

CNN기반의 온라인 수어통역 상담 시스템에 관한 연구

박원철¹, 박구락^{2*}

¹공주대학교 컴퓨터공학과 박사, ²공주대학교 컴퓨터공학부 교수

CNN-based Online Sign Language Translation Counseling System

Won-Cheol Park¹, Koo-Rack Park^{2*}

¹Ph.D, Division of Computer Engineering, Kongju National University

²Professor, Division of Computer Science & Engineering, Kongju National University

요약 청각장애인들은 수어통역 없이 상담서비스를 이용하기에는 어려움이 있다. 수어 통역사 인력이 많이 부족하여 수어 통역사까지 상담이 연결되는데 많은 시간이 걸리거나 연결이 되지 않는 경우가 많이 발생하고 있다. 이에 본 논문에서는 OpenCV와 CNN(Convolutional Neural Network)을 이용하여 수어를 영상으로 촬영하고 수어 동작을 인식하여 수어가 뜻하는 의미를 텍스트 형태의 데이터로 변환하여 사용자에게 제공하는 시스템을 제안한다. 상담사는 저장된 수어번역 상담내용을 열람하여 상담을 진행할 수 있다. 전문 수어 통역사가 없어도 상담이 가능하여 수어 통역사를 기다려야 하는 부담을 줄일 수 있다. 제안 시스템을 청각장애인 상담서비스에 적용할 경우 상담 효과의 향상과 향후 청각장애인 상담에 관한 학문적 연구를 촉진하는 계기가 될 것으로 기대된다.

주제어 : OpenCV, 합성곱 신경망, 수어, 청각장애인, 영상처리

Abstract It is difficult for the hearing impaired to use the counseling service without sign language interpretation. Due to the shortage of sign language interpreters, it takes a lot of time to connect to sign language interpreters, or there are many cases where the connection is not available. Therefore, in this paper, we propose a system that captures sign language as an image using OpenCV and CNN (Convolutional Neural Network), recognizes sign language motion, and converts the meaning of sign language into textual data and provides it to users. The counselor can conduct counseling by reading the stored sign language translation counseling contents. Consultation is possible without a professional sign language interpreter, reducing the burden of waiting for a sign language interpreter. If the proposed system is applied to counseling services for the hearing impaired, it is expected to improve the effectiveness of counseling and promote academic research on counseling for the hearing impaired in the future.

Key Words : OpenCV(Open Source Computer Vision), CNN(Convolutional Neural Networks), Sign Language, Hearing-Impaired Person, Image Processing

1. 서론

청각장애인들의 상담서비스 이용은 의사소통의 제약으로 상담서비스를 이용하는데 어려움이 있다. 일반적

으로 장애를 가진 사람은 비장애인들에 비해 낮은 자존감, 낮은 참여수준, 우울함 등 감정적인 부분에서 많은 어려움을 나타내고 있어 상담의 역할이 매우 중요하다고 할 수 있다[1]. 기존의 청각장애인을 위한 상담서비

*Corresponding Author : Koo-Rack Park(ecgrpark@kongju.ac.kr)

스는 전문 수어 통역사 또는 수어가 가능한 상담사와 연결이 되어야 상담이 가능하다. 그러나 전문 수어 통역사 인력의 숫자가 현저히 부족하고 항상 상담 진행을 위해 대기하는 인력의 수가 부족하여 청각장애인과 수어 통역사가 연결되기까지 시간이 오래 걸릴 뿐만 아니라 연결이 되지 않는 경우가 많아 청각장애인들이 상담 서비스를 이용하는데 많은 어려움이 있다[2]. 수어통역 이용경험이 있는 청각장애인들은 자신감이 생기고 주변을 더 잘 이해할 수 있었으며, 사회와의 소통의 폭이 넓어져 더 나은 삶을 누릴 수 있기 때문에 수어 통역은 굉장히 그 역할이 중요하다[3]. 이에 따라 청각장애인 수어 통역에 관련된 많은 연구들이 진행되고 있다. 스마트폰과 수화동작을 연동하여 음성과 텍스트로 통역해주는 수어통역 장갑과 청각장애인에게 소리 정보의 부재를 대신하여 장갑에 부착된 LED 조명의 시각적 효과를 통해 상황 인지 능력에 도움을 주는 IoT 기기에 관한 연구[4], 3D 모델링을 통해 손모양의 동작을 분석하여 수어를 통역하는 연구[5], 광학 흐름을 활용한 3D 수어 번역기에 관한 연구[6] 등 수어통역에 관련된 다양한 연구들이 진행되고 있으나, 수어통역 기술을 청각장애인 상담에 적용하는 연구는 미미한 상황이다. 특히 실용적인 측면과 휴대성 및 경제성 등의 제약적인 부분들로 인해 청각장애인들의 상담서비스 이용에 어려움이 있어 이를 개선시키기 위한 시스템이 필요하다. 본 논문에서는 OpenCV를 활용하여 청각장애인의 수어를 영상으로 촬영하여 수어 동작을 인식하고, CNN 기법을 활용하여 촬영된 수어 영상을 통역 및 번역하여 텍스트 형태의 데이터로 변환하고, 이를 상담사에게 제공하여 상담의 효율성을 높이는 시스템을 제안한다. 제안 시스템을 통하여 전문 수어 통역사가 없어도 청각장애인이 언제든지 상담서비스를 이용하는데 많은 도움이 될 것이며, 상담의 효과가 향상될 수 있어, 청각장애인들의 삶의 질 또한 높아질 것으로 기대된다.

