Visual Speech Recognition of Korean Words Using Convolutional Neural Network

Sung-Won Lee, Je-Hun Yu, Seung Min Park, and Kwee-Bo Sim

Department of Electronic and Electrical Engineering, Chung-Ang University, Seoul, Korea

Abstract

In recent studies, speech recognition performance is greatly improved by using HMM and CNN. HMM is studying statistical modeling of voice to construct an acoustic model and to reduce the error rate by predicting voice through image of mouth region using CNN. In this paper, we propose visual speech recognition (VSR) using lip images. To implement VSR, we repeatedly recorded three subjects speaking 53 words chosen from an emergency medical service vocabulary book. To extract images of consonants, vowels, and final consonants in the recorded video, audio signals were used. The Viola–Jones algorithm was used for lip tracking on the extracted images. The lip tracking images were grouped and then classified using CNNs. To classify the components of a syllable including consonants, vowels, and final consonants, the structure of the CNN used VGG-s and modified LeNet-5, which has more layers. All syllable components were classified, and then the word was found by the Euclidean distance. From this experiment, a classification rate of 72.327% using 318 total testing words was obtained when VGG-s was used. When LeNet-5 applied this classifier for words, however, the classification rate was 22.327%.

Keywords: Convolutional neural network, Human–robot interaction, Korean word recognition, Viola–Jones algorithm, Visual speech recognition

1. Introduction

Many people have an interest in service robots owing to developments in artificial intelligence (AI). Thus, researchers on robots or AI are developing diverse robots to recognize human expression, emotions, and speech. Such research is called human–robot interaction (HRI) [1, 2]. HRI can be applied to various fields such as factories, hospitals, and amusement parks.

To implement the HRI system, speech recognition is one of the important aspects. People require speech recognition because many products use this technology—e.g., vehicle navigation, voice recognition services of cell phones, and voice searching on the Internet. However, these programs for speech recognition have the problem of inaccuracy [3–6]. In the presence of noise, these programs cannot hear and analyze the command of the user. Thus, such programs are not used in emergency situations.

To overcome this problem, many ideas have been proposed by researchers. The solution to the speech recognition problem is visual speech recognition (VSR). Current speech recognition technology uses people’s voices. In contrast, VSR uses lip shapes to improve the accuracy of speech recognition.

Received: Jun. 5, 2018
Revised : Sep. 7, 2018
Accepted: Sep. 7, 2018

Correspondence to: Kwee-Bo Sim
(kbsim@cau.ac.kr)
©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 1. Structure of a Korean word.

Thus, VSR is used in human–computer interaction, speaker recognition, audio-visual speech recognition, sign language recognition, and video surveillance for convenience [5, 7]. VSR technology has two methods of approach: the visemic approach and the holistic approach. The visemic approach is the conventional and common method.

The viseme uses the phoneme of a word's mouth shapes. However, the holistic approach uses the whole word. Thus, the holistic method has a better result than the visemic approach [7]. However, the holistic method has not yet been made for the Korean language.

In Korean, the syllable of a word consists of three parts. Figure 1 shows the construction of a word. The word has three consonants, three vowels, and two final consonants. Moreover, a syllable generally consists of a consonant, vowel, and final consonant. The pronunciation of a syllable is also a sequence of a consonant, vowel, and final consonant [8]. However, the use of only the information from the consonant, vowel, and final consonant images cannot determine the correct word because the information of the consonant and final consonant is not exact and makes no difference except bilabial [9].

In this paper, a holistic approach and the images of a syllable were combined to solve this problem. To apply the holistic approach in Korean, the consonant, vowel, and final consonant parts of words were categorized. In addition, the words were classified by collecting the classification results of a syllable's components in the order of time.

Fifty-three Korean words were chosen for the holistic approach and recorded using a camera. The 53 Korean words were selected from an emergency medical service vocabulary book that was published by the National Medical Center in Korea [10]. Using the Viola–Jones detection algorithm, the lip shapes of subjects were found. From the lip shapes, the 53

words were classified by convolution neural network.

