Journal of the Korea Academia-Industrial cooperation Society
Vol. 21, No. 7 pp. 685-691, 2020

https://doi.org/10.5762/KAIS.2020.21.7.685
ISSN 1975-4701 / eISSN 2288-4688

밀리미터파의 손동작 인식 알고리즘에 관한 연구

남명우*, 홍순관
혜전대학교 전기과

Study on Hand Gestures Recognition Algorithm of Millimeter Wave

Myung Woo Nam*, Soon Kwan Hong
Dept. of Digital Electronics, Hyejeon College

요 약 본 논문에서는 77GHz를 사용하는 밀리미터파 레이더 센서의 반향 신호를 이용하여 손동작의 움직임을 추적한 후 얻어진 데이터로 0부터 9까지의 숫자들을 인식하는 알고리즘을 개발하였다. 손동작을 감지하여 레이더 센서로부터 얻어진 반향 신호들은 산란 단면적의 차이 등에 의해 불규칙한 점들의 군집형태를 보인다. 이들로부터 유효한 중심점을 얻기 위해 3차원 좌푯값들을 이용해 K-Means 알고리즘을 적용하였다. 그리고 얻어진 중심점들을 연결하여 숫자 형태의 이미지를 생성하였다. 얻어진 이미지와 스무딩 기법을 적용해 사람의 손글씨 형태와 유사하게 만든 이미지를 MNIST(Modified National Institute of Standards and Technology database)로 훈련된 CNN(Convolutional Neural Network) 모델에 입력하여 인식률을 비교하였다. 실험은 두 가지 방법으로 진행되었다. 먼저 스무딩 기법을 적용한 이미지와 적용하지 않은 이미지를 사용한 인식 실험에서는 각각 평균 77.0%와 81.0%의 인식률을 얻었다. 그리고 학습데이터를 확장(augmentation)한 CNN 모델의 실험에서는 스무딩 기법을 적용한 이미지와 적용하지 않은 이미지를 사용한 인식 실험에서 각각 평균 97.5%와 평균 99.0%의 인식률을 얻었다. 본 연구는 레이더 센서를 이용한 다양한 비접촉 인식기술에 응용이 가능할 것으로 판단된다.

Abstract In this study, an algorithm that recognizes numbers from 0 to 9 was developed using the data obtained after tracking hand movements using the echo signal of a millimeter-wave radar sensor at 77 GHz. The echo signals obtained from the radar sensor by detecting the motion of a hand gesture revealed a cluster of irregular dots due to the difference in scattering cross-sectional area. A valid center point was obtained from them by applying a K-Means algorithm using 3D coordinate values. In addition, the obtained center points were connected to produce a numeric image. The recognition rate was compared by inputting the obtained image and an image similar to human handwriting by applying the smoothing technique to a CNN (Convolutional Neural Network) model trained with MNIST (Modified National Institute of Standards and Technology database). The experiment was conducted in two ways. First, in the recognition experiments using images with and without smoothing, average recognition rates of 77.0% and 81.0% were obtained, respectively. In the experiment of the CNN model with augmentation of learning data, a recognition rate of 97.5% and 99.0% on average was obtained in the recognition experiment using the image with and without smoothing technique, respectively. This study can be applied to various non-contact recognition technologies using radar sensors.

Keywords : Millimeter Wave, Gesture Recognition, K-Means, Smoothing, Deep Learning

본 논문은 혜전대학교 교내연구과제로 수행되었음.
*Corresponding Author : Myung Woo Nam(Hyejeon College)
email: mwnam@hj.ac.kr
Received April 16, 2020 Revised May 6, 2020
Accepted July 3, 2020 Published July 31, 2020

