

멀티모달 신호처리를 위한 경량 인공지능 시스템 설계

김병수* · 이재학** · 황태호** · 김동순**

Design of Lightweight Artificial Intelligence System for Multimodal Signal Processing

Byung-Soo Kim* · Jea-Hack Lee** · Tae-Ho Hwang** · Dong-Sun Kim**

요 약

최근 인간의 뇌를 모방하여 정보를 학습하고 처리하는 뉴로모픽 기술에 대한 연구는 꾸준히 진행되고 있다. 뉴로모픽 시스템의 하드웨어 구현은 다수의 간단한 연산절차와 고도의 병렬처리 구조로 구성이 가능하며, 처리속도, 전력소비, 저 복잡도 구현 측면에서 상당한 이점을 가진다. 또한 저 전력, 소형 임베디드 시스템에 적용 가능한 뉴로모픽 기술에 대한 연구가 급증하고 있으며, 정확도 손실 없이 저 복잡도 구현을 위해서는 입력 데이터의 차원축소 기술이 필수적이다. 본 논문은 멀티모달 센서 데이터를 처리하기 위해 멀티모달 센서 시스템, 다수의 뉴론 엔진, 뉴론 엔진 컨트롤러 등으로 구성된 경량 인공지능 엔진과 특징추출기를 설계 하였으며, 이를 위한 병렬 뉴론 엔진 구조를 제안하였다. 설계한 인공지능 엔진, 특징 추출기, Micro Controller Unit(MCU)를 연동하여 제안한 경량 인공지능 엔진의 성능 검증을 진행하였다.

ABSTRACT

The neuromorphic technology has been researched for decades, which learns and processes the information by imitating the human brain. The hardware implementations of neuromorphic systems are configured with highly parallel processing structures and a number of simple computational units. It can achieve high processing speed, low power consumption, and low hardware complexity. Recently, the interests of the neuromorphic technology for low power and small embedded systems have been increasing rapidly. To implement low-complexity hardware, it is necessary to reduce input data dimension without accuracy loss. This paper proposed a low-complexity artificial intelligent engine which consists of parallel neuron engines and a feature extractor. A artificial intelligent engine has a number of neuron engines and its controller to process multimodal sensor data. We verified the performance of the proposed neuron engine including the designed artificial intelligent engines, the feature extractor, and a Micro Controller Unit(MCU).

키워드

Multimodal, Artificial Intelligence, Hardware Implementation, Feature Extraction
멀티 모달, 인공지능, 하드웨어 구현, 특징 추출

* 교신저자 : 전자부품연구원 SoC플랫폼연구센터

• Received : Aug. 03, 2018, Revised : Sep. 08, 2018, Accepted : Oct. 15, 2018

** 전자부품연구원(jhk507@keti.re.kr, taео@keti.re.kr, dskim@keti.re.kr)

• Corresponding Author : Byung-Soo Kim

SoC Platform Research Center, Korea Electronics Technology Institute

Email : bskim4k@keti.re.kr

• 접수일 : 2018. 08. 03

• 수정완료일 : 2018. 09. 08

• 게재확정일 : 2018. 10. 15

I. 서 론

최근 인간의 뇌를 모방하여 정보를 학습하고 처리하는 뉴로모픽 기술이 꾸준히 연구되고 있으며, 뉴로모픽 시스템에 대한 하드웨어 구현은 지난 몇 년간 급속한 발전을 이루었다[1-2]. 일반적으로 순차적으로 데이터를 처리하는 기존의 폰-노이만 구조에서는 시스템 클럭 주파수를 높이고, 연산 알고리즘을 최적화하고, 멀티-코어 구조를 도입하여 연산의 한계를 극복하려고 노력하였지만 전력 소비와 발열량이 급속히 증가하는 문제를 가지고 있다[3-4]. 이와 달리, 뉴로모픽 시스템은 수백, 수천 개의 뉴런이 병렬로 연결되어 있으며, 모든 뉴런이 동시에 입력을 받아 처리할 수 있는 구조로 구성이 가능하여 처리속도, 전력소비, 저복잡도 구현 측면에서 상당한 이점을 갖는다[5-6]. 그리하여 범용 프로세서의 컴퓨팅 성능이 엄청나게 증가했음에도 불구하고, 병렬구조의 뉴로모픽 하드웨어를 구현하는 것이 실시간 제어, 영상처리, 패턴 인식 등과 같은 실시간 처리, 데이터 집약적인 응용 프로그램을 처리함에 있어 유리함을 가지게 된다. 이렇게 초창기 뉴로모픽 기술에 대한 연구는 병렬성을 이용한 계산 속도의 향상 및 실시간 성능에 집중되었지만 최근에는 저전력, 소형 임베디드 시스템에 적용 가능한 뉴로모픽 기술에 대한 연구도 급증하고 있다[7].

