

특집논문 (Special Paper)

방송공학회논문지 제24권 제4호, 2019년 7월 (JBE Vol. 24, No. 4, July 2019)

<https://doi.org/10.5909/JBE.2019.24.4.573>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

ADAS용 다중화각 카메라를 이용한 객체 인식 향상

박 동 훈^{a)}, 김 학 일^{a)†}

Improved Object Recognition using Multi-view Camera for ADAS

Dong-hun Park^{a)} and Hakil Kim^{a)†}

요 약

완전한 자율 주행에 이르기 위해서는 주변 환경을 인지하는 인지 능력이 사람보다 뛰어나야 한다. 자율 주행에서 주로 사용되는 60° 협각, 120° 광각 카메라는 시야각에 따른 각각의 단점이 존재한다. 본 논문의 목적은 광각, 협각 카메라가 가진 각각의 단점을 극복하기 위하여, 다중화각 차량 전방 카메라 시스템을 이용하여 더 넓은 영역의 전방을 대상으로 더 정확히 객체를 인식할 수 있는 심층신경망 알고리즘을 개발하는 것이다. 광각, 협각 카메라로 취득된 데이터의 종횡비를 분석해 SSD(Single Shot Detector) 알고리즘을 수정하였고, 취득된 데이터를 학습하여 단안 카메라만을 사용할 때 보다 높은 성능을 달성하였다.

Abstract

To achieve fully autonomous driving, the perceptual skills of the surrounding environment must be superior to those of humans. The 60° angle, 120° wide angle cameras, which are used primarily in autonomous driving, have their disadvantages depending on the viewing angle. This paper uses a multi-angle object recognition system to overcome each of the disadvantages of wide and narrow-angle cameras. Also, the aspect ratio of data acquired with wide and narrow-angle cameras was analyzed to modify the SSD(Single Shot Detector) algorithm, and the acquired data was learned to achieve higher performance than when using only monocular cameras.

Keyword : Object Detection, Multi-Angle Camera, Deep learning, Light Weight, Vehicle camera system

a) 인하대학교 융합학과 미래차동차공학부(Future Vehicle Engineering, Inha University)

† Corresponding Author : 김학일 (Hakil Kim)

E-mail: hikim@inha.ac.kr

Tel: +82-32-860-7385

ORCID: <https://orcid.org/0000-0003-4243-3804>

※ 이 논문의 연구결과 중 일부는 “IPIU 2019”에서 발표한 바 있음.

· Manuscript received May 7, 2019; Revised July 8, 2019; Accepted July 8, 2019.

I. 서 론

완전한 자율 주행에 이르기 위해서는 주변 환경을 인지하는 인지 능력이 사람보다 뛰어나야 한다. 자율 주행에서 주로 사용되는 60° 협각, 120° 광각 카메라는 시야각에 따른 각각의 단점이 존재한다. 본 논문의 목적은 광각, 협각

카메라가 가진 각각의 단점을 극복하기 위하여, 다중화각 차량 전방 카메라 시스템을 이용하여 더 넓은 영역의 전방을 대상으로 더 정확히 객체를 인식할 수 있는 심층신경망 알고리즘을 개발하는 것이다. 광각, 협각 카메라로 취득된 데이터의 종횡비를 분석해 SSD(Single Shot Detector) 알고리즘을 수정하였고, 취득된 데이터를 학습하여 단안 카메라만을 사용할 때 보다 높은 성능을 달성하였다.

II. 개발 동향

운전자가 주행 도중 전방에 나타난 물체의 정체를 판단하는 일은 매우 중요하며, 주행 중 일어날 수 있는 다양한 환경을 빠르고 정확하게 인식할 수 있어야 한다. 최근, 딥러닝 기술을 자율 주행에 도입해 차량 인지 능력을 향상시키려는 시도가 늘고 있다. 데이터를 입력하고 학습하면 정보를 스스로 인지할 수 있기 때문이다. 이러한 딥러닝 기술을 차량에 적용하면 주차보조, 사각지대 감지, 충돌방지, 교통 표지판 인식^[1], 전방 객체 인식^{[2][3]} 등에 활용이 가능하다.