2. 관련연구

2.1 OpenCV

OpenCV(Open Source Computer Vision)는 실시간 이미지 프로세싱 라이브러리이다. Intel에서 1999년에 개발한 공개 라이브러리로서 프로그래밍 언어 C언어와 C++로 개발되었으며 Python, Java 및

MATLAB에 바인딩 되어 편리한 개발환경을 지원한다 [7]. Window, Linux, Android OS에서 구동이 가능하며 실시간 이미지 프로세싱에 중점을 둔 라이브러리이다[8]. 기초 영상처리에서부터 고급 수준의 영상처리까지 많은 양의 알고리즘들이 함수로 구현되어 있어 영상처리부분에 있어 유용한 기술들을 제공해 준다. OpenCV는 Intel CPU에 최적화 되어 있어 실행 속도가 빠르며, 오픈소스 특성상 다양한 소스들이 공유되고 있어 규칙에 맞는 코딩을 작성하면 OpenCV의 알고리즘에 자신이 만든 알고리즘을 추가할 수 있다[9]. OpenCV를 이용하여 다음과 같은 연구들이 진행되고 있다. OpenCV의 영상처리와 TensorFlow 딥러닝 영상처리 기반으로 기존의 등록된 실종자의 이미지와 새로 등록되는 해당 실종자의 이미지를 비교하여 이미지의 유사도를 추출하여 실종자에 관한 정보를 제공하는 연구[10], OpenCV를 기반으로 한 영상 개선, 영상 이진화, 템플릿 영상 정합, 특징점 추출 등의 영상처리 기법을 활용한 도로표지 영상에서 방향정보들의 자동 인식과 추출에 관한 연구[11]가 진행되고 있으며, OpenCV를 기반으로 RGB 색 채널과 Brute-Force Hamming 알고리즘, k-평균 클러스터링과, ORB를 이용한 안드로이드 기반 플랫폼의 꽃 인식 시스템에 관한 연구[12] 등 다양한 분야에서 OpenCV와 다양한 기술들을 접목한 연구들이 진행되고 있다. 본 논문에서는 오픈소스로 제공되고 있는 Teachable Machine 기술을 이용하여 시스템을 구현하였다.

2.2 CNN

CNN(Convolutional Neural Network)은 특정 객체를 이미지로부터 식별하는데 사용되는 인공신경망으로서 합성곱 연산을 적용하고, 신경망을 연산하여 이미지 처리에 특화된 모델로 자연어 해석, 이미지, 영상 인식 및 분류 등 다양한 분야에서 활용되고 있다[13]. 다음의 Fig. 1은 CNN의 구조를 나타낸 것으로 이미지의 특징을 추출하는 부분과 클래스를 분류하는 부분으로 나눌 수 있다.

