ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 20-03-13

j.inst.Korean.electr.electron.eng.Vol.24,No.3,777~782,September 2020
http://dx.doi.org/10.7471/ikeee.2020.24.3.777

도플러 레이다 및 음성 센서를 활용한 CNN 기반 HMI 시스템 설계 및 구현

Design and Implementation of CNN-based HMI System using Doppler Radar and Voice Sensor

오 승 현*, 배 찬 희*, 김 세 령*, 조 재 찬*, 정 윤 호*★

Seunghyun Oh*, Chanhee Bae*, Seryeong Kim*, Jaechan Cho*, Yunho Jung*★

Abstract

In this paper, we propose CNN-based HMI system using Doppler radar and voice sensor, and present hardware design and implementation results. To overcome the limitation of single sensor monitoring, the proposed HMI system combines data from two sensors to improve performance. The proposed system exhibits improved performance by 3.5% and 12% compared to a single radar and voice sensor-based classifier in noisy environment. In addition, hardware to accelerate the complex computational unit of CNN is implemented and verified on the FPGA test system. As a result of performance evaluation, the proposed HMI acceleration platform can be processed with 95% reduction in computation time compared to a single software-based design.

요 약

본 논문에서는 도플러 레이다와 음성 센서를 이용한 CNN 기반 HMI 시스템을 제안하고, 가속을 위한 하드웨어 설계 및 구현 결과를 제시한다. 단일 센서 모니터링의 한계를 극복하기 위해 제안된 HMI 시스템은 두 센서의 데이터를 융합 처리하여 분류 성능을 개선했다. 제안된 시스템은 다양한 노이즈 환경에서 단일 레이다 및 음성 센서 기반 분류기에 비해 3.5% 및 12% 향상된 성능을 나타냈다. 또한, CNN의 복잡한 연산부를 가속하기 위해 설계된 하드웨어를 FPGA 디바이스 상에서 구현 및 검증하였다. 성능 평가 결과, 제안된 HMI 가속 플랫폼은 단일 소프트웨어 기반 구조에 비해 연산 시간을 95% 단축 가능한 것을 확인하였다.

Key words : accelerator, convolutional neural network, FPGA, human machine interface, sensor fusion

Ⅰ. 서론

최근 사람의 움직임을 인지하고 기기를 다루는 HMI(human machine interface) 시스템은 스마트 가전, 스마트카 등 다양한 분야에서 필요성이 대두되고 있다[1]. 이 중 사용자의 명령을 내장된 센서로 감지하고 추출된 데이터를 학습하여 효율적으로 기기를 제어할 수 있는 지능형 HMI 시스템에 대한 연구가 활발히 진행되고 있다[2]. 대표적으로 레이다, 음성, 카메라, IMU 센서 기반 HMI 시스템이 우수한 성능을 보이며 다양한 기법이 제안되었다[3-5]. 하지만, 카메라로 추출된 이미지 데이터는 높은 연산량을 요구하여 일반적으로 저면적, 저전

* School of Electronics and Information Engineering, Korea Aerospace University
★ Corresponding author
E-mail : yjung@kau.ac.kr, Tel : +82-2-300-0133
※ Acknowledgment
This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-00056) and CAD tools were supported by IDEC.
Manuscript received Aug. 31, 2020; revised Sep. 14, 2020; accepted Sep. 16, 2020.

력 플랫폼으로 구현되는 HMI 시스템에 적합하지 않으며, IMU 센서는 사용자가 착용해야 하는 단점이 존재한다[4]. 따라서, 레이다 및 음성 센서는 우수한 성능과 더불어 연산량 및 사용자 편의성 측면에서 HMI 시스템에 가장 적합하다[5].

하지만, 단일 센서 정보를 학습 및 인식할 경우, 특정 환경에서 성능이 급격히 저하되는 한계가 있다[6-10]. 예를 들어, 레이다는 전파를 보내서 반사되는 신호로 정보를 획득하기 때문에 빛이 없는 환경에서 제한이 없다는 장점이 있지만, 클러터가 많은 환경에서는 제한이 발생한다[9]. 음성 센서는 사용자의 소리 정보를 통해 명령을 수행하여 빛이 없는 환경에 대한 제한이 없지만, 소음이 많은 환경에서 제한이 발생한다[10].