1. 서론

맨-머신 인터페이스(man-machine interface)에는 키보드, 마우스, 터치스크린 등의 기존 인터페이스 장비들이 아직 많이 사용되고 있다. 그러나 인공지능 기술의 발달과 가상현실 및 맨-머신 인터페이스에 관심이 높아지며 사람의 신체기관을 이용한 정보전달 방법들이 빠르게 상용화되고 있다. 인간에게 있어 손동작은 자연스러운 형태의 의사소통 방법이다. 사람의 신체를 이용해 컴퓨터와 자연스러운 상호작용을 목표로 하는 NUI(Natural User Interface) 분야에서 손동작 인식기술은 매우 편리하고 직관적이기 때문에 과거부터 현재까지 지속해서 연구되어 오고 있다[1-2]. 손동작 인식에는 모션 센서를 이용한 동작 인식 그리고 광학 센서를 이용한 영상인식 기술 등이 사용되고 있으나, 모션 센서는 높은 비용과 사용상의 어려움이 있고 광학 센서는 조명과 주변 장애물 등에 의해 인식에 문제가 발생한다. 최근 이러한 문제점들로부터 자유롭고 다양한 환경에서도 사용할 수 있는 밀리미터파 레이더 센서를 이용한 손동작 인식 연구가 진행되고 있다[3-4]. 고주파 영역의 단파장을 이용하는 밀리미터파 레이더 센서는 높은 분해능과 단순 장애물 통과, 저비용 등의 장점이 있어 단순한 손동작을 인식하고 이를 통해 기기를 작동시키는 데 사용되고 있다. 그러나 레이더 센서를 이용한 액티브 계측방법은 밀리미터파를 물체에 방사하여 산란된 반향 신호를 얻는 방식이므로 물체의 단면적 변화에 민감하게 반응한다. 이로 인해 움직이는 물체에서는 불규칙한 산란 신호를 얻게 되어, 주로 물체의 추적이나 단순 동작을 인식하는 데 사용되고 있다. 많은 분야에서 비접촉 방식으로 컴퓨터와 데이터를 주고받아야 하는 경우가 발생한다. 이때 간단한 동작으로 데이터를 전송하기에는 전송속도와 방법 등에 제약이 따른다. 따라서 손동작을 통해 숫자를 전달할 수 있다면 빠르고 정확하게 많은 정보를 전달할 수 있을 것으로 판단된다.

본 논문은 밀리미터파 레이더 센서를 사용하여 단순 동작 인식이 아닌 손동작으로 표현한 0부터 9까지 숫자를 높은 정확도로 인식할 수 있는 알고리즘을 개발하였다.

2. 손동작 인식 시스템

2.1 하드웨어 구성

본 논문에 사용된 하드웨어는 TI사의 mmWave 레이더 센서로 모델명은 iwr1443Boost이다. 77GHz 단파장 신호를 사용하며 3개의 Tx 안테나와 4개의 Rx 안테나로 구성되어 있다. mmWave 레이더 센서는 파장의 길이가 밀리미터(mm)로 매우 짧은 단파장을 사용하여 신호를 보내기 때문에 높은 분해능과 안테나의 크기를 줄일 수 있는 이점이 있다[5]. Fig. 1은 실험을 위해 사용한 레이더 센서와 센서의 안테나 구조이다.

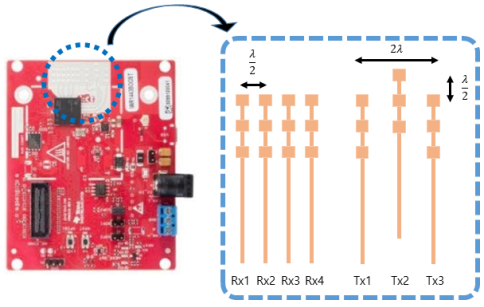


Fig. 1. iwr1443Boost & anntena layout

iwr1443Boost는 FMCW(Frequency Modulated Continuous Wave) 레이더 기술을 기반으로 하는 센서로 빌딩 자동화, 공장 자동화, 드론, 자재 취급, 교통 모니터링 및 감시와 같은 산업 응용 분야에서 저전력, 자체 모니터링, 초정밀 레이더 시스템으로 사용되고 있다. 그리고 PLL(Phase-Locked Loop) 및 AD 변환기가 내장된 안테나들과 복잡한 FFT(Fast Fourier Transform) 및 CFAR(Constant False Alarm Rate) 감지를 지원하는 하드웨어 가속기와 두 개의 ARM R4F 기반 프로세서가 포함되어 있다. 센서는 물체의 방위각과 고도, 속도 및 3차원 공간 좌표 (x, y, z)를 실시간으로 처리할 수 있다[5].

