. 6, 2020

[†] Dept. of Electronic Engineering, Pukyong National University (E-mail : milorenius@pukyong.ac.kr)

^{††} Dept. of Electronic Engineering, Pukyong National University (E-mail : hanhoon.park@pknu.ac.kr)

^{†††} Dept. of Electronic Engineering, Pukyong National University

※ This work was supported by a Research Grant of Pukyong National University (2019).

using cameras. Many vision-based researches have been reported on providing an effective heart rate monitoring system using either of PPG and BCG, including the PPG-based methods [3,4,5] and BCG-based methods [6,7,8]. The PPG-based methods commonly attempt to compute the average of pixels' intensities within certain regions of interest (ROI) in the human face. Variations of such pixels' intensities over time are then considered as raw heart rate-related signals, may containing some noises in real environments. The pixel intensities can vary depending on lighting conditions; thus the PPG-based methods are usually sensitive to unstable lighting conditions. For example, ambient lighting changes would highly influence those raw signals. On the other hand, the BCG-based methods commonly attempt to track some feature points on the human face within a period of time. Series of locations of such feature points are considered as raw heart rate-related signals. Therefore, the BCG-based methods are usually sensitive to voluntary head motions that are not originated from heartbeats. However, two approaches, PPG and BCG, can be alternatively used in the challenging conditions of each other, because the change in pixel intensity in the human face caused by voluntary head motions are small and the feature points used in BCG are rarely influenced by lighting changes. Hence, in this paper, we explore an effective way to combine PPG and BCG approaches to create a hybrid method that has better robustness to voluntary head movements and unstable lighting conditions than existing PPG-based methods and BCG-based methods.

The remaining parts of this paper are organized as follows. First, related works are introduced in Section 2. Then, the proposed combination method is explained in Section 3. In Section 4, experimental results are presented and analyzed. Finally, conclusion is made and possible future studies are discussed in Section 5.

2. RELATED WORK

Both PPG and BCG approaches have three main steps: frame-to-frame tracking and raw signal acquisition, signal processing and selection, and heart rate computation. In this section, we will discuss each step used in several notable PPG-based and BCG-based methods.

2.1 PPG-Based Methods

For both PPG-based and BCG-based methods, raw signals are acquired by tracking certain chosen ROIs within the human face. For PPG-based methods, the aim of tracking is to retrieve colors changed within the same ROI for all frames. For BCG-based methods, locations of tracked features in ROIs are directly considered as raw signals. One of early PPG-based methods, proposed in [9], used the whole facial region as ROI and computed the average pixel intensity for the red, green, and blue channel in each frame. Time series of those pixel intensities gave three raw signals. The method removed noise by applying the fourth order Butterworth filter, normalized the signals, and applied an independent component analysis (ICA) for blind source separation. Then, the method selected the signal that was the most similar to a pulse signal and computed the heart rate from the selected signal. Tasli et al. proposed to use facial landmarks to address facial expression problems [10]. They averaged intensity of the green channel of pixels to obtain the raw signal. Then, they detrended the signal using the technique proposed in [11] to remove noises caused by ambient lighting changes. Finally, they used both time and frequency domain information to estimate the heart rate. Jain et al. introduced an additional preprocessing step of cropping around the face [12]. After cropping the face video with a predefined size of $M \times M$, they considered only the red channel of pixels and built a matrix \mathbf{A} where each column represents a frame. Then, they applied a principal component analysis

(PCA) to obtain a matrix \mathbf{A}' , which was reconstructed with the most principal vectors, and computed a reconstruction error matrix \mathbf{A}_{PPG} by subtracting \mathbf{A}' from \mathbf{A} . Each column of \mathbf{A}_{PPG} was reshaped back to the original frame size of $M' \times M$. The PPG signal was obtained by averaging the pixel intensity in each reconstructed frame. After filtering the signal with a bandpass filter with a pass band of [0.5, 5] Hz, they proposed to only consider portion of the signal having significant visual features like peaks, foots, and dicrotic notches. Finally, they computed heart rate relatively to the number of detected peaks within a signal duration.