이에, 본 논문에서는 단일 센서의 제한적인 환경을 해결하기 위해 레이다와 음성 센서 정보를 융합한 지능형 HMI 시스템을 제안한다. 제안된 시스템은 CNN(convolutional neural network)을 통해 레이다 및 음성 데이터를 융합처리 하여 우수한 분류 성능을 나타낸다. CNN 구조는 다양한 노이즈 환경에서의 성능 평가를 통해 최소한의 복잡도로 높은 분류 성능을 보이는 모델을 선정하였으며, 연산 복잡도가 가장 큰 계층에 대한 하드웨어 구조 설계 및 가속화를 통해 최적의 연산 시간을 지원한다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 제안된 시스템에 사용되는 STFT(short time Fourier transform) 및 CNN에 대해 설명하며, Ⅲ장에서는 제안된 HMI 시스템 및 성능 평가 결과와 연산 가속화를 위한 하드웨어 구조설계 결과를 제시하고, 끝으로 Ⅳ장에서 본 논문의 결론을 맺는다.

Ⅱ. Backgrounds

1. STFT

시간 영역의 데이터를 주파수 영역으로 변환하기 위하여 DFT(discrete Fourier transform) 연산이 사용된다. 그러나 DFT 연산은 데이터 구간 전체에 대한 주파수를 반환하기 때문에 음성과 제스처와 같이 시간에 따라 구성 주파수가 달라지는 시간 의존적 데이터에 적용하기 어렵다. 이러한 시간 의존적 데이터는 시 구간을 짧게 나누는 윈도우 연산과 분할된 각 구간에 DFT 연산을 함으로써, 짧은 시 구간에 대한 주파수를 얻을 수 있는 STFT를 사용하여 시간 변화에 대한 주파수를 얻는다[11]. STFT는 식(1)으로 표현되고, ω는 윈도우 함수, τ는 윈도우 지연시간을 뜻한다.

$$X(\tau,f)=\int_{-\infty}^{+\infty} x(t)\cdot w(t-\tau)\cdot\exp(-j2\pi ft)dt \qquad (1)$$

STFT 결과로 얻은 함수 X를 절대값으로 표현한 것을 spectrogram이라 한다.

2. CNN

LeCun [12]에 의해 제시된 CNN은 학습한 필터를 입력 데이터와 convolution 연산함으로써 입력 데이터의 유효한 특징을 크기와 위치에 무관하게 추출할 수 있어, 특징의 크기와 위치 변형이 많은 이미지 인식에서 뛰어난 성능을 보인다[13, 14]. CNN은 특징 추출 계층인 convolution layer와 분류를 위한 계층인 FCL(fully connected layer)로 구성된다. Convolution layer는 학습된 필터와 입력 데이터를 convolution 연산하여 유효한 특징을 추출한다. FCL은 완전 연결된 노드들의 연산을 통하여 필터 연산으로 추출된 데이터를 하나의 class로 분류하는 역할을 수행한다.

Ⅲ. 제안된 HMI 시스템

1. 시스템 구조

그림 1과 같이 제안된 HMI 시스템의 CNN 분류기는 음성 센서를 이용하여 얻은 음성 command와 도플러 레이다를 이용하여 얻은 제스처 command를 입력으로 사용한다. 두 종류의 command 데이터에 대해 STFT 기반의 신호처리를 적용하여

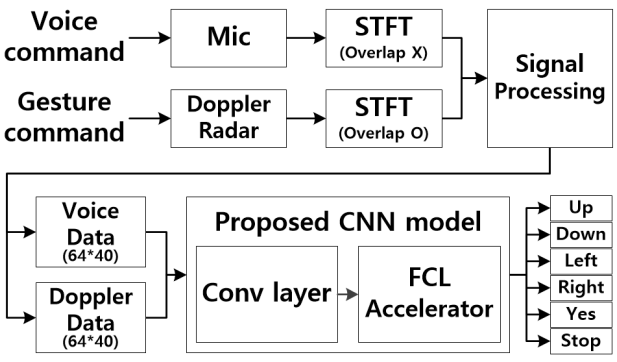

Fig. 1. Overall scheme of the proposed HMI system.
그림 1. 제안된 HMI 시스템 개요.

spectrogram 형태로 변환 후 CNN 분류기에 입력한다. 신호처리는 STFT 결과 중 불필요한 영역을 제거하여 CNN 분류기의 복잡도와 메모리 사용량을 줄이고, 서로 다른 종류의 spectrogram을 융합 학습 및 인식이 가능하도록 동일한 형태로 가공한다.