개발한 알고리즘이 실제 자율주행 차량이 사용할 수 있으려면 경량화 된 실시간 처리속도의 모델이 임베디드 환경에서 동작할 수 있어야 한다. 임베디드 환경에서 최적화하여 동작할 수 있는 딥러닝 모델 연구는 여전히 활발히 연구 중이다^[4].

카메라와 딥러닝 기술을 이용해 차량 인지 능력을 향상시키는 방법에는 단안(Mono)카메라 방식, 듀얼(Dual) 카메라 방식이 존재한다. 단안 카메라 방식은 듀얼 카메라에 비해 하드웨어 가격이 싸며 알고리즘이 단순한 장점이 존재하지만, 사고가 날 경우 대체 가능한 카메라가 없으며 시야각(FOV)이 고정된 단점이 있다. 위 방식의 단점을 극복하기 위해 본 논문에서는 다중화각 카메라를 이용한 전방 객체 인식 시스템을 소개한다. 또한, 각 카메라의 장점을 이용해 객체 인식 성능을 향상시킬 수 있는 방법을 제안한다. 본 논문에서 사용한 시스템 및 알고리즘은 앞서 소개한 실제 임베디드 환경에서 사용이 가능하도록 임베디드 보드인 NVIDIA Tx2를 플랫폼 기반으로 개발하였다.

III. 다중화각 카메라를 이용한 차량 전방 객체 인식 방법

1. 협각, 광각을 이용한 차량 전방 객체 인식

협각과 광각 카메라는 각각의 장단점이 존재한다. 협각 카메라는 가시거리는 길지만, 시야각 좁은 단점이 있고, 광각 카메라는 시야각은 넓지만 취득된 이미지에 존재하는 객체의 스케일이 작고 영상 왜곡이 존재한다는 단점이 있다.

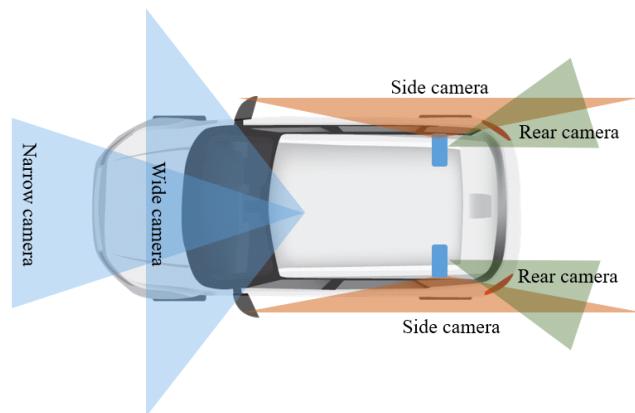


그림 1. 듀얼 카메리를 부착한 자율 주행 시스템
Fig. 1. Automotive driving system with multi-view camera

본 논문에서 사용한 시스템은 두 카메라의 장점을 모두 활용하고자 하였다. 그림1은 제안하는 다중화각 객체 검출 시스템을 나타낸다. 본 시스템은 속도가 저속(30km/h 미만) 일 때는 넓은 시야각으로 주변을 인지하기 위해서 광각 카메라 만을 이용하였으며, 고속(30km/h 이상) 일 때는 멀리 있는 물체를 인지하기 위하여 협각 카메라를 이용하여 전방 객체를 인지하도록 하였다.

2. 객체 인식을 위한CNN 알고리즘 선정

자율주행을 위한 객체 검출기는 실시간성을 보장할 수 있어야 하며, 저 메모리 환경에서 동작할 수 있어야 한다^[4]. 제안하는 방식의 도메인인 임베디드 보드 NVIDIA Tx2에서 동작하기 위해서 경량화 된 모델인 SSD InceptionV3를 사용하였다. SSD는 격자 기반의 알고리즘을 사용하여 겹

치는 영역에 대한 중복연산을 최소화하여 연산량을 크게 줄인 검출 알고리즘이다^[5].