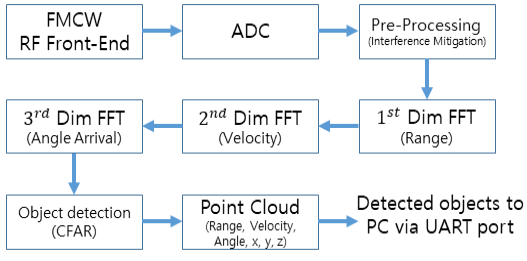


Fig. 2. iwr1443Boost received signal processing process

Fig. 2는 iwr1443Boost의 송신 안테나로 FMCW 레이더 신호를 송수신한 후 처리하는 과정을 나타내었다. 수신된 신호는 Range FFT와 Doppler FFT를 적용하여 물체의 거리와 속도, 각도를 계산하게 된다[3-4]. 물체의

거리와 속도 등을 구하는데 유용한 FMCW 신호는 주파수가 시간에 따라 선형으로 변하는 선형적인 주파수변조 방식을 뜻한다[6]. FMCW 신호를 생성하기 위해 iwr1443Boost의 펌웨어에 제공된 설정값들은 Table 1과 같다. 최대 객체 감지 거리는 9.62 m이며 UART 포트 통신속도를 고려하여 1초에 5번의 실시간 감지 데이터를 수신하도록 설정하였다.

Table 1. FMCW configurations for iwr1443Boost

Parameters	Value
Range resolution	0.047 m
Maximum unambiguous range	9.62 m
Maximum radial velocity	2.05 m/s
Radial velocity resolution	0.26 m/s
Frame duration	200 msec

2.2 소프트웨어 구성

레이더 센서로부터 얻어진 데이터를 가공하여 실시간으로 손동작을 추적하기 위해 매트랩을 사용하여 프로그램을 작성하였다. UART 포트를 통해 얻은 레이더 센서 데이터를 목적에 맞게 분류한 후, 3차원 공간 좌표(x, y, z)와 반향 신호의 크기, 객체의 속도정보를 얻었다. 얻어진 정보는 반향 신호들이 군집화되어있는 구조로 고정된 주변 사물들의 정보를 제거하기 위해 움직임이 있는 객체들의 정보만 수집하였다.

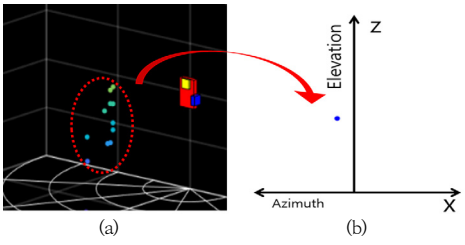


Fig. 3. Receiving data of the radar sense the hand gesture (a) detected moving objects (b) 2D center point of reflect data using K-Means algorithm

Fig. 3에 손의 움직임으로부터 얻어진 레이더 센서의 반향 신호들과 K-Means 알고리즘으로 얻어진 중심점을 나타내었다.

2.3 손동작 추적 알고리즘

손동작을 추적하기 위해서는 반향된 군집 신호들로부터 유효한 2차원 좌푯값을 얻어내고 이를 연결하여 숫자로 표현하는 알고리즘이 필요하다. 손동작을 추적하여 얻어진 군집화 데이터들은 손과 손목, 팔꿈치 등 전반적인 움직임이 추적된 데이터들이다. 이로 인해 움직임으로 얻어진 군집화 데이터들은 상황에 따라 복잡한 형태로 나타나게 된다. 이를 해결하기 위해 얻어진 데이터의 거리 정보를 분석하여 센서와 근접한 거리에 모여있는 신호만을 선택하여 K-Means 알고리즘을 적용하였다. 이는 손동작으로 숫자를 표시할 때 손이 레이더 센서로부터 제일 가깝게 위치할 수밖에 없는 신체 구조 때문이다. 이러한 중심점들로부터 얻어진 경로는 불규칙한 모양을 형성하였다. 구현된 알고리즘은 Fig. 4와 같다.