신호처리가 끝난 2채널 spectrogram 데이터는 제안된 CNN 분류기를 통해 학습 및 인식되며, 네트워크 구조는 그림 2와 같이 특징 추출을 위한 2 convolution layer와 분류를 위한 1 FCL로 구성된다. Convolution layer를 통해 추출된 특징 데이터는 최종적으로 FCL로 입력되고, 가장 큰 값을 갖는 class를 출력함으로써 인식 과정이 완료된다.

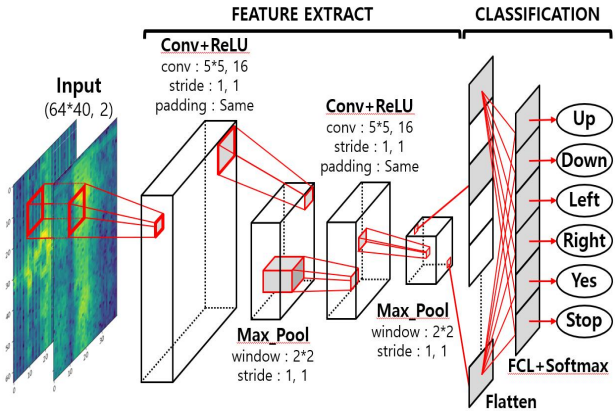

Fig. 2. Network structure of the proposed CNN classifier.
그림 2. 제안된 CNN 분류기 네트워크 구조

2. 센서 신호처리

(1) 도플러 레이다 센서

다양한 레이다 중 도플러 레이다는 타겟의 움직임이나 제스처에 의해 발생하는 도플러 효과를 이용해 속도를 측정한다[15]. 타겟이 레이더를 향해 다가오면 송신된 연속파보다 수신된 연속파의 주파수가 더 높아지고, 타겟이 레이더에서 멀어지면 송신된 연속파보다 수신된 연속파의 주파수가 더 낮아지는 현상이 도플러 효과이다. 도플러 효과를 통해 타겟의 속도를 계산할 수 있으며, 이를 STFT하면 짧은 시간 신호의 변화에 대한 도플러 주파수를 알 수 있다[16].

제안된 HMI 시스템에서는 중심주파수는 24GHz이며, bandwidth는 250MHz인 연속파를 사용하는 도플러 레이다를 이용해 사용자의 command로 활용될 손동작을 직접 추출하였다. 그림 3과 같이 손을 위로 swipe, 손을 아래로 swipe, 손을 왼쪽으로 swipe, 손을 오른쪽으로 swipe, 검지를 시계방향으로 계속 돌리는 동작, 손바닥을 레이더 정면으로 쭉 뻗는 동작으로 구성된 총 6개의 손동작을 up, down, left, right, yes, stop으로 정의하였다. 레이다 샘플링 주파수는 3000Hz로 3200개를 샘플링하여 STFT후 DC offset을 제거하여 도플러 주파수를 얻었다. STFT는 128 point hamming window에 64 point overlap을 적용하였고, FFT는 128 point로 진행하였다. STFT 결과를 주파수축에서 0Hz를 중심으로 64 point, 시간축에서 시작점부터 40 point를 잘라내어 spectrogram을 그림 4와 같이 얻었다.