2.2 BCG-Based Methods

As mentioned before, the focus of BCG-based methods is to accurately track locations of selected features within ROIs. The first practical BCG-based method was proposed by Balakrishnan et al. [6]. The method used the Viola-Jones face detector [13] that comes with OpenCV library [14] to retrieve a human face region. As the face detector returns a rough rectangular region, the method considered as ROI 90% heightwise and 50% widthwise centered at the middle. To avoid eyes blinking artifacts, the method removed sub-rectangle spanning from 25% to 50% heightwise from the ROI. Then, the method detected feature points within the ROI using OpenCV's good features to track (GFT) function. Those detected features were tracked using OpenCV's Lucas-Kanade optical flow. Time series of vertical locations of feature points constituted the raw BCG signal. The fifth order Butterworth filter with a bandpass of [0.75, 5] Hz was applied to the raw signal to remove noises. Then, PCA was applied to the filtered signal and only the five most dominant eigenvectors were considered in the following steps. Finally, the method selected the most periodical eigenvector and estimated heart rate from the frequency having the highest magnitude after transforming the selected signal to the frequency domain. Shan and

Yu proposed to reduce the ROI only within forehead regions in order to avoid feature points' motions caused by facial expressions [7]. They also used ICA instead of PCA for signal decomposition. To estimate the heart rate, they selected the frequency with the highest magnitude in the power spectra density. Haque et al. proposed combining the OpenCV's GFT and supervised descent method (SDM) [15] to detect feature points on the human face [8]. Feature points detected and tracked with SDM helped to address the problem of tracking loss that occurred during facial expressions or other rigid motions. After tracking feature points and obtaining a raw signal, they proposed to apply an eighth order Butterworth filter followed by a moving average filter to remove noises. They also applied PCA and a discrete cosine transform to the filtered signal to find the most periodic signal.

3. THE PROPOSED COMBINATION METHOD

In this section, the entire process for heart rate estimation is described; First, we introduce how raw PPG and BCG signals are obtained. Second, we introduce how the raw signals are processed and combined for a more robust heart rate estimation under challenging conditions. Third, we discuss the problem with selecting the wrong signal not related to heart rate, caused when using the existing signal selection methods. Finally, we introduce how heart rate is estimated from the heart rate-related signal. Fig. 1 shows the overall process of the proposed method.

3.1 Raw Signal Acquisition

We detect the facial region using the Viola-Jones detector and also detect facial landmarks using the SDM that comes implemented with contribution modules of OpenCV. After facial landmark detection, we define four different regions around eyes, forehead, and nose. Those regions are defined by grouping facial landmarks and founding

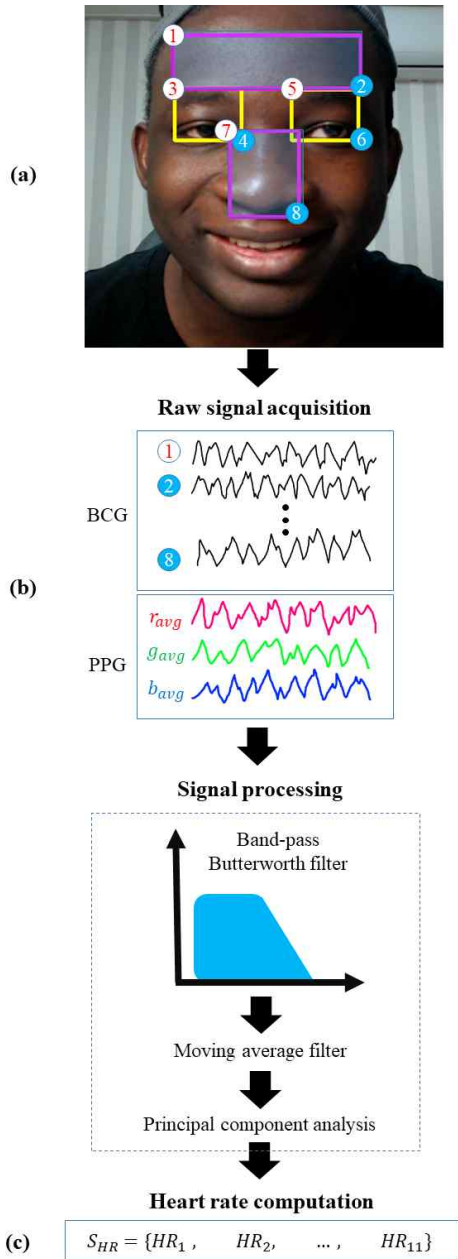


Fig. 1. Flowchart of the proposed method. (a) ROIs detected at the first frame and tracked throughout the subsequent frames, (b) raw signals acquired by: averaging pixel intensities on the red, green, and blue channels within selected ROIs (pink boxes) to construct raw PPG signals; tracking vertical locations of top-left corners (white circles) and bottom-right corners (blue circles) to construct raw BCG signals, and (c) heart rate computation for all signals reconstructed by PCA.

a 2D bounding box. Those bounding boxes constitute regions of interest (ROI) of the proposed method. Then, we track ROIs using the OpenCV median-flow region-based tracker. Vertical locations of left-top and bottom-right corners of tracked regions are considered as raw BCG signals. To obtain raw PPG signals, at each frame, we compute the average of pixels intensities for red, blue, and green channels within forehead and nose regions. In general, the red channel provides better PPG signals than the others [16], but the blue and green channels may contain complementary information [9]. We excluded eye regions to minimize the impact of facial non-rigid motions. Therefore, we obtain eight raw BCG signals and three raw PPG signals.