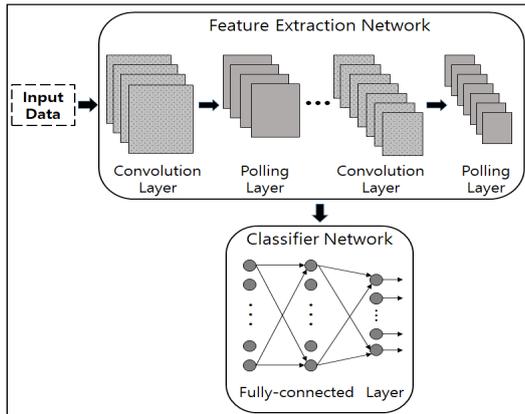


Fig. 1. Structure of Typical CNN

이미지의 특징을 추출하는 영역은 Convolution Layer와 Pooling Layer를 여러 겹 쌓는 형태로 구성된다. 입력층(Input Layer)에 입력된 이미지 데이터는 합성곱층(Convolutional Layer)에서 픽셀단위의 숫자로 변환되어 필터링 과정을 통하여 이미지의 특징을 추출한다. 추출된 특징들은 특성맵으로 저장하고, 필터의 크기와 연산 방법을 다르게 설정하여 이미지의 특징을 다양하게 추출할 수 있다[14]. Pooling Layer에서는 입력된 이미지의 크기를 조절하여 연산되는 용량을 감소시키고 다음 층으로 이미지의 특징을 전달한다. 완전 연결층(Fully Connected Layer)에서는 앞의 과정들을 반복적으로 실행한 후 3차원에서 1차원으로 값을 전환하고 식별하고자 하는 이미지가 일치하는지 확인하고 출력층(Output Layer)에서 식별 결과를 출력한다 [15]. 본 논문에서는 일반적인 CNN모델과 ResNet50, VGG-16등 특성 추출 성능이 검증된 주요 CNN모델의 구조를 적용하였다.

3. 제안 시스템

3.1 시스템 구성도

다음의 Fig. 2는 제안 시스템의 전체적인 시스템 구성도로서 크게 3부분으로 구성되어 있다. 상담을 요청하는 청각장애인들인 사용자, 상담을 진행하는 상담사들인 사용자가 시스템에 접속하기 위한 클라이언트 브라우저, 수어에 관한 대량의 데이터들을 기계학습을 통해 시스템에 데이터를 학습시키고, 사용자로부터 입력된 데이터를 처리해주는 시스템 서버, 학습된 데이터

및 입력된 데이터들을 저장하는 데이터베이스, 대용량 이미지 파일들을 효율적으로 관리하기 위한 별도의 스토리지를 따로 두고 구성하였다.

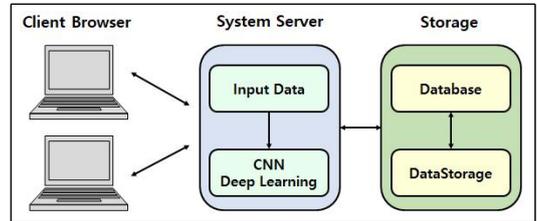


Fig. 2. Proposed Model Configuration

제안 시스템에서 사용한 플랫폼은 Open Source를 적극적으로 활용하여, 상호 운용성에 대한 향상 효과 및 유지보수 부분에서도 효과를 보기 위해, SpringMVC 프레임 워크를 기반으로 시스템을 구현하였다.

3.2 시스템 프로세스

다음의 Fig. 3은 제안 시스템의 전체적인 프로세스로서 크게 4단계로 구성되어 있다.

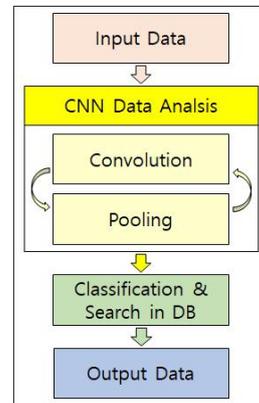


Fig. 3. System Process

첫째, 청각장애 사용자로부터 수어를 사용한 상담영상 데이터가 입력된다. 입력된 상담데이터는 연속적인 이미지 데이터로 변환되어 시스템에 전달된다. 둘째, 입력된 상담데이터는 CNN 데이터 분석 단계에서 각각 이미지 데이터에서 오브젝트의 특징들을 추출한다. 셋째, 추출된 오브젝트들을 분석하고 시스템이 학습한다. 학습을 통해서 기존의 데이터들과 비교하고 기존에 학

습된 모델에 대한 업데이트를 진행한다. 데이터베이스에 저장되어 있는 학습 모델들과 비교하고 일치하는 데이터를 검색한다. 넷째, 검색된 결과 데이터를 텍스트 데이터 형식으로 사용자에게 화면에 출력하여 제공한다.