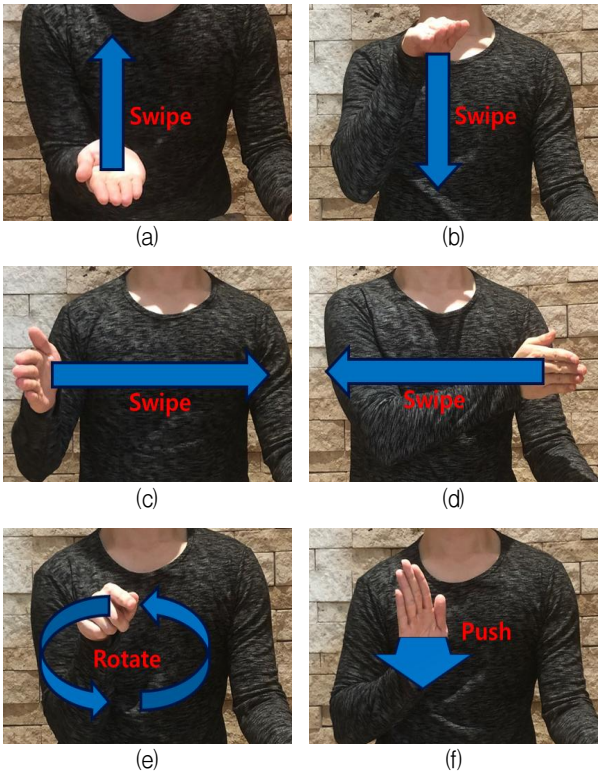

Fig. 3. Hand gesture examples : (a) up, (b) down, (c) left, (d) right, (e) yes, (f) stop.
그림 3. 손동작 예시 : (a) 위, (b) 아래, (c) 왼쪽, (d) 오른쪽, (e) 승인, (f) 정지

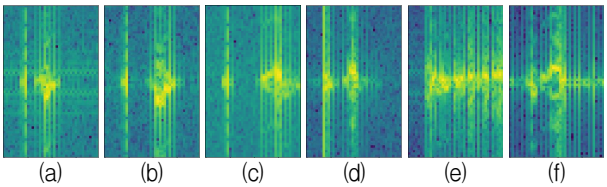

Fig. 4. Radar data spectrogram : (a) up, (b) down, (c) left, (d) right, (e) yes, (f) stop.
그림 4. 레이다 데이터 스펙트로그램 : (a) 위, (b) 아래, (c) 왼쪽, (d) 오른쪽, (e) 승인, (f) 정지

(2) 음성 센서

음성 command를 STFT하면 시간에 따른 주파수 성분을 구할 수 있다. 본 논문에서는 음성 command로 TensorFlow와 AIY에서 만든 speech command dataset을 사용하였다[17]. 음성 command는 총 6개 class로 구성되고 up, down, left, right, yes, stop으로 도플러 레이다로 추출한 데이터와 각각 매칭된다. 음성 데이터의 샘플링 주파수는 8000Hz이며, STFT는 overlap 없이 128 point hamming window를 적용하였고, FFT는 128 point로 진행하였다. STFT 결과를 주파수축에서 0Hz를 기준으로 64 point로 추출하였고, 시간축에서 음성의 최대 파워가 나타나는 시간을 중심으로 40 point로 crop하여 그림 5와 같은 spectrogram을 얻었다.

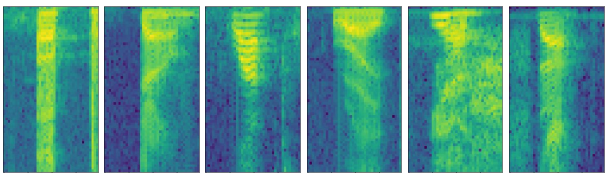

Fig. 5. Voice data spectrogram : (a) up, (b) down, (c) left, (d) right, (e) yes, (f) stop.

그림 5. 음성 데이터 스펙트로그램 : (a) 위, (b) 아래, (c) 왼쪽, (d) 오른쪽, (e) 승인, (f) 정지

3. 성능 평가 결과

성능 평가를 위한 데이터 셋은 레이다 command 6000개, 음성 command 6000개를 사용하였고, up, down, left, right, yes, stop 6개로 class를 구분하였다. 학습을 위해 사용된 데이터는 각각 5400(90%)개이고, 검증을 위한 데이터는 각각 600(10%)개이다. 학습은 cross entropy loss function과 adam optimizer를 사용하였으며 learning rate는 0.001, batch size는 200, epoch는 20으로 진행하였다.

제안된 HMI 시스템을 통해 레이다와 음성신호를 융합한 데이터에 대해 학습 및 분류를 수행하고, 그 결과를 단일 센서 시스템 결과와 비교하였다. 또한, 제한된 환경에서의 동작을 검증하기 위해 다양한 노이즈 환경에서 성능 평가를 수행하였다. 성능 평가 결과, 표 1과 같이 제안된 시스템은 단일 센서 시스템보다 노이즈 환경에서 평균 7.7% 우수한 성능을 보이는 것을 확인하였다.

Table 1. Simulation results in various environment.