3.2 Signal Processing and Selection

The main goal of this paper is to effectively combine the raw PPG and BCG signals so that information related to the heart rate may be conserved the most. With a frame rate of 25 fps, we gather 512 samples corresponding to 20.48 s for each raw PPG or BCG signal. Thus, the dimension of samples for eight raw BCG signals and three raw PPG signals corresponds to 8×512 and 3×512 , respectively. Then, we stack them and obtain a resulting raw signal samples of size 11×512 . Then, we independently apply a Butterworth filter with a pass band of [0.75, 3.5] Hz to each of the raw signals. The aim of the Butterworth filter is at removing high frequency noises usually related to fast lighting condition changes. In addition, we apply a moving average filter to smoothen the signals.

Finally, we perform the PCA on the filtered signal data. Although ideally the heart rate-related information should appear within the most dominant Eigenvectors, we observed in our preliminary experiments that it is not guaranteed. Hence, we use all the Eigenvectors to re-project the original data to obtain a transformed data. After performing the

PCA, conventional methods usually select the most periodic signal as the pulse signal. However, in our observation during preliminary experiments, the signal having the closest heart rate estimation to the ground truth was not always the most periodic one, especially on challenging conditions. Also, we could not find a reliable signal selection method in the literature and failed to develop it as well. Nevertheless, as the aim of this paper is to analyze accuracy of the combination method under challenging conditions, we focus on the percentage of cases where at least one candidate signal had an estimation of heart rate close to the ground truth. To this end, signals are manually selected after estimating the heart rate using the technique described in the next section.

3.3 Heart Rate Computation

The most popular way to compute the heart rate would be to transform the selected signal to the frequency domain by applying a fast Fourier transform and finding the frequency corresponding to the maximum power spectrum as done in [7]. In situations of strong noises under challenging scenarios, however, the maximum frequency is not always the one expressing the heart rate at best. Hence, we use a more advanced heart rate computation method that does not need the use of the maximum frequency. Using a moving average filter with a large window, we compute the trend of the selected signal that will be smoother and has only lower frequency components of the original signal. The trend signal divides the original signal into cycles. From each cycle we create sets of peaks and perform all combinations of peaks from different cycles. Then, the combination having the lowest variance for the average elapsed time between peaks is used for the heart rate computation.

We compute the heart rate in beats per minute (BPM) from the set of peaks p_i , $i = (1, 2, \dots, K)$, where K is the number of peaks in a combination, using Eqs. (1) and (2). This is originated from the

optimization tree proposed in [17] where the period with respect to the most dominant frequency is used to segment the signal into cycles. However, our method does not compute the most dominant frequency, which allows to be more robust in challenging conditions where that most dominant frequency would be too far from the heart rate-related frequency.

$$T_E = \frac{1}{K-1} \sum_{j=1}^{K-1} \{t(p_{j+1}) - t(p_j)\}, \quad (1)$$

$$\text{Heart_rate} = \frac{f_{\text{samp}} \times 60}{T_E}. \quad (2)$$

where T_E is the average elapsed time between peaks, $t(p_j)$ is the time at the j -th peak, and f_{samp} is the sampling rate.

4. EXPERIMENTAL RESULTS AND ANALYSIS

Experiments were conducted on a laptop setup using an Apple MacBook pro model 2015 with an Intel core i7. To get face videos, the webcam we used a webcam (Logitech C922). The Polar chest belt H7 was used as the ground truth device. 25 subjects of different ethnicities and ages ranging from 22 to 41 participated to our experiments. They were required to sit within 1 meter from the webcam while staring at it. To create challenging data, we asked subject to freely move their heads while staring at the camera. Moreover, we randomly changed the intensity of the indoor LED light sources to create challenging lighting conditions. In some cases, we combined head movements and lighting changes. As the proposed method focusses on scenarios of challenging conditions, we compare it with other related methods in such cases. We picked PPG-based methods proposed in [9] and [12], BCG-based methods proposed in [7] and [8], together with a variant of the proposed method that uses ICA instead of PCA. Moreover, we focus on showing that the proposed method always ensures that one of obtained candidate signals has an estimated heart rate close to the ground truth within

