ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 20-04-35
228

j.inst.Korean.electr.electron.eng.Vol.24,No.4,1148∼1155,December 2020
http://dx.doi.org/10.7471/ikeee.2020.24.4.1148

Real-time Smoke Detection Research with False Positive Reduction using Spatial and Temporal Features based on Faster R-CNN

Sang-Hoon Lee*, Yeung-Hak Lee**★

Abstract

Fire must be extinguished as quickly as possible because they cause a lot of economic loss and take away precious human lives. Especially, the detection of smoke, which tends to be found first in fire, is of great importance. Smoke detection based on image has many difficulties in algorithm research due to the irregular shape of smoke. In this study, we introduce a new real-time smoke detection algorithm that reduces the detection of false positives generated by irregular smoke shape based on faster r-cnn of factory-installed surveillance cameras. First, we compute the global frame similarity and mean squared error (MSE) to detect the movement of smoke from the input surveillance camera. Second, we use deep learning algorithm (Faster r-cnn) to extract deferred candidate regions. Third, the extracted candidate areas for acting are finally determined using space and temporal features as smoke area. In this study, we proposed a new algorithm using the space and temporal features of global and local frames, which are well-proposed object information, to reduce false positives based on deep learning techniques. The experimental results confirmed that the proposed algorithm has excellent performance by reducing false positives of about 99.0% while maintaining smoke detection performance.

Key words : deep learning; wavelet transform; smoke detection; false positive; temporal and spatial features

Ⅰ. Introduction

Fires come without notice, but if they are well prevented or discovered in the early stages, major disasters can be avoided. Fire is one of the greatest harm to humans of many catastrophes, either artificially or naturally. Because it causes serious economic losses and sometimes lose their precious lives, research on automatic fire detection or monitoring has been noted since the past to protect against casualties and property damage caused by fire.

The sensor operation is sometimes limited by the change in the surrounding environment.

In the event of a fire, it is difficult to detect when the air flow is directed in a different direction or when the excitation is low. And the UV detector used with the flame detection device

* Research Engineer, Gumi Electronics & Information Technology Research Institute
** SW Convergence Education Center, Andong National University
★ Corresponding author
E-mail : yhyi@anu.ac.kr, Tel : +82-54-820-6308
※ Acknowledgment
This work supported by the Disaster Safety Research and Development, 20010305, Development of high-stretch material-based chemical protection suits(Level B, C class) with activity and safety through wearer's bio-signal measurement and gas detection, funded by the Ministry of Public Administration and Security.
Manuscript received Dec. 3, 2020; revised Dec. 29, 2020; accepted Dec. 29, 2020.

will lose its sensitivity when it absorbs UV rays from the surrounding environment. Systems that detect the presence of smoke and fire-generated particles utilize chemical reactions by ionization.

In recent years, the development of CCTV technology and the improvement of computer performance has greatly improved the research on fire detection systems through video analysis over the last 10 years. Smoke detection using computer vision has been researched in various fields, and has many advantages over the method using conventional sensors[1]. Smoke appears before the start of a full-scale fire, so early detection can prevent the spread of fire. Until now, many researchers using different algorithms have detected smoke and fire flames based on video images. Most of these research and development algorithms focus on color and shape characteristics combined with the transient motion of smoke and flames. Gubbi [2] used wavelet transform and discrete cosine transform for feature extraction to detect smoke from an input image, and a support vector machine for classification.

Ko [3] extracted features such as color, ripple coefficient, motion vector, and nine feature vectors from the input image for smoke detection, and used a random forest for classification. Frizzi [1] developed an algorithm that detects smoke and fire using Tensorflow's deep convolutional neural network. Chen [4] and ÇeliK [5] established a color model for recognizing flame and fire smoke. This method extracts various feature extraction algorithms or multidimensional feature vectors and uses classification algorithms such as neural networks, Adaboost, and SVM. [6-7].

Another approach to fire and smoke detection is to use artificial intelligence algorithms (Deep Learning). In Deep Learning, the shape or expression of object data is trained, and smoke and fire are detected using the training result. Convolutional Neural Networks, used in Deep Learning, are a variant of deep learning training that can extract topological features.