기본 네트워크의 경우 사이즈와 파라미터 수가 더 적은 모델인 MobileNetV2를 사용하려고 하였으나^{[7][8]}, 각각의 모델을 Tx2상에서 테스트한 결과 FPS(Frame per second) 차이가 크지 않았다. 따라서 mAP(mean Average Precision) 가 상대적으로 더 높은 InceptionV3를 기본 네트워크로 사용하였다^[10]. 표1은 임베디드 보드에 적합한 경량화 모델을 비교한 것이다.^[6]

표 1. 경량화 모델의 크기, 파라미터 수, 성능 비교

Table 1. Size, number of parameters and performance comparison of lightweight CNN models

	MobileNetV2	InceptionV3
Size	14MB	45MB
Parameters	3,538,984	23,851,784
Top-1 Accuracy on ImageNet(mAP)	71.3%	77.9%
FPS	약 21FPS	약 20FPS

3. 객체 인식 성능 향상을 위한 협각, 광각 영상 정보 분석

SSD 알고리즘은 컨볼루션 필터에 의해 K개의 박스 정보를 출력하게 되는데, 미리 정해진 비율로 생성된 디폴트 박스 중 더 가까운 쪽으로 매핑이 된다^[5]. 이때 사전에 정하는 디

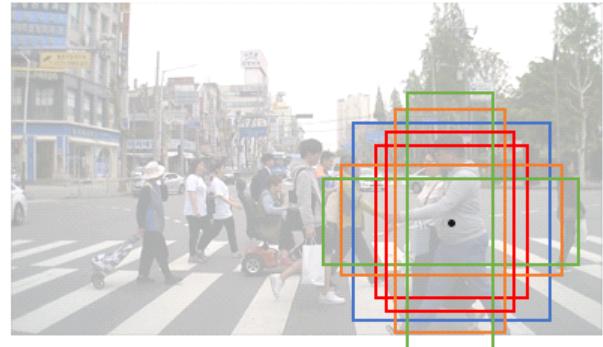


그림 2. 협각 카메라로 취득한 샘플 이미지

Fig. 2. Sample images acquired with a narrow camera

풀트(Default) 박스를 도메인 환경에 맞춰 변경할 수 있다.

그림 3은 협각 카메라로부터 취득된 국내 데이터의 종횡비 분포도이다. 히스토그램으로부터 커널 밀도 함수를 추정한 결과 자동차의 경우 1, 1.5, 보행자의 경우 1/2, 1/3, 자전거 행인의 경우 0.7, 1, 1/2에 분포가 높은 것을 확인할 수 있다.

그림 4는 광각 카메라로부터 취득된 국내 데이터의 종횡비 분포도이다. 그림3, 4의 각각 종횡비 분포를 확인할 때, 협각, 광각 데이터의 종횡비가 매우 유사함 확인할 수 있다.

국내 도로상 객체 크기 분포를 기반으로 SSD의 디폴트 박스 종횡비는 [1/3, 1/2, 1, 1.5, 2, 3]로 설정할 수 있다. 그림 2는 입의의 피처맵에서 그려지는 디폴트 박스를 영상에 투영한 그림이다. 입력 영상의 크기는 실시간성을 만족시키기 위해 640 360 크기로 변환하였다. 최소 검출 가능

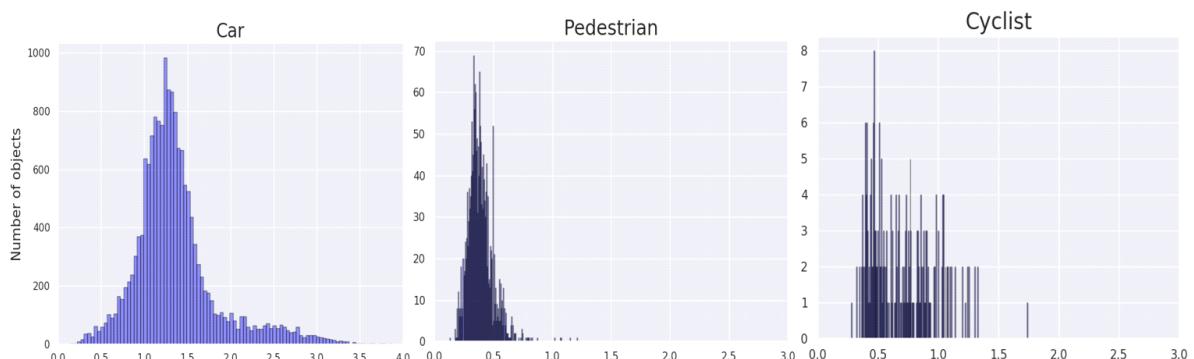


그림 3. 협각 카메라로부터 취득된 국내 도로 데이터 객체 별 종횡비(Width/Height)