2. Related Work

Previous research on speech recognition has focused on improving the accuracy. Therefore, many results on speech recognition have been proposed using VSR for decades. For our VSR, the classification algorithm, lip extraction, and VSR method were investigated. In VSR, lip tracking is important because the extraction of the lip can simplify classification and recognition. In the section on the VSR method, VSR methods from the literature are explained.

2.1 Convolutional Neural Network

For classification of Korean words, a convolutional neural network was used. The convolutional neural network is a powerful classification algorithm developed in 1998 by LeCun et al. [11]. However, the convolutional neural network attracted few researchers until a few years ago owing to its number of operations for classification. Now, given the development of computer hardware, many researchers have paid attention to convolutional neural network theory.

A convolutional neural network has three stages: the convolution layer, subsampling layer, and fully connected layer. The convolution and subsampling layers are used for feature extraction of an input image. In addition, the fully connected layer classifies an input image. This is the advantage of the convolutional neural network because it has no other feature extraction. Figure 2 shows a simple structure of a convolutional neural network, also known as the LeNet-5 model. However, this LeNet-5 was modified to increase the classification rates. Moreover, this LeNet-5 has more layers than conventional LeNet-5 [1, 2, 11].

The convolutional neural network has other structures in addition to LeNet-5. In 2012, AlexNet was introduced and won the 2012 ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). AlexNet used 2 GPUs to increase performance and

Figure 2. Example of a convolutional neural network.

obtained good results in classifying images [12]. In 2014, many structures using GPUs such as GoogLeNet and VGGNet were influenced by AlexNet. Moreover, GoogLeNet won the 2014 ILSVRC, and VGGNet ranked second [13–15].

2.2 Visual Speech Recognition Method

For VSR, various methods have been proposed. Most approaches used extracted mouth images. However, the proposed methods of various researchers differ regarding how to classify mouth images or extract the mouth from an image

In 1994, Bregler and Konig [16] introduced word classification using “Eigenlips”. The authors used the energy function of measured the image features and a contour model. To classify 2,955 German words, they used a multilayer perceptron and a hidden Markov model (HMM).

In 2011, Shin et al. [17] made an interface device for a vehicle navigation device. They used not only VSR but also audio-visual speech recognition (AVSR). AVSR was mostly used when the phonic data contained significant noise. Therefore, to overcome the noise problem, they used a robust lip tracker such as the Lucas–Kanade (LK) method and classification such as the hidden Markov model, artificial neural network, and k-nearest neighbor.

In 2015, Noda et al. [5] also used AVSR. Data for classification used Japanese speech video that was recorded by six males. They used the convolutional neural network for classification of lip images and used a multistream HMM for AVSR. The input sizes of the convolutional neural network were 16×16 , 32×32 , and 64×64 .

In 2015, Kumaravel [7] classified English words using a histogram method for feature and support vector machines for classification. For recognition, the data of English words were recorded using a camera. In addition, the video included images of 33 people including men and women.

2.3 Viola-Jones Object Detection Algorithm

The Viola–Jones object detection algorithm is one of the most popular algorithms. In 2001, Viola and Jones [18] developed this algorithm, whose advantages are good performance and fast processing speed. This algorithm consists of a Haar feature, integral image, Adaboost, and cascade classifier.

In 2016, Yu and Kim [1] used the Viola–Jones algorithm to extract and classify subject faces. The extracted faces of subjects were classified using the classifying algorithm. He also improved the performance of the Viola–Jones algorithm using

the convolutional neural network to find and extract the facial point [2].

3. Experimental Method

3.1 Database

In this paper, speech videos of Korean words were recorded as classifying data. The speech data of Korean words include speech video of three males speaking 53 words five times for training data and two times for testing data. Then, nine people (six males and three females) were recorded 53 words three times for training data and two times for testing data to check the effectiveness of VSR. In this experiment, six males and three females are that the Korean language is mother tongue.

To record the Korean speech of subjects, a smartphone video camera was used. The camera recorded the voice and image of the subjects with a video frame size of 1920×1080 . The recorded environment included a white wall for the background. Basic lighting was used without additional lights. The proposed experimental process is shown schematically in Figure 3.

3.2 Set of Words

To classify the words, we recorded the subjects speaking 53 words from an emergency medical service vocabulary [10]. The words were selected to test the use in emergency situations. The list of selected words in the experiment is shown in Table 1.