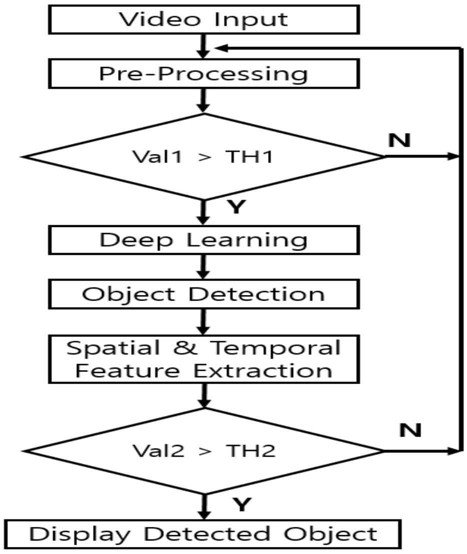


Fig. 1. Proposed algorithm included with preporcessing and post processing.

A new CNN with reduced error attenuation was proposed by LeCun [8], and Krizhevsky [9] applied a general CNN for object recognition. This advanced method is a method using artificial intelligence that dramatically improves the existing object recognition method and obtains normalized values using ReLU (Rectified Linear Unit), activation function, and dropout [10].

Common deep learning methods are done for moving or stationary objects. However, smoke and flames are atypical objects and are difficult to detect because they show a good resemblance to areas of reflective objects such as sunlight. In this paper, a deep learning algorithm is used to obtain a high detection rate, and a new algorithm is proposed to reduce false positive smoke detection, as shown in picture 1.

II. Faster r-cnn

There are several ways to distinguish objects. However, it may be difficult to distinguish objects in an image. Basically, deep learning using the R-CNN method goes through several steps. R-CNN creates a number of region regions by determining the initial segment of the image for the region in which the object is present. The second is to combine similar regions around each region using the greedy algorithm.

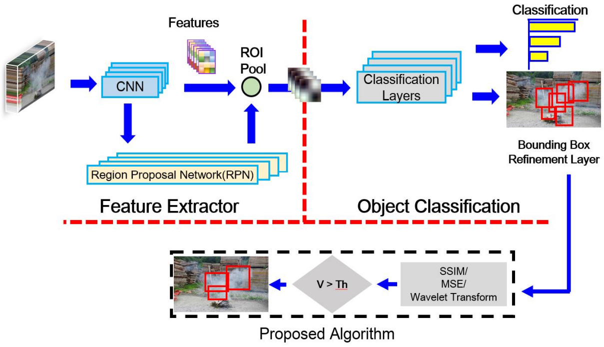


Fig. 2. Proposed Algorithm with Faster r-cnn.

The third is to propose the combined and enlarged region as the final region proposal and classify the selected region using SVM. Finally, using a linear regression model, the bounding box of the classified object sets the exact coordinates. The CNN for the training data proposed by this paper is divided into three parts. Figure 2 shows the overall flow of the proposed system.

The biggest problem with R-CNN is the excessive amount of association due to complex processes. In recent years, high-performance GPUs have become popular, so even deep neural nets can be processed quickly through GPU computation. However, since R-CNN's region proposal work and NMS algorithm work through the selective search algorithm are performed by CPU calculation, it consumes a lot of computation and time. In addition, since the classification of the region and the regression of the bounding box work together during SVM prediction, there is a disadvantage in that real-time analysis is difficult because computation and time are consumed a lot in model prediction.

This model was developed to solve the problem of R-CNN of Fast R-CNN. The operation method is similar to R-CNN, but region proposal works, but unlike RCNN, in Fast R-CNN, the entire image is first input as input of ConvNet. The image passes through ConvNet to extract a feature map, and this feature map extracts Regions of Interest (RoI) through region proposal based on selectice search. Afterwards, the selected regions go through the RoI Pooling layer. This process serves to downsize the regions for future prediction and convert them all to the same fixed size. As the last step, Softmax Classification and Bounding Box Regression are performed through a fully connected layer. Faster R-CNN is a model in which R-CNN and Fast R-CNN are developed to solve the excessive computational problem caused by region proposals. The selective search, which was used in the existing region proposal, was the main reason for increasing the computational amount and consuming a lot of time. So, in Faster R-CNN, the selective search algorithm was removed and a neural network called Region Proposal Networks (RPN) was added to predict region proposals. After that, the predicted region proposal goes through the RoI pooling layer similar to Fast R-CNN, and after fixing all regions to the same size, classification and bounding box recognition are performed.