Table 1. Percentages of containing heart rate-related one among the signals obtained by different signal acquisition methods

Challenging conditions	BCG [7]	BCG [8]	PPG [9]	PPG [12]	Proposed-ICA	Proposed-PCA
C0: None (ideal)	100%	100%	100%	100%	100%	100%
C1: Allowing lighting changes	94.50%	96.34%	75.20%	76.61%	97.32%	99.10%
C2: Allowing voluntary head movements	72.30%	75.40%	96.50%	94.68%	93.71%	98.41%
C3: C1 + C2	63.41%	71.32%	60.50%	65.05%	89.11%	93.35%

Table 2. Mean absolute errors in BPM

Challenging conditions	BCG [7]	BCG [8]	PPG [9]	PPG [12]	Proposed-ICA	Proposed-PCA
C0: None (ideal)	2.70	2.30	2.10	1.92	1.26	1.18
C1: Allowing lighting changes	3.20	3.31	11.71	10.60	1.99	1.45
C2: Allowing voluntary head movements	12.31	9.97	3.57	3.67	4.38	2.78
C3: C1 + C2	15.71	13.99	15.25	14.78	5.23	3.01

a certain threshold. In our experiments, we set the threshold value to 3 BPM. Table 1 and Fig. 2 show percentages of containing heart rate-related one among the signals obtained by each method in different scenarios. Table 2 and Fig. 3 show the mean absolute error of each method. First, in the ideal scenario where the subjects are asked to stay still without movements and also the lighting conditions are ideally set with a single LED source, all the methods had a heart rate estimation within 3 BPM away from the ground truth with a probability of 100%. In the second scenario with chal-

lenging lighting conditions, where we randomly moved an obstructing object between the light source and the subjects, PPG-based methods decreased much more in accuracy compared to BCG-based methods. In the third scenario where the subjects were asked to move their head freely while recording their faces, BCG-based methods' accuracies significantly decreased. Finally, in the fourth scenario where two challenging conditions were combined, both PPG-based and BCG-based methods experienced a significant decrease in their accuracy; however, the proposed method still had the high percentages (> 93%) of extracting heart

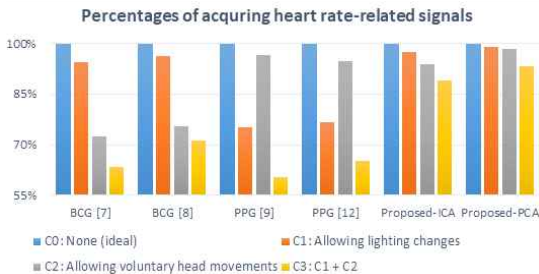


Fig. 2. Percentages of containing heart rate-related one among the signals obtained by each signal acquisition method in different challenging conditions.

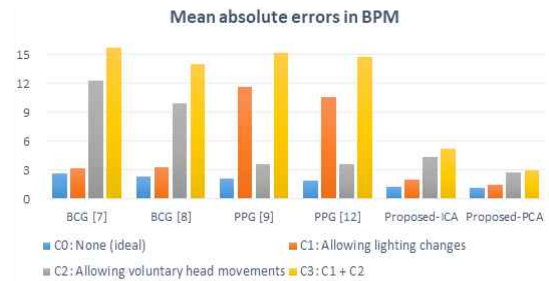


Fig. 3. Mean absolute errors in BPM of different signal acquisition methods in different challenging conditions.

rate-related signals and maintained low errors (< 3 BPM) in all the challenging conditions. In the proposed method, using ICA instead of PCA also yielded acceptable performance, but using PCA yielded better performance, especially in more challenging conditions.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a vision-based method that measures heart rate from face videos. The proposed method aimed at addressing challenging scenarios where lighting conditions are dynamically changing and the subjects freely move their heads, by combining PPG and BCG signals using PCA. As PPG signals were robust to voluntary head movements and BCG signals were robust to lighting changes, the PCA-based combination method minimized the erroneous heart rate measurement in the challenging conditions. As a result, it allowed to estimate heart rate in higher accuracy than conventional PPG-based or BCG-based methods.

As aforementioned, we could not find a reliable way to select a heart rate-related one among candidate signals obtained by applying PCA or ICA to the BCG and PPG signals. Thus, our future study will focus on developing a new and reliable signal selection method.