4. VSR Method

To classify the words using lip shape, the proposed VSR method shown in Figure 4 was used in the experiment. First, each

Figure 3. The method of recording the Korean speech.

Table 1. Selected 53 words in emergency medical service vocabulary

Korean	Pronunciation	English
가려움	garyeoum	itch
가슴	gaseum	chest
간호사	ganhosa	nurse
감각이상	gamgag isang	paresthesia
경련	gyeonglyeon	convulsion
경찰	gyeongchal	police
고름	goleum	pus
고열	goyeol	high fever
고혈압	gohyeol-ab	high blood pressure
골절	goljeol	fracture
구급차	gugeubcha	ambulance
구토	guto	throw up
긴급	gingeub	emergency
내장	naejang	guts
뇌진탕	noejintang	concussion
당뇨	dangnyo	diabetes
도와주세요	dowajuseyo	help
사고	sago	accident
살려주세요	sallyeojuseyo	please spare
설사	seolsa	diarrhea
소생	sosaeng	revival
소생술	sosaengsul	resuscitation
식중독	sigjungdog	food poisoning
신고	singo	notify
실신	silsin	faint
심폐소생술	simpye sosaengsul	CPR
어지럼	eojeleom	dizziness
앰블런스	embyulleonseu	ambulance
열	yeol	fever
염증	yeomjeung	Inflammation
응급실	eung-geubsil	emergency room
응급치료	eung-geub chilyo	first aid
의사	uisa	doctor
의식	uisig	consciousness
맥박	maegbag	pulse
멀미	meolmi	motion sickness
목구멍	mogumeong	throat
무감각	mugamgag	stupor
무기력	mugilyeog	lethargy
무의식	muuisig	unconscious
발열	bal-yeol	fever

발작	baljag	seizure
병원	byeong-won	hospital
빈혈	binhyeol	anemia
장염	jang-yeom	enteritis
저혈압	jeohyeol-ab	hypotension
전화	jeonhwa	telephone
주사	jusa	injection
지혈	jihyeol	hemostasis
진통	jintong	throes
환자	hwanja	patient
화상	hwasang	burn
환자	hwanja	patient

Figure 4. The framework for VSR.

speech sound in the recorded video was categorized by consonants, vowels, and final consonants using audio speech signal analysis. The images of consonants, vowels, and final consonants were extracted using the categorized speech sounds. A tracking algorithm found the lip images in the extracted images. Using the lip images of consonants, vowels, and final consonants, a classifier was trained and tested. The words were then classified by the results of the classifier output.

Figure 5. The framework for VSR.

4.1 Categorization of Images

To categorize the recorded video, the video sounds were used. First, the Daum PotEncoder video encoding program was used to extract the video sounds. In the experiment, the categorized images were divided into consonants, vowels, and final consonants using the extracted sound and MATLAB. In MATLAB, the audio data had a threshold of 0.8 to eliminate noise. To find frames with information of consonant images, the starting points of each syllable were used. The final consonant images were found using the endpoint of each syllable. The vowel images were extracted using the mean value of each starting point and endpoint of the syllables. Figure 5 is an example of finding each image using the MATLAB programming and sound file of the recorded video.

4.2 Lip Tracker

To track the lip images, the Viola–Jones algorithm was used. Using the Viola–Jones algorithm, the faces of subject images are extracted. The lip images are then found using the extracted facial images and the Viola–Jones algorithm. The process to find the lip images is shown in Figure 6.

The extracted facial images were resized to $400 \times 400 \times 3$ (RGB data) in the case of VGG-s. The lip images were also resized to $224 \times 224 \times 3$ (RGB data) because the input size of VGG-s is $224 \times 224 \times 3$. However, the video data were encoded into 100 MB in the case of LeNet-5. The video was then resized to $272 \times 480 \times 3$ by the encoding. In the video, the faces of subjects were extracted and then resized to 300×280 . For the input size of LeNet-5, the mouth images extracted from the faces were resized to $32 \times 32 \times 3$. To extract fixed lip images, the minimum and maximum sizes were decided for VGG-s and LeNet-5.