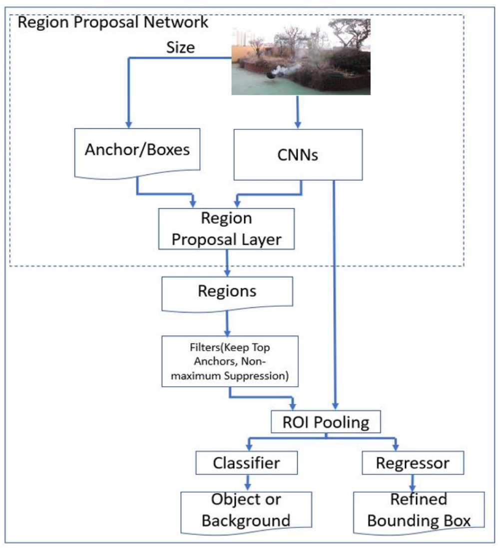


Fig. 3. The Architecture of Faster R-CNN.

III. Features of Smoke

1. Structure Similarity

Structural Similarity (SSIM) [15] is a method to measure the similarity between the original image and the distortion due to compression and transformation. This shows higher accuracy than the Mean Square Error (MSE) and Peak Signal Noise Ratio (PSNR) methods. This is an evaluation of the test image (X) for the original image (Y) to quantify the visual similarity. A value close to 1.0 means that the test image is similar to the original image, and a value close to 0.0 means that the test image is very different from the original image. The SSIM formulas are defined as follows,

$$l(x,y) = \frac{2\mu_x\mu_y + K1}{\mu_x^2 + \mu_y^2 + K1} \tag{1}$$

$$m(x,y) = \frac{2\sigma_x\sigma_y + K2}{\sigma_x^2 + \sigma_y^2 + K2} \tag{2}$$

$$n(x,y) = \frac{\sigma_{xy} + K3}{\sigma_x\sigma_y + K3} \tag{3}$$

where μ_x and μ_y are mean of pixels, σ_x and σ_y are standard deviation, and σ_{xy} is covariance. *K1, K2,* and *K3* are constant for preventing the denominator and numerator from becoming zero. $l(x,y)$ is the relationship of the brightness difference, $m(x,y)$ is the contrast difference, and $n(x,y)$ is the similarity of the structural change between x and y. The structural similarity is shown in Equation (4).

$$SSIM = [l(x,y)]^{\alpha}[m(x,y)]^{\beta}[n(x,y)]^{\gamma} \tag{4}$$

where α, β, and γ represent the importance of each term, 1.0 was used in this paper.

2. RGB Color Histogram

In general, smoke has a grayish color (dark gray, gray, light gray and white). And black smoke is produced by unburned material or combustion at high temperatures, which means that a lot of fires have already occurred. In this paper, it is important to detect smoke first in order to extinguish a fire in the beginning, so focus on smoke detection and set the conditions as shown in Equation 5 to use the smoke color from gray to white.

$$C = (R+G+B)/3,\ C_{Th1} < C < C_{Th2} \tag{5}$$

where C is output image. This research set the C_{Th1} is minimum value 80 and C_{Th2} is maximum value 250. The average image C is histogramized into 256 bins (0 to 255) for each pixel. Depending on the input image size, the values stored in each bin of the histogram are normalized and the sum is obtained as Equation (6).

$$H_S = \sum_{i=80}^{250}(b_i/(h\times w)) \tag{6}$$

where H_S is the RGB color histogram result value. b_i means the histogram bins from 80 to 250, h and w is height and width for input image. The grayish color is distributed intensively between 80 and 250.

3. Wavelet Transform

The characteristic of a wavelet is a building block for constructing or expressing a signal or function, which can easily localize a signal in time-frequency. The wavelet can know not only the frequency component but also the location information of the low and high frequency components on the time axis. A circular wavelet with a real value can be expressed as Equation (7).

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{x-b}{a}\right),\ \Psi(t)\in L^2(R), a,b\in R \tag{7}$$

where a and b are real numbers, a is expansion and contraction, and b is movement.