표 1. 다양한 환경에서의 성능평가 결과

SNR	Radar	Mic	Fusion	Performance Comparison	
				vs. Radar	vs. Mic
0dB	88.6%	75.6%	92.5%	+3.9%	+16.9%
3dB	89.5%	77.8%	94.0%	+4.5%	+16.2%
6dB	91.6%	81.3%	94.6%	+3.0%	+13.3%
10dB	92.0%	85.5%	96.0%	+4.0%	+10.5%
13dB	92.3%	85.8%	94.8%	+2.5%	+9.0%
16dB	92.5%	87.0%	96.8%	+4.3%	+9.8%
20dB	94.3%	88.3%	96.8%	+2.5%	+8.5%

4. 가속 하드웨어 구조 설계

제안된 HMI 시스템의 연산 시간을 줄이기 위해 연산 복잡도가 가장 높은 CNN 분류기 내부 FCL에 대한 가속 하드웨어 구조 설계를 진행하였다. 그림 6은 설계된 CNN 가속 하드웨어와 통합 시스템 검증을 위한 Xilinx PYNQ-Z1 FPGA 기반 SoC 플랫폼의 구조도로, MCU(micro control unit)와 H/W IP간의 AMBA 버스 통신을 위한 AXI interface 및 설계된 FCL 연산 가속기로 구성된다.

설계된 FCL 연산 가속기는 Verilog-HDL로 작성하였으며, FPGA 기반 구현 결과, 표 2와 같이 최대 110.57MHz의 동작 주파수에서 연산 가능함을 확인하였다. 테스트 데이터 셋에 대한 검증 결과, 100 샘플 데이터에 대해 제안된 가속 하드웨어 기반 HMI 시스템은 76.9ms로 기존 software 기반 시스템 대비 95.6% 감소된 연산 시간을 나타냈다.

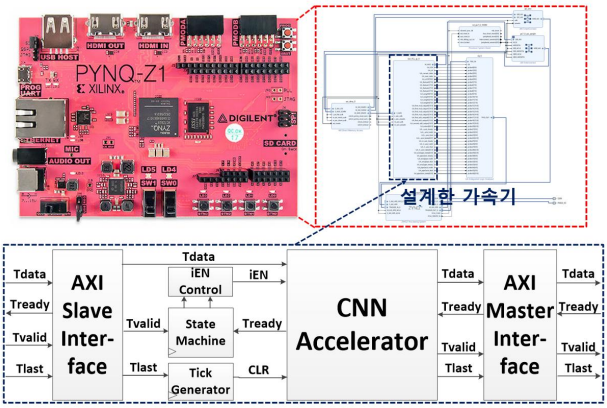

Fig. 6. Block diagram of the proposed FCL accelerator based HMI system.

그림 6. 제안된 FCL 가속 하드웨어 기반 HMI 시스템 블록도

Table 2. Verification results of proposed HMI system.
표 2. 제안된 HMI 시스템의 검증 결과

Parameter	Value
Target FPGA	XC7Z020-1CLG400C
Operating Frequency	110.57MHz
HW Execution Time	0.07695s
SW Execution Time	1.70821s

Ⅳ. 결론

본 논문에서는 센서 융합을 통해 분류 성능을 향상시킨 CNN 기반 HMI 시스템을 제안하였고, 이의 가속화를 위한 하드웨어 구조 설계 결과를 제시하였다. 제안된 HMI 시스템은 기존 단일 센서 기반 시스템에 비해 평균 7.7% 높은 성능을 보였으며, 가속 하드웨어 적용으로 95.6% 향상된 처리 속도를 나타냈다.

References

[1] J. Yu, Z. F. Wang, “A video, text, and speech-driven realistic 3-D virtual head for human-machine interface,” *IEEE Trans. Cybernetics,* vol.45, no.5, pp.977-988, 2015.
DOI: 10.1109/TCYB.2014.2341737

[2] Y. Zhang et al, “Static and dynamic human aram/hand gesture capturing and recognition via multiinformation fusion of flexible strain sensors,” *IEEE Sensors Journal*, vol.20, no.12, pp.6450-6459, 2020. DOI: 10.1109/jsen.2020.2965580

[3] M. Kim, J. Cho, S. Lee, Y. Jung, “IMU sensor-based hand gesture recognition for human machine interfaces,” *MDPI Sensors*, vol.19, no.18, pp.1-13, 2019. DOI: 10.3390/s19183827