3.3 시스템 사전학습

제안 시스템의 학습을 위한 사전 훈련작업으로 국립국어원에서 온라인을 통하여 제공하고 있는, 한국어사어 사전에 있는 일상생활에서 많이 사용되는 수어 단어 300개를 선별하여, 영상으로 촬영하고 시스템에 학습 시켰다. 촬영하는 영상의 Frame Rate는 초당 24개의 프레임을 촬영하기 위해 일반적으로 영화 촬영 시 표준이 되는 24fps(Frame Per Second)를 설정하여 촬영을 진행하였다. 촬영된 영상은 연속된 이미지로 시스템에 입력이 되며 시스템의 학습 진행 속도 향상을 위해 이미지 크기를 가로, 세로 224픽셀로 고정하였다. 영상을 촬영하는 동안 입력되는 이미지들은 설정한 고정 사이즈로 시스템에 입력된다. 다음의 Fig. 4는 입력된 이미지 데이터들의 Labeling 과정의 일부분이다.

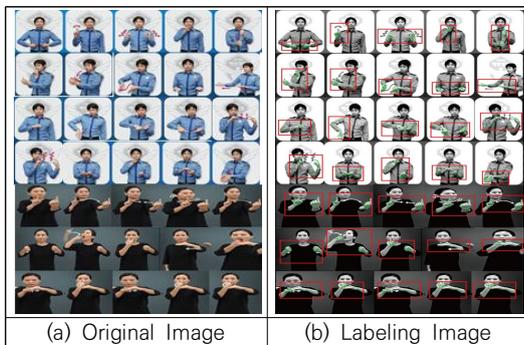


Fig. 4. Part of System's Image Labeling Process

(a)의 경우 원본영상을 시스템이 설정에 맞는 데이터로 변환된 이미지이다. (b)의 경우 이미지의 특징추출을 위해 OpenCV에서 제공되는 이진화 기법을 사용하여, 변환된 이미지들의 색을 인식하여 전처리 작업을 수행하였다. 이진화 처리된 데이터는 오픈소스로 제공되고 있는 Teachable Machine의 기술을 활용하여 이미지 내의 신체 영역에 대한 객체를 생성하고, 손부분의 오브젝트 영역을 사각형으로 영역을 지정하고, 손의 마디 부분에 객체를 생성하여 데이터화 하고, 이를 데이터베이스에 저장한다. Labeling이 완료되고 난 후,

시스템은 학습을 진행한다. 다음의 Fig. 5는 제안 시스템의 CNN 모델링 알고리즘의 일부분이다.

```
Epoch = 800
Batch_size = 32
Learning_rate = 0.001
With session(graph = g) as session :
    Execution session_data_reset_function()
    for i in range (30) :
        batch_data, batch_label =
            batch(img_list, Batch_size),
            l = session.run([train, loss]),
            feed_dict = {x: batch_data
                        y: batch_label })
        print(i, 1)
    saver.save(session, "", global_step = i + 1)
```

Fig. 5. Part of System Learning CNN Algorithm

시스템이 학습을 진행하는데 있어 전체 데이터 셋에 대한 학습과정 1회를 완료한 상태를 나타내는 epochs와 batch마다 데이터의 크기를 조절하는 batch size, 학습속도를 조절해주는 learning rate 항목들의 값을 각각 epoch = 800, batch Size = 32, learning rate = 0.001로 설정하고 학습을 수행하였을 때, 가장 높은 정확도를 나타내었으며, 본 연구에서는 5 Depth 레이어를 구성하여 실험을 진행하였다. 다음의 Fig. 6은 입력된 데이터들을 시스템이 학습을 진행함에 있어 validation loss의 변화를 나타낸 그래프이다.

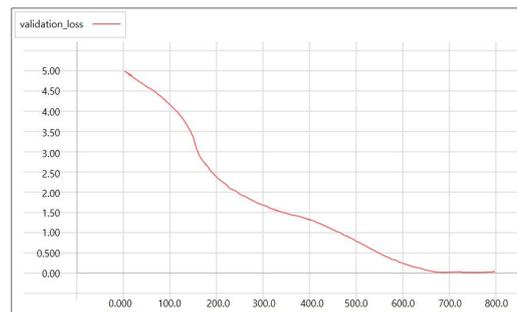


Fig. 6. Graph of Change in Validation Loss of System

시스템이 학습을 시작하여 622회까지 점점 하락하였고, 631회부터 학습이 진행되는 동안 validation loss는 완만하게 줄어드는 것을 확인할 수 있었다. 최종 학습 횟수인 800회가 완료된 후 validation loss = 0.0016297241 데이터를 획득하였다.