Fig. 3. Aspect ratios of domestic road data acquired from narrow-angle camera (Width/Height)

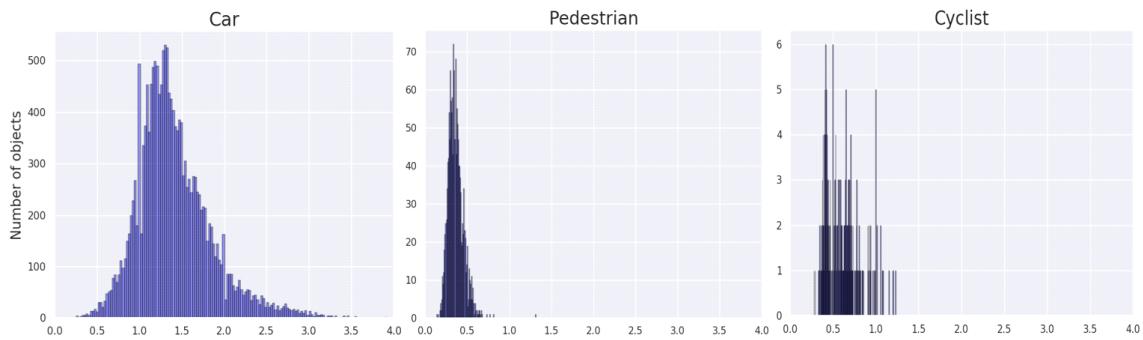


그림 4. 광각 카메라로부터 취득된 국내 도로 데이터 객체 별 종횡비(Width/Height)
Fig. 4. Aspect ratios of domestic road data acquired from wide-angle camera (Width/Height)

객체의 픽셀 크기는 차량의 크기를 2m로 가정하고, 최대 60m 거리에서 검출이 가능하도록 하기 위해 핀홀(Pinhole) 모델을 가정하여 수식 1을 통해 정하였다.

$$h_{\min} = \frac{H}{Z} f_{\text{Camera}} \quad (1)$$

h_{\min} : 이미지 상에서 객체 최소 검출 가능 높이(px)

H : 실제 차량의 크기, 2m.

Z : 카메라 초점에서부터 객체까지의 거리, 60m.

f_{camera} : 카메라의 초점거리, 635px

정해진 최소 검출 가능 크기를 고려하여 SSD의 스케일 요소를 [21, 36.75, 64.31, 112.55, 196.96, 360]으로 설정하였다^[5]. 각 스케일 요소는 수식2에 근거하여 SSD 알고리즘 각 특징맵(Feature Map) 셀마다 존재하는 Anchor Box의 Width, Height를 결정하게 된다.

$$\begin{aligned} w_n^l &= scale^l * \sqrt{\text{aspect ratio}^l} \\ h_n^l &= scale^l / \sqrt{\text{aspect ratio}^l} \end{aligned} \quad (2)$$

l : l 번째 레이어, $l = 1, 2, 3, 4, 5, 6$

w_n^l : l 번째 계층의 n 번째 디폴트 박스의 너비

h_n^l : l 번째 계층의 n 번째 디폴트 박스의 높이

s^l : l 번째 계층의 스케일 요소

a_n^l : l 번째 계층의 n 번째 디폴트 박스의 종횡비

IV. 실험 결과 및 분석

1. 실험 환경 구축

각 카메라로부터 취득한 데이터는 연속된 영상으로부터 취득한 것이기에 높은 상관관계를 가지고 있다. 학습 데이터와 평가 데이터를 분할할 때에는 각 데이터의 상관관계를 최소화하기 위하여, 무작위로 추출하였다^[11].

협각 카메라로부터 취득한 데이터는 총 3,000장이며 트레이닝 데이터 개수는 2,400장, 테스트 데이터 개수는 600장으로 분할하였다. 동 시간대에 취득한 광각 카메라로 취득한 데이터도 동일하게 분할하였다. 각 카테고리마다 제공되는 인스턴스는 아래 표 2에 표시하였다.

