

## mmWave 레이더 포인트 클라우드를 이용한 손동작 인식 경량화 딥러닝 모델

문지영, 이수진, 이소영, 손봉기\*, 강지현  
덕성여자대학교, 서원대학교\*

{moon09, sojinlee, dlthdud0112, jhkang}@duksung.ac.kr, \*bksohn@seowon.ac.kr

## Light-weight Deep Learning Model for Hand Gesture Recognition using Point-cloud of mmWave Radar

Jiyoung Moon, Soojin Lee, Soyoung Lee, Bong-Ki Son\*, and Jiheon Kang  
Duksung Women's Univ. Seowon University\*

## 요약

본 논문에서는 멀티채널 mmWave 레이더를 이용한 손동작 인식 시스템에서 딥러닝 모델의 입력으로 사용되는 3D 복셀 형태의 포인트 클라우드를 2D 이미지 형태로 투영 변환하여 딥러닝 모델의 연산량과 메모리 효율성을 개선한 경량화 모델을 제안한다. 포인트 클라우드의 희소한(sparse) 데이터 행렬에 따른 메모리 비효율성과 낮은 학습율을 유발하는 3D 복셀 표현 대신 투영 좌표축의 위치 정보를 포함한 2D 이미지로 차원 축소하여 딥러닝 모델의 입력으로 사용하였고, 제스처 분류 모델은 특징 추출을 위한 CNN과 시계열 특성을 고려한 양방향 LSTM을 결합하여 손동작을 인식 및 분류하였다. 기존 3D 복셀 기반 시스템과의 성능 및 메모리 효율성 비교 실험을 통해 제안 방식의 효율성을 검증하였다.

## I. 서론

최근 레이더 센서 기반의 손동작, 행동 인식 시스템은 사생활 침해 우려가 있는 카메라 기반 방식을 대체할 수 있는 수단으로 활용되고 있다[1]. 특히 레이더로 수집한 포인트 클라우드를 활용하는 경우 3차원 공간에서 직접적인 기하(위치 및 형태) 표현을 활용할 수 있다는 장점이 있다.

기존의 포인트 클라우드 기반 손동작 및 행동 인식에 대한 연구는 주로 3D 복셀을 입력 벡터로 사용하지만, 복셀 표현은 실질적으로 불필요한 정보(sparse data)를 다수 포함하며 입력 데이터가 커져 많은 컴퓨팅 자원(메모리, 연산량)이 필요하며, 불필요한 데이터로 인해 학습율이 낮아지는 문제가 발생한다[2].

따라서 본 논문에서는 모바일 또는 임베디드 장치에 적용 가능한 실시간 손동작 인식 시스템을 위해 손동작 인식 딥러닝 모델의 입력 벡터를 3D에서 2D로 차원 축소하고, 줄어든 입력에 따라 학습 파라미터를 감소시켜 적은 학습데이터에서도 효율적으로 동작하는 경량화된 딥러닝 모델을 제안한다.

## II. 본론

## A. 레이더 센서를 통한 포인트 클라우드 처리

포인트 클라우드 수집에 사용한 레이더 센서 모듈은 3개의 Tx, 4개의 Rx 안테나가 내장된 Texas Instruments사의 IWR6843AOPEVM이다. 레이더 센서가 물체로부터 반사된 신호를 처리하여 레이더 센서 위치 기반

상대적인 3차원 위치 정보를 계산하고, 레이더 센서의 거리/각도 분해능을 고려하여 좌표계로 변환하여 x, y, z 좌표 형식의 포인트 클라우드를 얻어낸다.

본 연구는 근거리에서 수행되는 손동작 인식을 목표로 하므로, 제스처가 발생할 것이라 예상되는 관심 영역을 그림 1과 같이 설정하여 필터링하였다. 과 같이 레이더의 정면(y축)으로부터 0~1m, 좌우(x축)로 -0.5~0.5m, 상하(z축)로 -0.5~0.5m 내에 있는 ROI(Region of Interest) 포인트만을 추출하여 활용하였다.

본 연구에서 손동작은 레이더의 정면에서 수행되므로 y축에 대한 움직임이 비교적 적다. 따라서 포인트 클라우드를 x-z 평면에 투영하여 32x32 크기의 1채널 이미지로 차원 축소하여 표현하였다.

y좌표값은 투영된 포인트의 픽셀 위치를 결정하는데 활용하였고, 일반적인 투영 기법에 의해 y축 위치 정보가 유실되는 것을 방지하기 위해 y좌표값이 작을수록 1에 가까운 값을, 멀수록 0에 가까운 값을 갖도록 정규화하였다. 또한 투영 위치가 겹치는 포인트의 경우 값을 합하여 2D 이미지의 픽셀값을 부여하였다. 레이더와 거리가 가까운 포인트일수록, 일직선상에 여러 포인트가 존재할수록 유의미한 신호로 표현하고, y축 값의 손실을 최소화하기 위함이다.