인공 신경망에 대한 입력 데이터의 표현은 인공 신경망의 성능에 상당한 영향을 줄 수 있는 중요한 요소이다. 입력데이터가 클래스에 따라 중요한 차이를 나타내도록 표현한다면, 이를 분류하는 인공 신경망을 보다 쉽게 설계할 수 있으며 좋은 성능을 나타낼 수 있으므로 입력데이터를 변환하는 특징 추출은 인공 신경망의 핵심적인 문제이다. 또한 고차원의 데이터를 사용하는 경우 인공신경망의 크기, 학습 시간이 증가하는 문제가 발생하므로 작은 인공 신경망으로 높은 정확도를 얻을 수 있도록 입력 데이터의 차원 축소가 필수적이다[8-9].

본 논문에서는 멀티모달 센서 데이터를 처리할 수 있는 센서 시스템, 다양한 어플리케이션을 처리할 수 있는 경량 인공지능 엔진과 이를 위한 특징추출기를 제안하고, 제안된 하드웨어의 성능을 소프트웨어와 비교 검증하였다. 제안하는 경량 인공지능 엔진은 병렬 처리 특성으로, 입력 데이터 셋에 관계없이 일정 시간

내에 멀티모달 신호처리(인식, 학습, 등)를 수행하며, 그 성능을 MCU와 FPGA를 연동하여 검증하였다.

II. 경량 인공지능 시스템

본 논문에서 제안하는 멀티모달 신호처리를 위한 경량 인공지능 시스템은 멀티모달 센서 시스템과 경량 인공지능 엔진으로 구성되어 있으며, 하드웨어 구조는 그림 1과 같다. 멀티모달 센서 시스템은 다양한 센서 신호를 입력받아 경량 인공지능 엔진 또는 특징 추출기로 보정된 데이터를 전달한다. 이 센서 데이터는 특징추출기에서 변환되고, 뉴런 엔진, 뉴런 엔진 발화 뉴런 검출기로 구성된 인공지능 엔진은 입력 데이터에 대한 학습, 인지 과정을 처리한다.

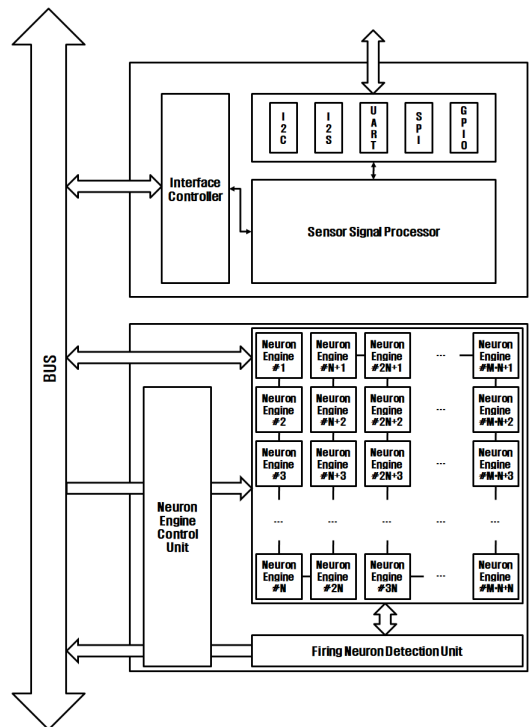


그림 1. 경량 인공지능 엔진 하드웨어 구조
Fig. 1 Lightweight ai engine hardware architecture

경량 인공지능 엔진은 MCU와의 통신을 위한 인터페이스, 128개의 뉴런 엔진을 제어하기 위한 뉴런 엔진 컨트롤러와 병렬 뉴런 엔진으로 구성되어 있다.