Figure 6. Process to extract the face and lip images using the Viola–Jones detection algorithm: (a) extracted image using the audio data, (b) face detection and extraction using the Viola–Jones detection algorithm, (c) mouth detection using the Viola–Jones detection algorithm, and (d) extracted mouth.

Table 2. The grouping method to classify the vowel images

Number	Korean	English
1	ㅏ, ㅑ	a, ya
2	ㅓ, ㅕ	eo, yeo
3	ㅗ, ㅛ	o, yo
4	ㅜ, ㅠ	u, yu
5	ㅡ	eu
6	ㅣ	I
7	ㅐ, ㅒ, ㅖ	e, ae, ye
8	ㅚ	oe
9	ㅟ	wi

4.3 Grouping the Pronunciations

To group the pronunciations, the vowels, consonants, and final consonants of the syllables depended on the lip shapes and pronunciations. The method to group the vowel is shown in Table 2. The number in Table 2 is the label order.

The consonant consisted of bilabials such as ‘m’, ‘b’, and ‘p’ as well as no bilabial. The final consonant consisted of the bilabial, the other final consonant, and no final consonant. The components of the consonant and final consonant had the labels.

Figure 7. Results of grouping method and the lip images of the vowels, consonants, and final consonants.

The results of categorized lip images are shown in Figure 7.

To classify the lip images, the convolutional neural network is used. The structure of the convolutional neural network is VGG-s, which was developed by the University of Oxford. Using the results of the lip tracker, the lip shape images are trained. The performance of the classification is then checked using the testing lip images. The proposed method of classifying the Korean language is shown in Figure 8. The structure for word classification consists of the bounds of the vowels, consonants, and final consonants. The consonant and final consonant bounds in Figure 8 are used to classify consonants such as ‘m’, ‘b’, and ‘p’. In the case of the final consonant bound, the convolutional neural network also classified the lack of a final consonant. The vowel bound finds vowels ‘a’, ‘e’, ‘i’, ‘o’, and ‘u’. The number of times the 53 words were reproduced was then classified using Euclidean distance with the desired labels, which are sets of each component’s labels, and estimated labels, which are results of classification.

The Euclidean distance to calculate difference with desired and resulting output could be described as below:

$$\sqrt{(t_{1c} - o_{k1c})^2 + (t_{1c} - o_{k1c})^2 + (t_{1c} - o_{k1c})^2 + \dots + (t_{1c} - o_{k1c})^2 + (t_{1c} - o_{k1c})^2 + (t_{1c} - o_{k1c})^2}, \quad (1)$$

where t_n and O_{kn} is resulting and desired output. The inferior letter n is the number of letters of a word. Inferior letter k means a label number from one of 53 words. Inferior letters c , v , and f is consonant, vowel, and final consonant. From results of (1), we selected the answer that has minimum value.

For classification, the VGG-s and modified LeNet-5 of the

Figure 8. Method of word classification.

convolutional neural network structure are used with MatConvNet based on MATLAB [19]. In the case of LeNet-5, the Daum PotEncoder program was used for video encoding owing to its input size of $32 \times 32 \times 3$. The input image size of LeNet-5 was $32 \times 32 \times 3$ (RGB data), and the size of VGG-s was $224 \times 224 \times 3$ (RGB data). The number of consonant output nodes was two, the number of final consonant nodes was three, and the number of vowel nodes was nine.

5. Experiment Result

The 1,989 lip images were used for consonant, final consonant, and vowel training data. For testing classification of the consonant, final consonant, and vowel images, 804 images were used. The structures of LeNet-5 and VGG-s have 50 iterations for training. To train the consonant using VGG-s and LeNet-5, 5,967 images of mouth shapes were used. In addition, 2,412 images were used for testing the consonant. For training and testing the final consonant and vowel, the same images were used. Training and testing data were extracted from the recorded speech video of three subjects. Figure 9 shows the classification results of the test images. The classification results of each subject are shown in Table 3.

From these results, the performance of these classifiers could be compared. Moreover, we could find a more powerful VGG-s for VSR. Using only the lip images, 72.327% of the total classification rates were obtained. In the result, subject 2 has an 80.189% rate, which is the highest value using VGG-s. When VGG-s was used, the lowest value was 65.094%. However, a maximum value of 24.528% and minimum value of 19.811%

Figure 9. Classification results of pronunciation.