4. 실험 및 고찰

전처리 작업을 통하여 제안 시스템의 학습 과정이 완료된 후, 임의의 회화 데이터 50개를 구성하여 시스템에 입력하여 테스트를 진행하였다. 다음의 Fig. 7은 테스트 데이터를 입력하였을 때 시스템으로부터 출력되는 화면이다.

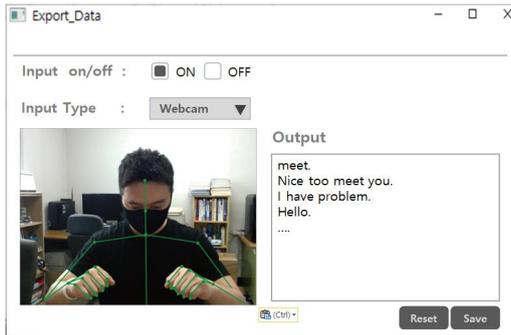


Fig. 7. Sign Language Processing Progress Screen

데이터 입력 시 화면에 출력되는 Input on/off 항목의 기능을 통하여 웹캠 또는 카메라의 사용여부를 결정할 수 있다. Input Type 버튼을 통해 입력하려는 데이터를 실시간 영상 또는 파일 업로드 중 하나를 택하여 입력방식을 설정할 수 있다. 화면 우측에는 입력된 데이터를 시스템이 분석한 결과가 텍스트로 출력되는 영역이다. 출력된 데이터를 저장 및 초기화 버튼을 사용하여 원하는 작업을 진행할 수 있다. 다음의 Table. 1은 테스트 결과를 정리한 것으로, 테스트 데이터를 시스템에 입력하여 획득한 결과이다.

Table 1. The Result of Test

Division	Accuracy [%]	Loss
Validation	98.48	0.0016297241
Training	97.14	0.0024
Testing	90.23	0.0312

Validation과 Training 데이터의 정확도 측정결과는 각각 98.48%, 97.14%로 나타났으며, 손실함수 loss의 값은 0.0016297241, 0.0024로 나타났다. 테스트 데이터의 결과로는 90.23%의 정확도를 보였으며, loss 값은 0.0312로 도출되었다. 일반적인 문장의 표현에 대해서는 97%이상의 정확도를 나타내었으나, 인간의

심리적 또는 감성적인 함축된 표현에 대해서는 입력한 예시 데이터와 같이 분석되지 않았고, 통역 부분에 있어 다소 정확도가 떨어지는 것을 확인할 수가 있었다.

5. 결론 및 향후 연구방향

본 논문에서는 수어 통역사 인력부족으로 인해 청각 장애인들의 상담 서비스 이용에 관한 불편함에 대한 문제점을 개선하고자, OpvenCV와 CNN을 이용하여 수어 통역 없이도 청각장애인들에게 상담서비스를 이용할 수 있는 시스템을 제안한다. 전처리 학습과정으로 온라인에서 제공되는 수어 데이터들을 CNN기술을 이용하여 시스템에 학습시킨 후 임의의 데이터를 입력하여 테스트한 결과 90.23이라는 정확도를 나타내었다. 입력되는 데이터 중 일반적인 데이터에 비해 좀 더 심도 깊은 대화 내용인, 인간의 내적인 표현 또는 함축적 의미를 담고 있는 표현에 대해서는 통역의 정확도가 다소 떨어지는 것을 확인할 수 있었다. 향후에는 함축적인 표현이나 인간의 심리적인 표현들에 대해서 많은 데이터를 수집하고, 표정과 행동 등 다양한 상담에 관련된 데이터를 수집하여 점진적인 기계학습의 발전을 통하여, 국내 수어뿐만 아니라 전 세계의 수어 데이터를 표준화 하여, 이를 자동으로 통역해주는 연구와 같이 모든 청각장애인들의 삶의 질을 향상시킬 수 있는 연구들이 계속 되어야 할 것이다.

REFERENCES

- [1] H. E. Kim, H. J. Koh, D. I. Kim & S. M. Choi. (2019). Research Trends and Support Measures in Counselling Service for Children and Adolescent with Disabilities in Korea. *Asian Journal of Education*, 20(3), 831-852. DOI : 10.15753/aje.2019.09.20.3.831
- [2] Y. I. Cho, B. C. Yoon & U. J. Min. (2012). A Qualitative Study for a Educational Sign Language Interpretation Service of Deaf Students at G High School. *The Journal of Special Children Education*, 14(3), 237-255. DOI : 10.21075/kacs.2012.14.3.237
- [3] J. R. Park. (2010). A qualitative study on the Deaf for Meaning of the Sign Language Experience and Sign Language interpretation services Experience based on narrative inquiry. *Social Science Research Review*, 26(4), 93-122.