MCU에서 버스 인터페이스를 통해 병렬 뉴런 엔진을 제어하기 위한 레지스터 셋, Feature 데이터를 저장하기 위한 레지스터 셋을 설정하면, 병렬 뉴런 엔진 및 컨트롤러에서 설정된 레지스터 값과 입력 Feature 데이터를 읽어 들어 멀티모달 신호에 대하여 학습 및 인식 과정을 수행한다. 해당 동작이 완료되면 Done 신호를 MCU에 전송하여 제어 레지스터를 리셋하고 다음 동작을 수행하기 위한 대기모드로 이동한다.

2.1 멀티모달 센서 시스템

다양한 센서를 지원하고 인공지능 엔진의 확장성을 위해 제안된 경량 인공지능 시스템에서는 그림 2와 같이 I2C, I2S, UART, SPI, GPIO 등의 다양한 통신 인터페이스를 지원하는 멀티모달 센서 시스템을 설계하였다. 멀티모달 센서 시스템으로 입력된 다양한 센서 데이터는 멀티모달 센서 시스템 내에서 1차적으로 보정되며, 이 보정된 센서 데이터는 특징 추출기 또는 뉴런 엔진의 입력으로 사용된다.

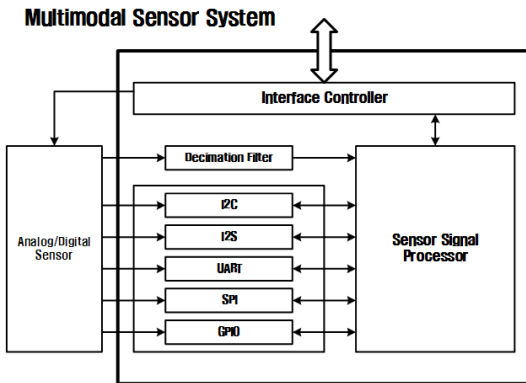


그림 2. 멀티모달 센서 시스템 블록다이어그램
Fig. 2 Multimodal sensor system block diagram

2.2 뉴런 엔진 컨트롤러

뉴런 엔진 컨트롤러의 상태 다이어그램은 그림 3과 같다. 뉴런 엔진 컨트롤러의 상태는 컨트롤 레지스터의 값과 네트워크의 상태를 나타내는 플래그 레지스터 값에 따라서 변화한다. 상태 머신은 10가지의 상태로 구성되어 있으나 동작에 따라 분류하면 크게 5가지 상태로 이루어졌다. 그 상태는 입력 데이터를 뉴런 엔진에 전송하는 상태, 뉴런 엔진이 입력 데이터를 학

습하는 상태, 입력된 데이터를 분류하는 상태, 학습되어진 네트워크 정보를 뉴런 엔진에 전송하는 업데이트 상태, 뉴런 엔진의 정보를 외부로 전송하는 상태이다. 각 상태별로 뉴런 엔진의 연산이 이루어지며, 연산이 종료 하게 되면 IDLE 상태로 돌아오며 Done 신호를 발생하는 구조이다.

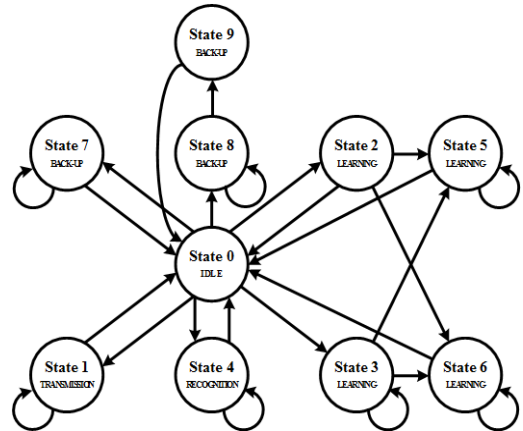


그림 3. 뉴런 엔진 컨트롤러의 상태 다이어그램
Fig. 3 State diagram of the neuron engine controller

2.3 뉴런 엔진

제안하는 뉴런 엔진의 하드웨어는 입력 Feature 데이터와 거리를 계산하기 위한 연산기, Threshold, 입력 데이터 셋의 클래스, 타입 등을 저장하는 레지스터, 현재 뉴런의 Feature 데이터를 저장하고 있는 메모리와 뉴런의 상태 및 동작을 제어하는 컨트롤 유닛으로 구성되어있으며, 그림 4는 제안하는 뉴런 엔진의 구조를 나타내고 있다.