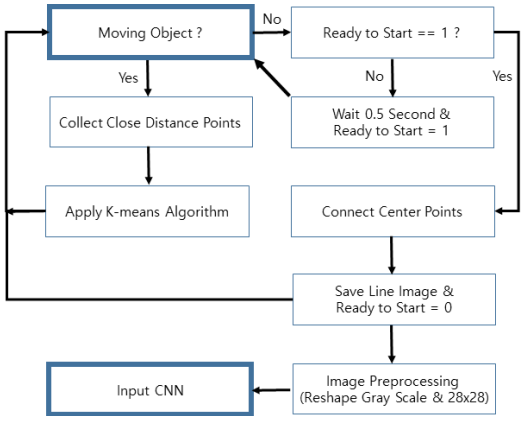


Fig. 4. Hand gesture recognition algorithm

2.4 MNIST로 학습된 CNN 모델

MNIST는 손으로 쓰인 28×28 크기의 0부터 9 사이의 숫자 이미지로 이루어진 대형 데이터베이스이며, 머신러닝(machine-learning)을 위해 일반적으로 사용된다[7]. 본 논문에서는 손동작으로 구해진 경로가 사람이 손으로 쓴 숫자들의 형태와 유사하다는 점에 주목하고 MNIST를 인식을 위한 학습데이터로 사용하였다. 사용된 CNN 모델은 TensorFlow 오픈소스 라이브러리를 이용하여 구성하였으며, 28×28 입력 이미지를 3×3 컨벌류션 레이어(convolution layer)의 필터와 풀링 레이어(pooling layer)에 반복적으로 통과시킨 후 기존의 MLP(multi-layer perceptron)와 softmax 함수를 통해 이미지를 분류하였다[8-10]. Fig. 5는 실험에 사용된 CNN 모델의 구조이다. 사용된 CNN 모델은 60,000개의 데이터로 구성된 MNIST를 이용하여 99.52%의 인식률을 얻을 수 있도록 학습을 시킨 후 실험에 사용하였다.

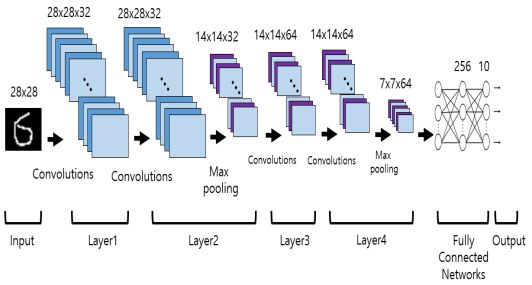


Fig. 5. Convolution neural network structure

3. 손동작 인식 실험

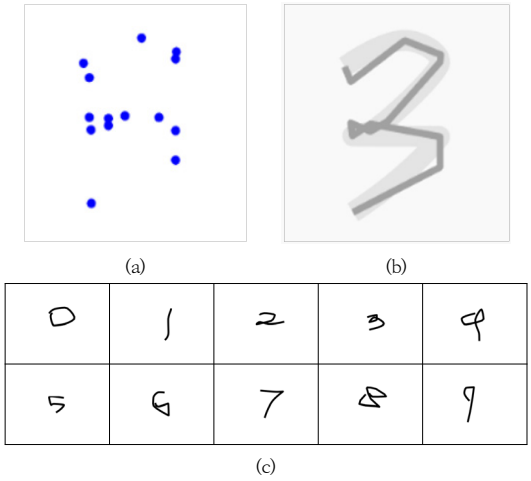
손동작 인식을 위해 레이더 센서 반향 신호들의 3차원 정보를 이용해 중심점을 추출하였고, 중심점들을 연결하여 숫자 형태를 생성한 후 인식 실험에 사용하였다. 그리고 불규칙한 숫자 형태를 보완하기 위해 스무딩 기법을 적용한 결과를 이용해 추가 실험을 하였다.