표 2. 차량 전방 카메라로부터 취득한 객체 수

Table 2. Number of objects acquired from vehicle front camera

		Car	Pedestrian	Cyclist
Wide angle 120°	Train data	17,446	1,536	230
	Test data	1,893	62	9
Narrow angle 60°	Train data	16,881	1,979	247
	Test data	1,629	107	15

2. 비교 실험 결과

본 실험에서는 협각, 광각 카메라 모두에서 Scale- Invariant 한 검출 성능을 내고자 하였고, 영상 왜곡에도 강인한 검출 성능을 얻기 위해 협각, 광각 카메라로부터 취득한 데이터

를 모두 학습하였다. 실험의 공정성을 확보하기 위해 두 데이터를 모두 학습한 경우는 6000장 중 랜덤하게 3000장을 추출하여 학습을 진행하였다. 아래는 비교 실험 결과를 기술한 것이다.

그림 5는 협각 카메라로부터 얻어진 데이터 만을 학습했을 경우 얻어진 객체 검출 샘플 이미지이다. 작은 물체를 검출하지 못하거나, 그림과 같이 오 검출하는 경우가 생겼다. 반면, 그림 6은 협각, 광각 카메라로부터 얻어진 데이터를 모두 학습했을 경우 얻어진 객체 검출 샘플 이미지이다. 오 검출이 크게 줄었을 뿐만 아니라, 작은 물체도 검출하는 모습을 확인할 수 있다.

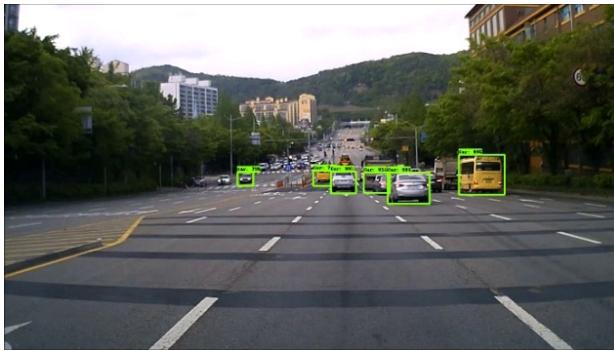


그림 5. 협각 카메라 데이터 학습 후 검출 영상
 Fig. 5. Detection image after learning the data from narrow-angle

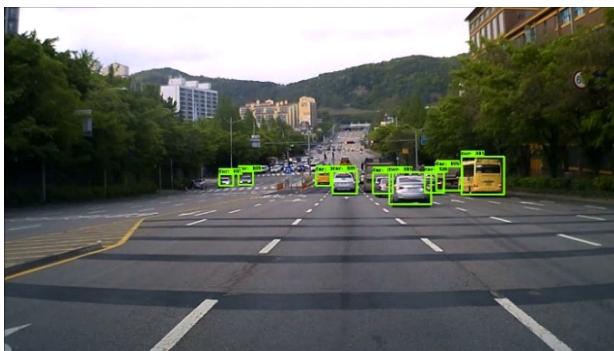


그림 6. 협각, 광각 카메라 데이터 학습 후 검출
 Fig. 6. Detection image after learning the data from wide, narrow-angle

그림 7은 광각 카메라로부터 얻어진 데이터만을 학습했을 경우 얻어진 객체 검출 샘플 이미지이다. 이미지상의 작은 물체에 대해서 검출하지 못하는 결과가 나타나지만 협각, 광각 데이터를 모두 학습한 그림 8은 작은 물체 또한

검출이 용이한 것을 확인할 수 있다.



그림 7. 광각 카메라 데이터 학습 후 검출 영상
 Fig. 7. Detection image after learning the data from wide-angle

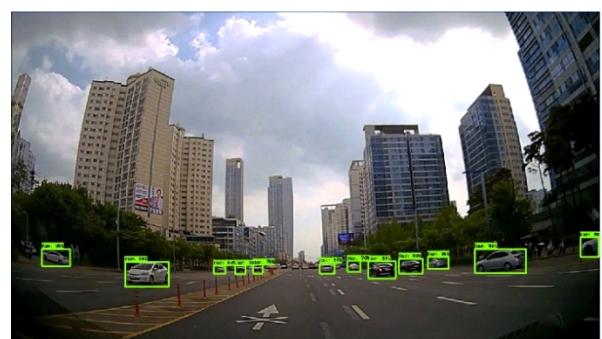


그림 8. 협각, 광각 카메라 데이터 학습 후 검출 영상
 Fig. 8. Detection image after learning the data from wide, narrow-angle