각각의 뉴런 엔진은 하나의 체인으로 연결되어 컨트롤러와 가장 근접한 뉴런 엔진부터 순차적으로 학습되며, 각 뉴런 엔진별로 컨트롤러를 통해 다음 뉴런의 활성화 가능 여부를 알려준다. 모든 뉴런 엔진의 출력 값은 발화 뉴런 검출기를 통해 뉴런 엔진과 뉴런 컨트롤러간의 통신이 이루어지며 학습 및 인식 동작을 수행한다. 이 때 출력은 뉴런 컨트롤러 또는 출력 레지스터로 전송되어 MCU에서 그 값을 사용하게 된다. 뉴런 컨트롤러에서는 입력 Feature 데이터를 뉴런 엔진에 전송하는 Feature 전송, 뉴런 엔진의 학습,

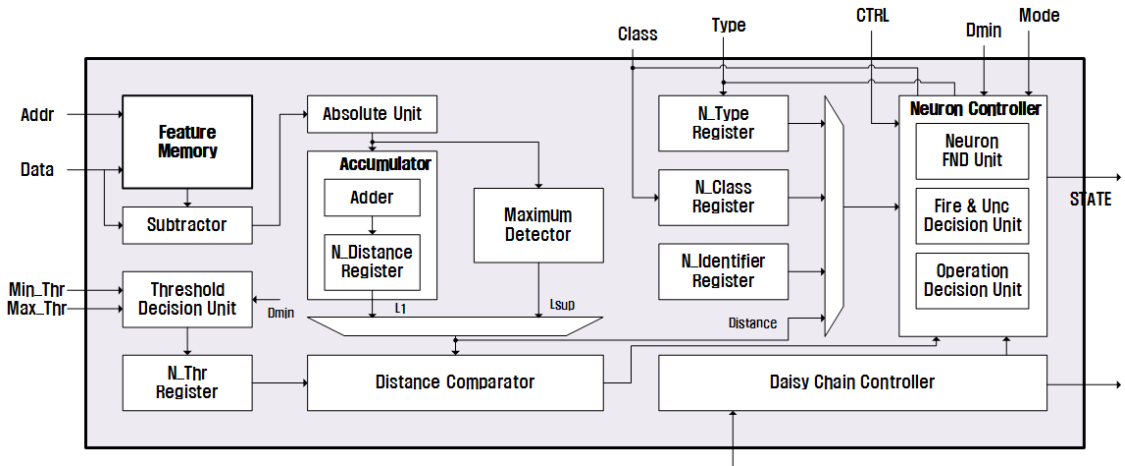


그림 4. 뉴런 엔진의 하드웨어 구조
Fig. 4 Neuron engine hardware architecture

입력 Feature 데이터를 분류하는 인식, 미리 학습된 정보를 뉴런 엔진으로 전송하는 업데이트, 뉴런 엔진의 정보를 외부로 전송하는 백업 등 총 5가지의 상태에 따라 뉴런 엔진을 제어한다. 각 상태별로 동작이 완료되면 초기 상태로 이동하며 Done 신호를 생성하며, 인식 동작이 완료되는 경우에는 클래스, 거리 정보 등을 출력한다.

III. 실험결과

제안하는 경량 인공지능 엔진의 검증을 위해 다양한 테스트 셋에 대한 검증을 진행하였다. MATLAB 시뮬레이터로부터 테스트 벡터 파일을 추출하고, 사용하는 뉴런 엔진의 수를 최소화 할 수 있는 최적의 Threshold값을 결정하였다. Threshold값 결정을 위해 일반적인 머신러닝에서 사용되는 데이터 셋을 사용하여 학습을 진행하였으며, 그림 5에 데이터 셋에 따른 정확도와 발화되는 뉴런의 개수가 나타나 있다[10]. 제안된 하드웨어 구조는 Verilog-HDL을 이용하여 설계하였으며, Xilinx사의 Vivado Too를 사용하며 세부 연산 모듈을 디자인하고 회로를 합성하였다. 타겟 디바이스는 Virtex-7 (xc7v2000t)로 하였으며 RTL 시뮬레이션을 통해 출력 벡터 파일을 추출, 비교하여 로직 검증을 수행하였다. 표 1은 128개의 뉴런 엔진으