[4] L. Baraldi, F. Paci, G. Serra, L. Benini, R. Cucchaiara, “Gesture Recognition Using Wearable Vision Sensors to Enhance Visitors’ Museum Experience,” *IEEE Sensors Journal*, vol.15, no.5, 2015. DOI: 10.1109/JSEN.2015.2411994

[5] S. Skaria, A. A. Hourani, M. Lech, R. J. Evans, “Hand-Gesture Recognition Using Two-Antenna Doppler Radar with Deep Convolutional Neural Networks,” *IEEE Sensors Journal*, vol.19, no.8, pp.3041-3048, 2019.
DOI: 10.1109/JSEN.2019.2892073

[6] Z. Wang, Y. Wu, Q. Niu, “Multi-sensor fusion in automated driving: a survey,” *IEEE Access*, vol.8, pp.2847-2868, 2019.
DOI: 10.1109/ACCESS.2019.2962554

[7] F. Garcia, D. Martin, A. Escalera, J. M. Armingol, “Sensor fusion methodology for vehicle detection,” *IEEE Intelligent Transportation System Magazine*, vol.9, no.1, pp.123-133, 2017.
DOI: 10.1109/MITS.2016.2620398

[8] F. Hafeez et al., “Insights and strategies for an autonomous vehicle with a sensor fusion innovation: a fictional outlook,” *IEEE Access*, vol.8, pp. 135162-135175, 2020.
DOI: 10.1109/ACCESS.2020.3010940

[9] D. Kang, D. Kum, “Camera and radar sensor fusion for robust vehicle localization via vehicle part localization,” *IEEE Access*, vol.8, pp.75223-75236, 2020. DOI: 10.1109/ACCESS.2020.2985075

[10] L. Mezai, F. Hanchouf, “Score-level fusion of face and voice using particle swarm optimization and belief functions,” *IEEE Trans. Human-machine System*, vol.45, no.6, pp.761-772, 2015.
DOI: 10.1109/THMS.2015.2438005

[11] C. Mateo, J. Talavera, “Short-time Fourier transform with the window size fixed in the frequency domain,” *Digital Signal Processing,* vol.77, pp.13-21, 2018.

[12] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vo.86, no.11, pp.2278-2324, 1998.

[13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2016, pp.770-778.

[14] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. *ICLR*, 2015.

[15] Z. Gu et al, “Blind separation of doppler human gesture signals based on continuous wave

radar sensors," *IEEE Trans. Instrumentation and Measurement,* vo.68, no.7, pp.2659-2661, Feb. 2019.
[16] E. M. Lima et al., "Analysis of the influence of the window used in the short-time Fourier transform for high impedance fault detection," In *ICHQP,* Oct. 2016, pp.350-355.
[17] P. Warden, "Speech commands: a dataset for limited-vocabulary speech recognition," *arXiv Computation and Language*, pp.1-11, Apr. 2018.

BIOGRAPHY

Seunghyun Oh (Member)

2020 : BS degree in School of Electronics and Information Engineering, Korea Aerospace University.
2020~present : MS degree course in School of Electronics and Information Engineering, Korea Aerospace University.

Chanhee Bae (Member)

2015~present : BS degree course in School of Electronics and Information Engineering, Korea Aerospace University.

Seryeong Kim (Member)

2017~present : BS degree course in School of Electronics and Information Engineering, Korea Aerospace University.

Jaechan Cho (Member)

2015 : BS degree in School of Electronics and Information Engineering, Korea Aerospace University.
2017 : MS degree in School of Electronics and Information Engineering, Korea Aerospace University.
2017~present : Ph.D degree course in School of Electronics and Information Engineering, Korea Aerospace University.

Yunho Jung (Member)

1998 : BS degree in Department of Electrical and Electronic Engineering, Yonsei University.
2000 : MS degree in Department of Electrical and Electronic Engineering, Yonsei University.
2005 : Ph.D degree in Department of Electrical and Electronic Engineering, Yonsei University.
2005~2007 : Senior Engineer, Samsung Electronics.
2007~2008 : Research professor, Institute of Information Engineering, Yonsei University.
2008~present : Professor, School of Electronics and Information Engineering, Korea Aerospace University