REFERENCE

- [1] E.J. Benjamin, P. Muntner, A. Alonso, M.S. Bittencourt, C.W. Callaway, A.P. Carson et al., "Heart Disease and Stroke Statistics-2019 Update: A Report from the American Heart Association," *Circulation*, Vol. 139, No. 10, pp. e56-e528, 2019.
- [2] A.B. Hertzman and J.B. Dillon, "Applications of Photoelectric Plethysmography in Peripheral Vascular Disease," *American Heart Journal*, Vol. 20, No. 6, pp. 750-761, 1940.
- [3] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote Heart Rate Measurement from Face Videos under Realistic Situations," *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4264-4271, 2014.
- [4] 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," *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2396-2404, 2016.
- [5] A. Lam and Y. Kuno, "Robust Heart Rate Measurement from Video Using Select Random Patches," *Proceeding of the IEEE International Conference on Computer Vision*, pp. 3640-3648, 2015.
- [6] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting Pulse from Head Motions in Video," *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3430-3437, 2013.
- [7] L. Shan and M. Yu, "Video-based Heart Rate Measurement Using Head Motion Tracking and ICA," *Proceeding of International Congress on Image and Signal Processing*, Vol. 1, pp. 160-164, 2013.
- [8] M.A. Haque, R. Irani, K. Nasrollahi, and T.B. Moeslund, "Heartbeat Rate Measurement from Facial Video," *IEEE Intelligent Systems*, Vol. 31, No. 3, pp. 40-48, 2016.
- [9] W. Verkruysse, L.O. Svaasand, and J.S. Nelson, "Remote Plethysmographic Imaging Using Ambient Light," *Optics Express*, Vol. 16, No. 26, 21434-21445, 2008.
- [10] H.E. Tasli, A. Gudi, and M. den Uyl, "Remote PPG Based Vital Sign Measurement Using Adaptive Facial Regions," *Proceeding of IEEE International Conference on Image Processing*, 2014.
- [11] M.P. Tarvainen, P.O. Ranta-aho, and P.A. Karjalainen, "An Advanced Detrending Me-

- thod with Application to HRV Analysis,” *IEEE Transactions on Biomedical Engineering*, Vol. 49, No. 2, pp. 172–175, 2002.
- [12] M. Jain, S. Deb, and A.V. Subramanyam, “Face Video Based Touchless Blood Pressure and Heart Rate Estimation,” *Proceeding of IEEE International Workshop on Multimedia Signal Processing*, pp. 1–5, 2016.
- [13] P. Viola and M.J. Jones, “Robust Real-time Face Detection,” *International Journal of Computer Vision*, Vol. 57, pp. 137–154, 2004.
- [14] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, O’Reilly Media, Sebastopol, California, 2008.
- [15] X. Xiong and F. De la Torre, “Supervised Descent Method and Its Applications to Face Alignment,” *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 532–539, 2013.
- [16] N.V. Hoan, J.H. Park, S.H. Lee, and K.R. Kwon, “Real-time Heart Rate Measurement Based on Photoplethysmography Using Android Smartphone Camera,” *Journal of Korea Multimedia Society*, Vol. 20, No. 2, pp. 234–243, 2017.
- [17] J.P. Lomaliza and H. Park, “Improved Peak Detection Technique for Robust PPG-based Heart Rate Monitoring System on Smartphones,” *Multimedia Tools and Applications*, Vol. 77, No. 13, pp. 17131–17155, 2018.



Jean-Pierre Lomaliza

He received the M.S. degree in electronic engineering from Pukyong National University, Busan, Korea, in 2016. He is currently a Ph.D. student at Pukyong National University. His research interests include augmented reality and computer vision.



Hanhoon Park

He received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from Hanyang University, Seoul, Korea, in 2000, 2002, and 2007, respectively. From 2008 to 2011, he was a post-doctoral researcher with NHK Science and Technology Research Laboratories, Tokyo, Japan. In 2012, he joined the Department of Electronic Engineering, Pukyong National University, Busan, Korea, where he is currently a Professor. His current research interests include augmented reality, human-computer interaction, and affective computing.



Kwang-Seok Moon

He received the B.S., M.S., and Ph.D. degrees in electronics engineering from Kyungpook National University, Seoul, Korea, in 1979, 1981, and 1989, respectively. In 1988, he was a student researcher with Tokyo University, Tokyo, Japan. From 1997 to 1998, he was a visiting professor with Jackson State University, MS, USA. In 1990, he joined the Department of Electronic Engineering, Pukyong National University, Busan, Korea, where he is currently a professor. His current research interests include image information processing and adaptive signal processing.