Table 3. Classification results of each subject (unit: %)

Structure		Subject1	Subject2	Subject3
VGG-s	Consonant	93.657	95.149	94.030
	Final consonant	76.493	82.836	72.761
	Vowel	79.478	89.552	74.254
	Total	83.209	89.179	80.348
LeNet-5	Consonant	96.269	94.776	94.776
	Final consonant	36.940	24.627	7.090
	Vowel	48.507	36.194	22.388
	Total	60.572	51.866	41.418

Figure 10. Classification results of 53 words by three subjects.

were obtained when LeNet-5 was used. In addition, the total classification rate was 22.327%.

In order to check the performance of this algorithm used VGG-s, nine subject's videos that included three subjects in previous experiments was used. Three times for training data and two times for testing data of nine subject's 53 words videos (six males and three females) were used in this experiment. Total training images were 11,850 and testing lip's images

Figure 11. Classification results of 53 words.

were 7,962. VGG-s have 30 iterations for training. Figure 10 shows the results of the classification rates of 53 words by three subjects. From these results, subject 2 has 48.0769% that is highest value and average value is 32.9554%.

6. Conclusion

The classification rates show that using VGG-s that is the structure of the convolutional neural network was the better method than the LeNet-5 structure of the convolutional neural network for the visual speech recognition. In addition, the performance of this algorithm was checked by nine subject's videos. However, there was the ambiguity of images in the case of the final consonant because not having a final consonant's images and a final consonant without bilabial images make no difference. The vowel images also have similar lip's shape. And, we knew that the label's order is important because of the similar lip's shape at the word classification. If the label is randomly decided, the results will be different of word classification and will not be good because the Euclidean distance was used. In future research, other classification algorithms such as GoogLeNet, deep belief network, and restricted Boltzmann machine will be used for classification of Korean words. We plan to implement and apply a new algorithm for accurate detection and extraction of the lip's images. Furthermore, studies on reducing the delay time needed for the training of convolutional neural network algorithm will be conducted. The results of this research showed that a machine can recognize the person's speech. With further this research and experiments, it will be able to assist the speech-impaired person and the elderly that hard to speak using this technology. Furthermore, it will be possible to help the people in emergency situations with noise and the crime prevention. Thus, visual speech recognition has the potential to be adopted in various human-robot interaction area and assistance

devices for rehabilitation.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

- [1] J. H. Yu and K. B. Sim, "Face classification using cascade facial detection and convolutional neural network," *Journal of Korean Institute of Intelligent Systems*, vol. 26, no. 1, pp. 70-75, 2016. <https://doi.org/10.5391/jkiis.2016.26.1.070>
- [2] J. H. Yu, K. E. Ko, and K. B. Sim, "Facial point classifier using convolution neural network and cascade facial point detector," *Journal of Institute of Control, Robotics and Systems*, vol. 22, no. 3, pp. 241-246, 2016. <https://doi.org/10.5302/j.icros.2016.15.0156>
- [3] J. Li, L. Deng, Y. Gong, and R. Haeb-Umbach, "An overview of noise-robust automatic speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 745-777, 2014. <https://doi.org/10.1109/TASLP.2014.2304637>
- [4] Z. Zhou, G. Zhao, X. Hong, and M. Pietikainen, "A review of recent advances in visual speech decoding," *Image and Vision Computing*, vol. 32, no. 9, pp. 590-605, 2014. <https://doi.org/10.1016/j.imavis.2014.06.004>
- [5] K. Noda, Y. Yamaguchi, K. Nakadai, H. G. Okuno, and T. Ogata, "Audio-visual speech recognition using deep learning," *Applied Intelligence*, vol. 42, no. 4, pp. 722-737, 2015. <https://doi.org/10.1007/s10489-014-0629-7>
- [6] M. L. Seltzer, D. Yu, and Y. Wang, "An investigation of deep neural networks for noise robust speech recognition," in *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, 2013, pp. 7398-7402. <https://doi.org/10.1109/ICASSP.2013.6639100>
- [7] S. S. Kumaravel, "Visual speech recognition using histogram of oriented displacements," M.S. thesis, Clemson University, Clemson, SC, 2015.
- [8] K. O. Yi, "The internal structure of Korean syllables: rhyme or body?," *Korean Journal of Experimental & Cognitive Psychology*, vol. 10, no. 1, pp. 67-83, 1998.
- [9] Y. M. Kwon, "Development of bilabialization (labialization)" *Korean Linguistics*, vol. 47, pp. 93-130, 2010.
- [10] National Emergency Medical Center, *Emergency Medical Dictionary*. Seoul: National Emergency Medical Center, 2005.
- [11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998. <https://doi.org/10.1109/5.726791>
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.
- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. Available <https://arxiv.org/abs/1409.1556>
- [14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, 2015, pp. 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, 2015, pp. 3431-3440.
- [16] C. Bregler and Y. Konig, "'Eigenlips" for robust speech recognition," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Adelaide, Australia, 1994, pp. 669-672. <https://doi.org/10.1109/icassp.1994.389567>
- [17] J. Shin, J. Lee, and D. Kim, "Real-time lip reading system for isolated Korean word recognition," *Pattern Recognition*, vol. 44, no. 3, pp. 559-571, 2011. <https://doi.org/10.1016/j.patcog.2010.09.011>
- [18] P. Viola and M. Jones, "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, Kauai, HI, 2001, pp. 511-518. <https://doi.org/10.1109/CVPR.2001.990517>