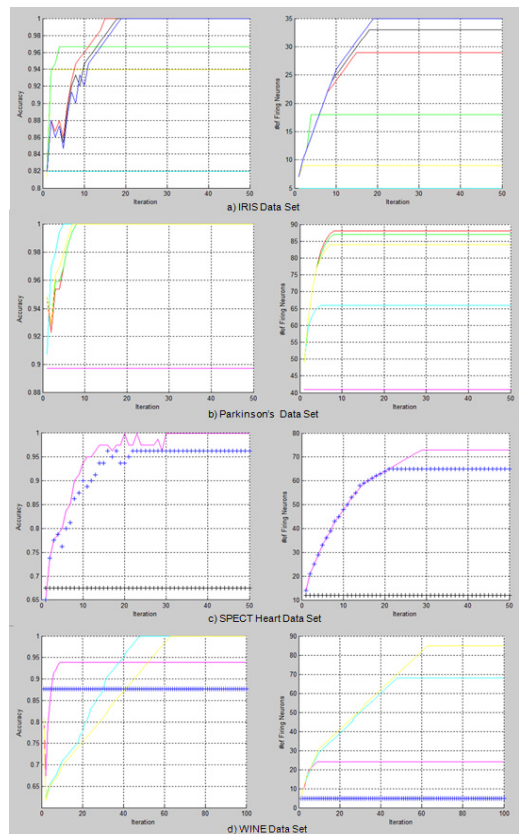


그림 5. 뉴런 엔진의 매트랩 시뮬레이션 결과
Fig. 5 Neuron engine MATLAB simulation result

로 구성된 인공지능 엔진을 하드웨어로 구현한 결과이며, 그림 6은 경량 인공지능 시스템을 검증하기 위해 제작된 FPGA 플랫폼 보드이다.

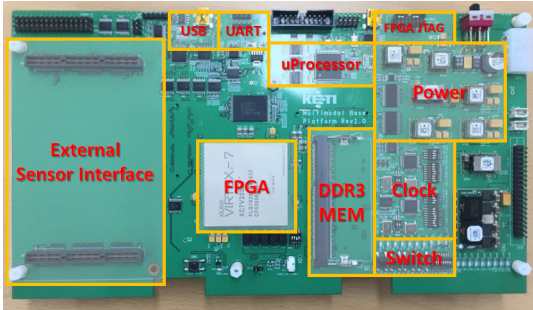


그림 6. 경량 인공지능 시스템 FPGA 검증 플랫폼
Fig. 6 FPGA verification platform of the lightweight artificial intelligence system

표 1. 경량 인공지능 엔진 FPGA Utilization
Table 1. FPGA utilization of the lightweight artificial intelligence engine

Implementation Specification	Description
Target Device	Xilinx Virtex-7(xc7v2000t)
Number of FFs	18,565 / 2,443,200(0.76%)
Number of LUTs	31,522 / 1,221,600(2.58%)
Number of BRAMs	64 / 1292(4.95%)

IV. 결 론

본 논문은 멀티모달 센서 데이터를 처리하기 위한 경량 인공지능 시스템 설계에 대한 연구이다. 제안한 경량 인공지능 시스템은 멀티모달 센서 데이터를 처리하기 위한 멀티모달 센서 시스템과 입력 데이터 셋을 가공할 수 있는 특징추출기, 데이터를 학습, 인지할 수 있는 인공지능 병렬 뉴런 엔진으로 구성되어 있다.

제안된 멀티모달 센서 시스템의 설계를 통해 다양한 데이터를 입력받을 수 있으며, 센서 고유의 특성에 의한 출력과 측정값의 차이를 보정하여 보다 정확한 데이터의 취득이 가능하며, 특징추출기를 통해 입력된 데이터의 가공이 가능하여 인공지능 엔진에 적합한

데이터를 제공할 수 있다. 또한 복잡한 연산이 아닌 단순 거리 계산 기반의 뉴런 엔진의 병렬화를 통해 경량 인공지능 엔진을 구현하였으며, 제안하는 경량 인공지능 시스템은 Threshold 조절을 통해 다양한 어플리케이션에 대한 학습 및 인식이 가능함을 검증하였다.