In this paper, the filter processing method proposed by Mallat [8], which is used in actual image processing, is processed in two directions, horizontal or vertical, in the filter process. The decomposition of the discrete signal is obtained by down-sampling and coefficients of the low-pass filter and high-pass filter as shown in Equations (8) and (9). In this paper, the filter processing method proposed by Mallat [8], which is used in actual image processing, is processed in two directions, horizontal or vertical, in the filter process. The decomposition of the discrete signal is obtained by down-sampling and coefficients of the low-pass filter and high-pass filter as shown in Equations (8) and (9).

$$c_j(n)=\sum_{k=0}^{N-1}h(k-2n)c_{j+1}(k) \tag{8}$$

$$d_j(n)=\sum_{k=0}^{N-1}g(k-2n)c_{j+1}(k) \tag{9}$$

In this paper, the method of applying the wavelet transform to the appearance of the smoke is by convolution with low-pass filter coefficients and down-sampling. The Mallat algorithm applied in the paper is shown in Figure 4. After performing $H(x)$, a low-frequency band filtering process, and down-sampling the selected bounding box using deep learning, $f_L(x,y)$, a low-frequency component, and $G(x)$, a high-frequency band filtering process, are performed. It can be decomposed into the component $f_H(x,y)$.

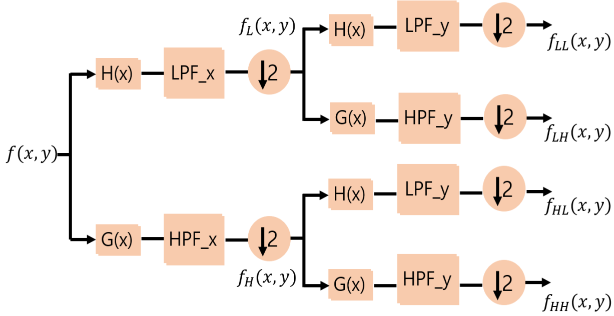


Fig. 4. Two-stage discrete wavelet decomposition.

$f_{LH}(x,y)$ is a horizontal high-frequency component and contains horizontal edge information of smoke. $f_{HL}(x,y)$ is a vertical high frequency component and contains vertical edge information that a smoke has. $f_{HH}(x,y)$ is a diagonal high-frequency component and has edge information in a diagonal direction taking into account the horizontal and vertical edge information of a smoke. Figure 5 shows the wavelet decomposition results for smoke.

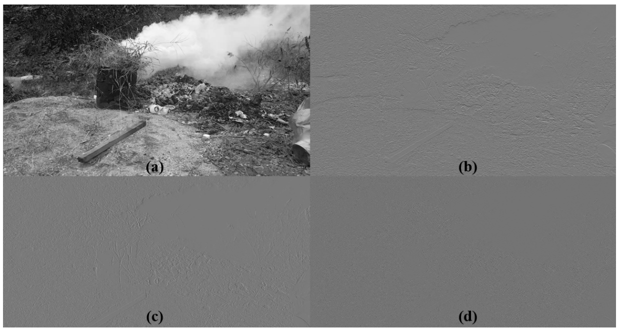


Fig. 5. Wavelet decomposition result of smoke image, (a) original image, (b) horizontal decomposition, (c) vertical decomposition, and (d) diagonal decomposition.

Figure 5 shows the wavelet stage one process, where the continuous operation up to the desired decomposition step for wavelet decomposition is achieved through a low frequency band to obtain a stage two wavelet decomposition image.

Ⅳ. Experimental Results

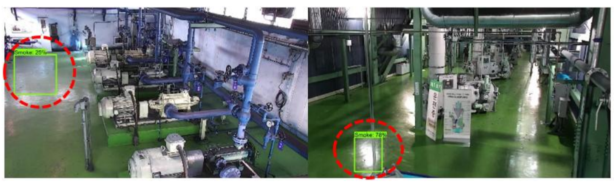
We proposed a new algorithm that uses a camera to detect smoke in real time and uses color histograms and similarities between full and local frames to reduce smoke false positives. The system used in this study was tested in an end-user computer environment consisting of an Intel Core i7-7700 (3.5GHz), memory 16G, and Ge-force TITAN-X. The smoke database used in this study was video recorded directly from the general ground and factory. The video recording devices are mobile phone camera, raspberry pi camera and Cannon G5.