- [19] A. Vedaldi and K. Lenc, "MatConvNet: convolutional neural networks for Matlab," in *Proceedings of the 23rd ACM International Conference on Multimedia*, Brisbane, Australia, 2015, pp. 689-692. <https://doi.org/10.1145/2733373.2807412>

Sung-Won Lee received his B.S. degree in electrical and electronic engineering from Seokyeong University, Seoul, Korea, in 2015. He received M.S. degree in electrical and electronic engineering from Chung-Ang University, Seoul, Korea, in 2017. He is currently pursuing the Ph.D. degree in electrical and electronics at Chung-Ang University, Seoul, Korea. His research interests includes IoT, sensor network, embedded, security algorithm.

E-mail: sungwon8912@cau.ac.kr

Je-Hun Yu received his M.S. degree in electrical and electronic engineering from Chung-Ang University, Seoul, Korea, in 2017. His research interests includes brain-computer interface, intention recognition, emotion recognition, intelligent robot, intelligence system, Internet of Things, and big data.

E-mail: yjhoon651@cau.ac.kr

Seung Min Park received B.S., M.S., and Ph.D. degrees from the Department of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Korea, in 2010, 2012, and 2019, respectively. In 2017 and 2018, he joined the Department of Electrical Electronics Engineering, Chung-Ang University, as a Lecturer. His current research interests include machine learning, brain computer interface, pattern recognition, intention recognition and deep learning. Dr. Park was a recipient of the prizes for best paper from the Korean Institute of Intelligent Systems Conference in 2010, 2011, 2015, 2016, 2018 and Student Paper Award from the 13th International Conference on Control, Automation and Systems in 2013. He was the Session Chair of the 7th International Conference on Natural Computation and the 8th International Conference on Fuzzy Systems and Knowledge Discovery (ICNC & FSKD '11) held in Shanghai, China. He is a member of the Korean Institute of Intelligent Systems (KIIS) and Institute of Control, Robotics and Systems (ICROS). He became an IEEE member in 2018.

E-mail: sminpark@cau.ac.kr

Kwee-Bo Sim received the Ph.D. degree in electronic engineering from University of Tokyo, Japan, in 1990. From 2007 to 2007, he was a President of Korean Institute of Intelligent System. Since 1991, he has been an Professor with Department of Electrical and Electronic Engineering, Chung-Ang University, Seoul. His research interests includes artificial life, ubiquitous robotics, intelligent system, soft computing, big-data, deep learning, and recognition. He is a member of the IEEE, SICE, RSJ, IEEK, KIEE, KIIS, KROS, IEMEK, and is an ICROS Fellow.

E-mail: kbsim@cau.ac.kr