Fig. 6. Hand gesture for recognition

실험은 Fig. 6과 같이 레이더 센서에서 1 m정도 떨어진 거리에서 손으로 숫자를 그리는 방식으로 진행되었다. 실험결과 허공에 손바닥 또는 손끝으로 숫자를 그려도 차이가 없었으며, 너무 빠르지만 않다면 손 움직임의 속도도 문제 되지 않았다. 실험에 사용된 레이더 센서는 움직임이 있는 물체만을 감지하도록 설정되어 있다. 이는 주변의 고정된 사물 정보를 제거하기 위함이다. 먼저 0.5초 동안 손의 움직임이 감지되지 않으면 손동작 입력 상태로 전환하도록 프로그램하였다. 이때부터 손으로 숫자를 허공에 그리게 되면 레이더 센서가 움직임을 감지하고 UART 포트로 데이터를 전송해준다.

실험을 위해 레이더 센서 앞에서 0부터 9까지 숫자별로 40번씩 손동작으로 숫자를 그렸다. 이렇게 얻어진 총 400개의 숫자 이미지들을 인식률 실험을 위한 데이터로 사용하였다.

3.1 스무딩 기법 적용



(a) (b)

(c)

Fig. 7. Hand gesture path result
(a) center points (b) 3 using original & smoothing path (c) digits between 0 and 9 using smoothing path

Fig. 7은 레이서 센서의 수신 데이터를 가공해 얻은 손동작 경로이다. Fig. 7 (a)는 K-Means 알고리즘으로 얻어진 중심점들이다. 그리고 이것들을 이어서 생성한 숫자는 Fig. 7 (b)의 얇은 선이며 불규칙한 형태를 보인다. 이러한 현상은 직선 또는 복잡한 형태를 띠는 1, 3, 4, 8 등의 숫자에서 심하게 나타났다. 본 실험에서는 이것을 보정하여 사람의 손글씨 형태와 유사한 형태를 보일 수 있도록 스무딩 기법을 적용한 후 인식률을 비교해 보았다. Fig. 7 (b)의 두꺼운 선은 스무딩 기법을 적용한 결과이다. Fig. 7 (c)는 0부터 9까지 스무딩 기법을 적용하여 얻어진 숫자 이미지들이다. 스무딩 기법은 구분적 3차 에르미트 보간법(PCHIP: Piecewise Cubic Hermite Interpolating Polynomial)을 사용하였다[11].

인식률 실험을 위해 생성한 총 400개의 숫자 이미지들을 숫자별로 스무딩 기법이 적용된 40개의 이미지와 적용되지 않은 10개로 각각 분류하였다. 그리고 이들을 CNN 모델에 입력하여 얻은 결과는 Table 2와 같다. 스무딩 기법이 적용된 이미지는 평균 77.0%, 적용되지 않은 이미지는 평균 81.0%의 인식률을 보였다. 특히 사람이 작성한 손글씨의 형태와 다른 형태를 보이는 숫자 4는 0과 8, 9 등으로 오인식되었으며, 숫자 6은 5와 3으로 오인식되었다. 사람의 손글씨 형태와 유사하도록 스무딩 기법을 적용한 실험결과는 원본 이미지를 사용한 경

우보다 낮게 나왔다. 이는 스무딩 기법이 적용되면서 숫자 형태의 불규칙성과 독창성이 감소하였고 이는 결과적으로 CNN 모델의 변별력을 감소시킨 것으로 판단된다. 이로 인해 스무딩 기법을 적용한 이미지들은 8과 9에서 0, 9, 8과 같은 오인식 결과가 더 많이 발생하였다.

Table 2. Hand gesture recognition result of CNN

Digits	Recognition rate(%)		Digits	Recognition rate(%)	
	Original	Smoothing		Original	Smoothing
0	80.0	85.0	5	80.0	85.0
1	100.0	92.5	6	50.0	45.0
2	90.0	95.0	7	100.0	95.0
3	100.0	90.0	8	80.0	67.5
4	40.0	40.0	9	90.0	75.0

3.2 CNN 모델 학습데이터 확장

MNIST 데이터베이스는 사람들이 작성한 손글씨 60,000개로 구성되어 있으며 작성된 숫자들은 이미지 중앙에 배치된 형태이다. 그리고 글씨 형태가 기울어짐이 크지 않은 정규화된 형태를 띠고 있다. 하지만 본 실험에서 손동작 경로로 얻어진 숫자들은 기울어짐과 숫자의 크기 등이 다양한 형태를 이루고 있다. 이를 보완하기 위해 숫자별로 스무딩 기법이 적용된 40개의 인식률 실험 이미지 중 20개를 CNN 모델 추가 학습을 위해 사용하였고, 나머지 20개로는 인식률을 구하는 실험을 추가로 진행하였다.