The proposed algorithm proceeds in three steps. The first is the labeling of the data set using the LabelImg program. Second, it is a training course on labeled images. Labeled images are converted to TFRecord file format to prevent video performance degradation, and the entire training data is classified and saved into 70% training data and 30% verification data. As the basic model for training, FASTER-CNN ResNet (Deep Residual Network), which represents the smallest number of objects and the highest detection rate, was used, and 21,230 smoke images were used. The third process is the extraction of the training model. Each check pointer contains meta information about the model in Tensorflow model file format. However, since there is a lot of unnecessary information in the ".meta" file, the final ".pb" file is created by excluding unnecessary data in the ".meta" file and combining the weights.

In this paper, we used a factory recorded video images as the experimental data. Figure 6 shows an example of a continuous frame of video used in the experiment. Basic smoke detection experiment was performed using ".pb" file based on Fater R-CNN model. Figure 7 shows smoke detection results included true positive and false positive using general deep learning.

Fig. 6. Example of the frame sequence of test video.

(a)



(b)

Fig. 7. The experimental results using the Faster R-CNN, (a) smoke detection, (b) the false positive.

Figure 7 (a) shows the smoke detection experiment results using various videos. The detection threshold of Faster R-CNN was 30% or more. Figure 7 (b) shows the false positive detection results. Various types of false positives appeared in several videos used in the experiment. These false positives could be divided into two types. First, it is detected like smoke by reflection of sunlight. Second, if the facilities inside and outside the factory have a shape or color similar to smoke, it was found as a false positive. Table 1 shows the smoke detection results for the videos.

Table 1. The results of video test using general Faster R-CNN.

Videos	Type	True Positive	False Positive
Video 1	Smoke	1307	0
Video 2	Smoke	1293	0
Video 3	Smoke	761	0
Video 4	Non-smoke	0	23
Video 5	Non-smoke	0	474

Video 1, Video 2 and Video 3 are smoke video. Video 4 and Video are non-smoke video (inside the factory). As shown in Table 1, inside the plant, false positives appear in several places due to the influence of sunlight. False Positive (FP) is a case where non-smoke objects are detected as a smoke.

$$Precision = TP/(TP+FP) \tag{10}$$

In order to reduce false positives generated in Video 4 and Video 5, we use the following characteristics. From the global check, we extracted the motion characteristics as preprocessing using equation (10), (11). Since there is a motion when a smoke occurs, if a block of moving pixels are generated, it is registered as a smoke candidate area. If the smoke candidate frame status is 1, proposed algorithm is performed, as shown Figure 1.

$$\begin{aligned} S_k &= SSIM(f_i, f_j) \\ M_k &= MSE(f_i, f_j) \\ A_k &= diff(f_i, f_j) \end{aligned} \tag{11}$$

$$V_G = \begin{cases} 1 & if\ S_k \le th1,\ M_k \ge th2,\ A_k > th3 \\ 0 & else \end{cases} \tag{12}$$

Second, it is local check. In this paper, we determine the final smoke region using the color histogram H, SSIM index, mean square error (MSE), and wavelet transform with other frames as below equation,

$$V_G = \begin{cases} 1 & if\ S_k \le th4,\ M_k \ge th5, A_k \ge th6, H_S > th7 \\ 0 & else \end{cases} \tag{13}$$

$$V_{WT} = \begin{cases} 1 & if\ W_{HL-LH} \le th8,\ W_{HH} \ge th9 \\ 0 & else \end{cases} \tag{14}$$

$$V_S = \begin{cases} 1 & if\ V_G, V_{WT} \\ 0 & else \end{cases} \tag{15}$$

where k means frames, from $th1$ to $th9$ are threshold value by experiment. We compared the local region (bounding box area) of interest using the 3 out of 10 frames (first, middle, and last frames).

Table 2. The results of video test using prosed algorithm.

Videos	Type	True Positive	False Positive
Video 1	Smoke	1442	0
Video 2	Smoke	1292	0
Video 3	Smoke	789	0
Video 4	Non-smoke	0	0
Video 5	Non-smoke	0	19

Table 2 shows the experimental results using the proposed algorithm. Video 1 and Video 3 increased true positives. And in Video 2, one frame was detected less. This is not a big problem since you are missing one frame. In non-smoke video, significant false positives were reduced by using the proposed algorithm. The case of video4 is completely gone, and video5 still has a problem to be solved.

The reason false positives do not disappear in the Video 5 experiment is because the object is detected as smoke. This is the case where the shape of this part is deformed and detected as smoke. It is necessary to adjust the final threshold considering various factors, or an algorithm using additional color, wavelet and feature analysis for the smoke.