결과적으로 두 데이터를 모두 학습한 네트워크는 협각

표 3. 제안된 방식의 성능분석

Table 3. Performance analysis of proposed method

Train	Test	Narrow-angle camera data	Wide-angle camera data
Narrow angle camera data	mAP	59.62%	38.08%
	AP_{car}	84%	70%
	AP_{ped}	69%	34%
	AP_{cyc}	26%	10%
Wide angle camera data	mAP	41.85%	43.75%
	AP_{car}	78%	83%
	AP_{ped}	28%	38%
	AP_{cyc}	20%	10%
Wide & Narrow angle camera data	mAP	61.70%	47.67%
	AP_{car}	86%	85%
	AP_{ped}	69%	48%
	AP_{cyc}	30%	10%

데이터, 광각 데이터 모두 강인한 결과를 얻었다. 아래의 표 3는 정량적으로 평가한 검출 성능 지표이다.

V. 결 론

본 논문은 다중화각 카메라를 이용한 자율주행 시스템의 객체 인식 성능을 높이고자 국내 도로에서 취득한 데이터를 분석하고 Domain Specific한 정보를 최대한 활용하는 방법을 제안하였다.

주요 검출 객체인 Car, Pedestrian, Cyclist의 종횡비를 기반으로 SSD Default Box의 Scale을 결정하였으며, 협각, 광각 카메라의 데이터를 모두 학습하여 작은 물체의 검출 성능을 높이면서 영상 왜곡에도 강인한 성능을 얻을 수 있었다.

제안한 방식은 협각 카메라 취득한 데이터에서 평가했을 경우 mAP=61.70%, 광각 카메라에서 취득한 데이터에서 평가했을 경우 mAP=47.67%를 얻어, 하나의 카메라에서 취득한 데이터를 학습한 경우보다 높은 성능을 얻었다.

참 고 문 헌 (References)

- [1] Sebastian Houben, Johannes Stallkamp. "Detection of traffic signs in real-world images: The German traffic sign detection benchmark" The 2013 International Joint Conference on Neural Networks, IJCNN.
- [2] M. Cordts, M. Omran, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, 2016
- [3] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite" Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.3354-3361, 2012.
- [4] Son, K. Choi, N. Song, and D Lee, "Real-Time Dynamic Simulation of Vehicle and Occupant Using a Neural Network" Transactions of KSSE, Vol.10, No2, pp.132-140,2002.
- [5] Wei Liu, Dragomir Anguelov, "SSD: Single Shot MultiBox Detector" In ECCV 2016.
- [6] Iljoo Baek, Albert Davies, "Real-time Detection, Tracking, and Classification of Moving and Stationary Objects using Multiple Fisheye Images" arXiv:1803.06077.
- [7] Dumitru Erhan, Christian Szegedy, "Scalable Object Detection using Deep Neural Networks" In CVPR 2014.
- [8] Andrew G Howard, Menglong Zhu, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications" In CVPR 2017.
- [9] Mark Sandler, Andrew Howard, "MobileNetV2: Inverted Residuals and Linear Bottlenecks", In CVPR 2018.
- [10] Christian Szegedy, Vincent Vanhoucke, "Rethinking the Inception Architecture for Computer Vision" In CVPR 2015.
- [11] Xiaozhi Chen, Kaustav Kundu, "3D Object Proposals for Accurate Object Class Detection" In NIPS 2015

저 자 소 개

박 동 훈

- 2016년 : 인하대학교 전자공학과
- 2018년 ~ 현재 : 인하대학교 대학원 미래자동차공학과 석사과정
- ORCID : <https://orcid.org/0000-0001-9021-2784>
- 주관심분야 : 컴퓨터 비전, 딥러닝, 센서퓨전



저자소개



김 학 일

- 1983년 : 서울대학교 제어계측공학과(공학사)
- 1985년 : Purdue 대학교 전기/컴퓨터공학과(공학석사)
- 1990년 : Purdue 대학교 전기/컴퓨터공학과(공학박사)
- 1990년 ~ 현재 : 인하대학교 정보통신공학과 교수
- ORCID : <https://orcid.org/0000-0003-4243-3804>
- 주관심분야 : 패턴인식, 컴퓨터비전, 바이오인식, 로봇비전, 의료영상처리