추가 학습을 위해 사용된 20개 이미지는 숫자 형태가 이미지 중앙에 위치하도록 조정한 후, 학습데이터 확장을 위해 좌우 기울기를 0°~20°까지 임의로 회전한 형태와 글자 크기를 100%~80%까지 임의로 줄인 형태, 그리고 상하좌우로 0%~5%씩 임의로 글자의 중심점을 옮긴 형태들로 새롭게 생성하였다[12].

Table 3은 28×28 크기로 정규화된 원본 이미지와 추가 학습에 사용된 스무딩 기법을 적용한 이미지, 학습데이터 확장용 이미지들이다.

학습데이터 확장을 위해 0부터 9까지 숫자별로 20개의 이미지를 랜덤하게 변형하여 숫자별로 400개의 이미지로 증가시킨 후 학습데이터로 사용하였다. 이는 머신러닝이 유사한 형태를 가지는 다수의 데이터로 훈련할 경우 높은 인식률을 보이기 때문이다. 숫자별로 400개씩 생성한 학습데이터 확장용 이미지 총 4000개와 MNIST 데이터 60,000개를 이용하여 CNN 모델을 인식률 99.29%까지 학습시켰다.

Table 3. 28×28 pixel grayscale augmentation images of hand gesture single digits between 0 and 9

Original	Smooth	Augmentation images		

Table 4는 새롭게 학습된 CNN 모델을 이용한 인식 실험결과이다. 실험에는 0부터 9까지 스무딩 기법이 적용된 20개의 이미지와 적용되지 않은 이미지 10개를 각각 사용하였다. 스무딩 기법이 적용된 이미지로는 평균 97.5% 인식률을 얻었고 적용되지 않은 이미지로는 평균 99.0%의 인식률을 얻었다. 레이더 센서로부터 얻어진 손동작 경로들이 학습데이터 확장을 통해 새롭게 CNN 모델에 학습되어 높은 인식률을 얻을 수 있었다. 그러나 스무딩 기법을 적용한 이미지들은 숫자 형태의 불규칙성이 사라져 이번 실험에도 원본 이미지보다 낮은 인식률을 보였다. 사람의 눈으로 판단할 때는 스무딩 기법이 적용된 숫자 형태가 인식하기 편했으나 CNN 모델에서는 반대의 결과를 보였다.

Table 4. Hand gesture recognition result of new trained CNN

Digits	Recognition rate(%)		Digits	Recognition rate(%)	
	Original	Smoothing		Original	Smoothing
0	100.0	100.0	5	100.0	95.0
1	100.0	100.0	6	100.0	90.0
2	100.0	100.0	7	100.0	100.0
3	100.0	100.0	8	90.0	90.0
4	100.0	100.0	9	100.0	100.0

4. 결론

본 연구에서는 레이더 센서를 이용하여 손동작을 추적한 후 0부터 9까지의 숫자를 인식하는 알고리즘을 구현하였다. 레이더 센서로부터 얻어진 반향 신호들은 손의 움직임에 따라 산란 단면적의 차이가 발생하여 불규칙한 점들의 군집형태로 나타난다. 얻어진 점들의 3차원 좌표 값들을 K-Means 알고리즘에 입력하여 유효한 중심점들을 추출한 후 연결하여 숫자 형태의 이미지로 생성하였다. 그리고 CNN 모델에 입력하여 인식률을 구하였다. 인식률 향상을 위해 얻어진 숫자들의 형태에 스무딩 기법을 적용하였고, CNN 모델 훈련에 학습데이터 확장기법을 사용하였다. 실험결과 CNN 모델 훈련에 학습데이터 확장기법을 적용하고, 스무딩 기법은 적용하지 않은 이미지가 99%로 가장 높은 인식률을 보였다. 이는 CNN 모델의 특성상 스무딩이 적용된 숫자 경로보다 불규칙한 특성을 보이는 경로들이 학습과 특징 추출에서 유리하게 작용한 것으로 판단된다. 본 실험의 결과로 레이더 센서와 손동작을 이용해 숫자 데이터를 높은 정확도로 전달 가능함이 확인되었고, 비접촉 데이터 전송 기술이 필요한 현장에 적용 가능함을 제시하였다.