Ⅴ. Conclusion

The start of a fire always starts with a small ember. These small embers create smoke and large flames that take our precious lives and lose our valuable wealth. In this paper, we studied a new smoke detection algorithm that improves false positives while maintaining the performance generated when real-time smoke detection based on deep learning for surveillance cameras. In general, the deep learning technique using the shape of an object frequently generates false positives that a general object is detected as fire or smoke. To solve this problem, we first used motion detection using a three-frame difference algorithm, and then applied frame similarity using SSIM and MSE. Second, the Faster R-CNN algorithm was applied to find the postponement candidate region for the detected frame. Third, spatial and temporal methods were used to extract local features; The final smoke area was determined using an RGB color histogram, frame similarity, MSE, and wavelet transform that represents the edge well. As a result of the experiment, it was found that the proposed algorithm significantly reduced false positives while maintaining the detection rate compared to conventional deep learning.

As future work, more diverse experimental environments will need to be applied, and research on various algorithm development to reduce false positives is required.

References

[1] B. C. Ko, S. J. Han, and J. Y. Nam, “Modeling and Formalization of Fuzzy Finite Automata for Detection of Irregular fire Flames,” *IEEE Transaction on Circuits and System Video Technology*, vol.21, no.12, pp.1903-1912, 2011. DOI: 10.1109/TCSVT.2011.2157190

[2] J. Gubbi and S. Marimuthu, “Smoke detection

in video using wavelets and support vector machines," *Fire Safety Journal*, vol.44, no.8, pp.1110-1115, 2009. DOI: 10.1016/j.firesaf.2009.08.003

[3] B. C. Ko, J. Y. Kwak, and J. Y. Nam, "Wildfire smoke detection using temporospatial features and random forest classifiers," *Optical Engineering*, vol.51, 2012. DOI: 10.1117/1.OE.51.1.017208

[4] T. Chen, P. Wu, and Y. Chiou, "An Early Fire-Detection Method Based on Image processing," *2004 International Conference on Image Processing*, pp.1707-1707, 2004. DOI: 10.1109/ICIP.2004.1421401

[5] T. B. CelikT, Y. Dedeoglu, and A. E. Cetin, "Wavelet Based Real-time Smoke Detection in Video," *EUSIPO 2005*, 2005.

[6] C. Yu, Z. Mei, and X. Zhang, "A Real-time Video Fire Flame and Smoke Detection Algorithm," *Procedia Engineering*, vol.62, pp.891-898, 2013. DOI: 10.1016/j.proeng.2013.08.140

[7] S. Frizzi, R. Kaabi, M. Bouchouicha, J. Ginoux, E. Moreau, and F. Fnaiech, "Convolutional Neural Network for Video Fire and Smoke Detection," *IECON 2016*, 2016. DOI: 10.1109/IECON.2016.7793196

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Hsffner, "Gradient-based Learning Applied to Document Recognition," *Proceeding of IEEE*, vol.86, no.11, pp.2278-2324, 1998. DOI: 10.1109/5.726791

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep convolutional neural Network: Advanced in Neural Information Processing System 25," *26th Annual Conference on Neural Information Processing Systems*, pp. 1106-1114, 2012.

[10] G. E. Hinton, N. Srivastava, A. Krizhevsky, S. Ilya, and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *Clinical Orthopedics and Related Research 2012*, pp.1-18, 2012.

BIOGRAPHY

Sang-Hun Lee (Member)

2002 : BS degree in Rehabilitation Engineering, Daegu University.
2005 : MS degree in Biomedical Engineering, Yeungnam University.
2011 : PhD degree in Electronic Engineering, Yeungnam University.
2012~ : Research Engineer, Gumi Electronics & Information Technology Research Institute

Yeung-Hak Lee (Member)

1988 : BS degree in Electronic Engineering, Yeungnam University.
1991 : MS degree in Electronic Engineering, Yeungnam University.
2003 : PhD degree in Electronic Engineering, Yeungnam University.
1991~1995 : Research Engineer, LG Precision.
2005~2006 : Postdoc, Cardiff University
2010~2016 : Professor, Kyungwoon University
2017~2018 : Research Engineer, Andong National University
2019~ : Professor, Andong National University