감사의 글

위 논문은 “2018년 한국전자통신학회 봄철학술대회 우수논문”입니다. 본 논문은 2018년도 산업통산자원부 및 산업기술평가관리원 연구비 지원에 의한 연구임(‘10073122, 오픈플랫폼을 고려한 지능형 멀티모달 처리 SoC 모듈 서비스 개발’).

References

- [1] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, “A million spiking-neuron integrated circuit with a scalable communication network and interface,” *Science*, vol. 345, no. 6197, 2014, pp. 668-673.
- [2] B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J. M. Bussar, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen, “Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations,” *Proc. of the IEEE*, vol. 102, no. 5, 2014, pp. 699-716.
- [3] K. Kim, “Optimal Structures of a Neural Network Based on OpenCV for a Golf Ball Recognition,” *J. of the Korea Institute of Electronic Communication Sciences*, vol. 10, no. 2, Feb. 2015, pp. 267-273.
- [4] Y. Lee and P. Moon, “A Comparison and Analysis of Deep Learning Framework,” *J. of the Korea Institute of Electronic Communication*

- Sciences*, vol. 12, no. 1, Feb. 2017, pp. 115-122.
- [5] C. Schuman, T. Potok, R. Patton, J. Birdwell, M. Dean, G. Rose, and J. Slank, "A survey of neuromorphic computing and neural networks in hardware," *arXiv preprint arXiv:1705.06963*, 2017, pp. 1-8.
- [6] S. Soman, Jayadeva, and M. Suri, "Recent trends in neuromorphic engineering," *Big Data Analytics*, vol. 1, no. 15, Dec. 2016, pp. 1-19.
- [7] M. Suri, V. Parmar, A. Singla, R. Malviya, and S. Nair, "Neuromorphic hardware accelerated adaptive authentication system," *Symp. Series on Computational Intelligence*, Dec. 2015, pp. 1206-1213.
- [8] T. Baltrušaitis, C. Ahuja, and L. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," *arXiv preprint arXiv:1705.09406*, 2017, pp. 1-20.
- [9] A. Benediktsson and J. Sveinsson. "Feature extraction for neural network classifiers," *n Neurocomputation in Remote Sensing Data Analysis*, 1997. pp. 97-104.
- [10] A. Asuncion and D. Newman, "UCI Machine Learning Repository," 2007.

저자 소개



김병수(Byung-Soo Kim)

2008년 인하대학교 정보통신공학과 졸업(공학석사)
2013년 인하대학교 정보통신공학과 졸업(공학박사)

2013년~현재 전자부품연구원 SoC플랫폼연구센터 선임연구원

※ 관심분야 : 인공지능 시스템 설계, SoC 설계



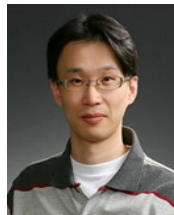
이재학(Jea-Hack Lee)

2011년 아주대학교 대학원 전자공학과 졸업(공학석사)

2017년 아주대학교 대학원 전자공학과 졸업(공학박사)

2017년~현재 전자부품연구원 SoC플랫폼연구센터 선임연구원

※ 관심분야 : SoC 설계, 유무선 통신시스템



황태호(Tae-Ho Hwang)

2000년 한국외국어대학교 컴퓨터공학과 졸업(공학석사)

2013년 한국외국어대학교 컴퓨터공학과 졸업(공학박사)

2000년~현재 전자부품연구원 SoC플랫폼연구센터 수석연구원

※ 관심분야 : 실시간 운영체제, 뉴로모픽 컴퓨팅, 이기종 컴퓨팅



김동순(Dong-Sun Kim)

1999년 인하대학교 전자재료공학과 졸업(공학석사)

2005년 인하대학교 미디어시스템학과 졸업(공학박사)

2011년~2014년 인하대학교 정보통신공학과 겸임교수

1999년~현재 전자부품연구원 SoC플랫폼연구센터 센터장

※ 관심분야 : 임베디드 하드웨어, 멀티미디어 SoC 설계