향후 레이더 센서 반향 신호들의 중심점을 좀 더 정확히 추출하는 방법과 스무딩 기법을 보완한 인식률 향상 방법 그리고 이를 통해 작은 손동작에서도 숫자를 인식할 수 있는 연구를 진행할 예정이다.

References

[1] S. Mitra and T. Acharya, "Gesture Recognition: A Survey", in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 37, no. 3, pp. 311-324, May 2007.
DOI: http://dx.doi.org/10.1109/TSMCC.2007.893280

[2] H. C. Yoon and J. S. Cho, "Hand Feature Extraction Algorithm Using Curvature Analysis For Recognition of Various Hand Gestures", Journal of The Korea Society of Computer and Information, Vol.20, No.5, pp.13-20, May 2015.
DOI : http://dx.doi.org/10.9708/jksci.2015.20.5.013

[3] Y. S. Lee, "Study on the Hand Gesture Recognition System and Algorithm based on Millimeter Wave Radar", Journal of Korea Institute of Information, Electronics, and Communication Technology, Vol.12, No.3, pp.251-256, Dec. 2019.
DOI : http://dx.doi.org/10.17661/jkiiect.2019.12.3.251

[4] P. Molchanov, S. Gupta, K. Kim and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing", 2015 IEEE Radar Conference (RadarCon), Arlington, VA, pp. 1491-1496, Jun. 2015.
DOI: http://dx.doi.org/10.1109/RADAR.2015.7131232

[5] http://www.ti.com/lit/ds/symlink/iwr1443.pdf (accessed Dec. 01, 2019)

[6] http://www.ti.com/lit/an/swra553a/swra553a.pdf (accessed Dec. 01, 2019)

[7] http://yann.lecun.com/exdb/mnist/ (accessed March 10, 2020)

[8] https://www.tensorflow.org/ (accessed March 10, 2020)

[9] Jürgen Schmidhuber, "Deep learning in neural networks: An overview", Neural Networks, Vol.61, pp.85-117, Jan. 2015.
DOI: http://dx.doi.org/10.1016/j.neunet.2014.09.003

[10] M. J. Kang. "Comparison of Gradient Descent for Deep Learning" Journal of the Korea Academia-Industrial cooperation Society, Vol.21, No.2, pp.189-194, Feb. 2020.
DOI: http://dx.doi.org/10.5762/KAIS.2020.21.2.189

[11] Fritsch, F. N. and R. E. Carlson. "Monotone Piecewise Cubic Interpolation", SIAM Journal on Numerical Analysis. Vol. 17, pp.238-246, 1980.
DOI: http://dx.doi.org/10.1137/0717021

[12] https://github.com/franneck94/MNIST-Data-Augmentation (accessed March 10, 2020)

남 명 우(Myung Woo Nam) [종신회원]

- 1994년 2월 : 서울시립대학교 전자전기컴퓨터공학부(공학석사)
- 2001년 8월 : 서울시립대학교 전자전기컴퓨터공학부(공학박사)
- 2003년 3월 ~ 현재 : 혜전대학교 전기과 교수

〈관심분야〉
반도체소자, 디지털신호처리, 마이크로프로세서

홍 순 관(Soon Kwan Hong) [종신회원]

- 1994년 2월 : 서울시립대학교 전자전기컴퓨터공학부(공학석사)
- 2001년 8월 : 서울시립대학교 전자전기컴퓨터공학부(공학박사)
- 2003년 3월 ~ 현재 : 혜전대학교 전기과 교수

〈관심분야〉
반도체 소자, 전기기판(PCB), 디지털신호처리