[4] S. M. Koo, I. G. Jang & Y. S. Son. (2018). An Open Source Hardware based Sign Language Interpreter Glove & Situation Awareness Auxiliary IoT Device for the Hearing Impaired. *KIISE Transactions on Computing Practices*, 24(4), 204-209.
DOI : 10.5626/KTCP.2018.24.4.204

[5] D. Rempel, M. J. Camilleri & D. L. Lee. (2017). The design of hand gestures for human-computer interaction: Lessons from sign language interpreters. *International Journal of Human-Computer Studies*, 72(10-11), 728-735.
DOI : 10.1016/j.ijhcs.2014.05.003

[6] R. Sigit & D. R. Kartika. (2016). 3D Sign language translator using optical flow. *International Electronics Symposium(IES)*, 262-266.
DOI : 10.1109/elecsym.2016.7861014

[7] G. T. Lee, T. H. Kim, J. H. Cho & M. K. Moon. (2017). Development of Parking Management System Using OpenCV for Handicapped Person. *Proceedings of the HCI Society of Korea Conference*, 1187-1189.

[8] S. H. Lee & H. M. Ahn. (2012). Piano practice using OpenCV and the Android application project. *Proceedings of the Korean Society of Computer Information Conference*, 20(2), 267-268.

[9] J. H. Kim, S. Y. Shim, B. J. Oh, J. A. Lee, H. W. Choi & W. O. Cha. (2006). Implementation of an educational interactive game using OpenCV and color recognition. *Proceedings of the Korea Multimedia Society Conference*, 30-34.

[10] Y. T. Baek, S. H. Lee & J. S. Kim. (2017). Intelligent missing persons index system Implementation based on the OpenCV image processing and TensorFlow Deep-running Image Processing. *Journal of the Korea Society of Computer and Information*, 22(1), 15-21.
DOI : 10.9708/jksoci.2017.22.01.015

[11] G. H. Kim, K. S. Chong & J. H. Youn. (2013). Automatic Recognition of Direction Information in Road Sign Image Using OpenCV. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, 31(4), 293-300.
DOI : 10.7848/ksgpc.2013.31.4.293

[12] K. C. Kim & C. Yu. (2017). Flower Recognition System Using OpenCV on Android Platform. *Journal of the Korea Institute of Information and Communication Engineering*, 21(1), 123-129.
DOI : 10.6109/jkiice.2017.21.1.123

[13] LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning. nature*, 521(7553), 436-444.
DOI : 10.1038/nature14539

[14] T. Bluche, H. Ney & C. Kermorvant. (2013). Feature extraction with convolutional neural networks for handwritten word recognition. *In 2013 12th International Conference on Document Analysis and Recognition. IEEE*, 285-289.
DOI : 10.1109/ICDAR.2013.64

[15] N. K. Lee, J. Y. Kim & J. H. Shim. (2021). Empirical Study on Analyzing Training Data for CNN-based Product Classification Deep Learning Model. *The Journal of Society for e-Business Studies*, 26(1), 107-126.
DOI : 10.7838/jsebs.2021.26.1.107

박 원 철(Won-Cheol Park)

[정회원]



- 2011년 2월 : 공주대학교 컴퓨터공학과(공학사)
- 2013년 2월 : 공주대학교 대학원 멀티미디어공학과(공학석사)
- 2018년 8월 : 공주대학교 대학원 컴퓨터공학과(공학박사)

- 관심분야 : 영상처리, 자동화 시스템, 머신러닝, IT 컨버전스
- E-Mail : pwcfrog@gmail.com

박 구 락(Koo-Rack Park)

[정회원]



- 1986년 2월 : 중앙대학교 전기공학과(공학사)
- 1988년 2월 : 숭실대학교 전자계산학과(공학석사)
- 2000년 2월 : 경기대학교 전자계산학과(이학박사)

- 1991년 ~ 현재 : 공주대학교 컴퓨터공학부 교수
- 관심분야 : IT 컨버전스, 정보통신, 머신러닝, 전자상거래
- E-Mail : ecgrpark@kongju.ac.kr