하나의 손동작이 1.5초 길이라고 가정하여 시퀀스를 분할하였고, 투영 이미지 시퀀스를 10프레임의 시간 윈도우를 간격으로 두고 27프레임씩(18FPS) 슬라이딩하여 하나의 동작 시퀀스로 처리하였다. 최종적으로 27x32x32 형태의 이미지 시퀀스 데이터를 제스처 분류 모델의 입력과 학습데이터로 사용하였다.

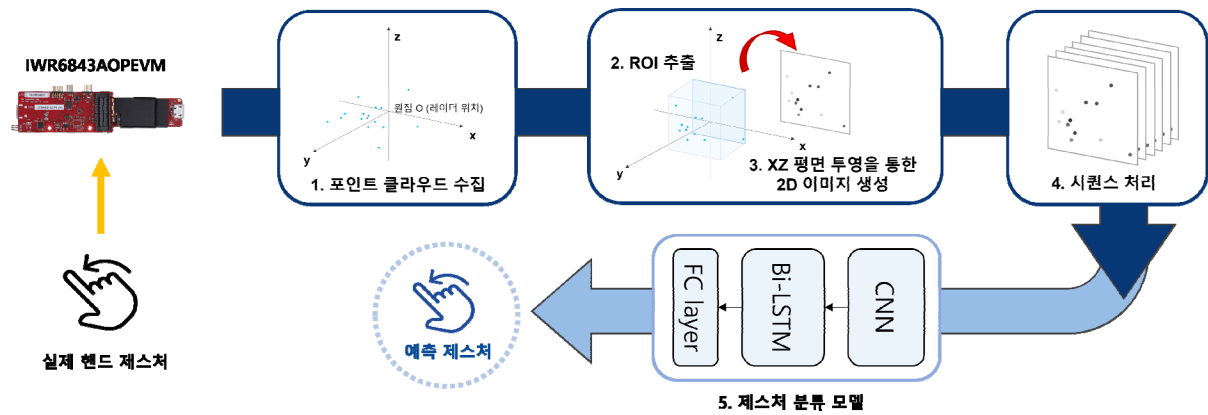


그림 1. 레이더 포인트 클라우드 2차원 투영 이미지 기반 손동작 인식 시스템 구성도

## B. 손동작 인식 및 분류 모델 및 성능 평가

제스처 분류 모델로는 차원 축소된 포인트 클라우드의 특징 벡터 추출을 위해 CNN 과 시계열 특성을 고려한 양방향 LSTM 을 결합한 딥러닝 모델을 활용하였다.

CNN 은 3 개의 Convolution 계층과 Max Pooling 계층을 사용하였고 Convolution 연산에는 3x3 크기의 필터를 32 개 사용하고, pooling 연산에는 2x2 크기의 필터를 적용하였다. CNN 을 통해 데이터의 공간적 특징을 추출하고 나면 각각 16 개의 노드로 구성된 양방향 LSTM 을 거쳐 데이터의 시간적 특징이 함께 추출된다. 이후 완전 연결 계층을 통해 데이터의 시간적 특징을 기반으로 8 개의 제스처 중 하나로 분류한다.

2 명의 참가자가 [그림 2]에 표현된 8 개의 제스처를 수행하여 총 80 분의 데이터를 수집하였고, 학습 데이터와 테스트 데이터는 9 대 1 로 분할하여 사용하였다.

본 연구에서 제안하는 투영 기법과 비교할 시스템은 복셀 기반 동작 인식 시스템이다[1]. ROI 추출 및 투영 과정 대신 복셀화를 수행하는 점 외에는 동일한 조건으로 시스템을 구성하고 학습하여 평가하였다.

두 시스템의 성능 비교 결과는 표 1 과 같다. 성능 지표 중 메모리 사용량은 테스트 데이터를 통한 분류 모델 평가 시 필요로 하는 실행시간에서의 메모리 사용량을 의미한다. 차원 축소된 2D 투영 이미지에 의해 학습파라미터 수가 줄었고 그에 따라 상대적으로 적은 데이터셋에서도 학습이 잘 수행되는 효과를 확인할 수 있었다.

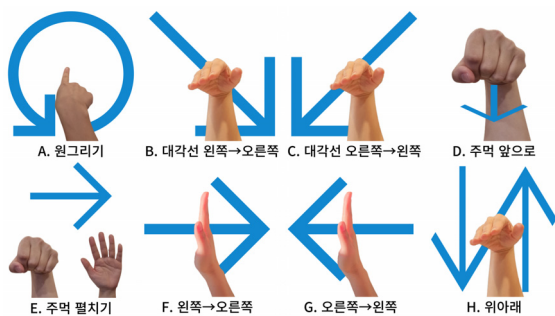


그림 3. 실험에 사용한 8 가지 손동작 종류

[표 1] 복셀 기반 시스템과 투영 기반 시스템의 성능 비교

성능 지표	3D 복셀 [1]	2D 투영이미지 [제안방식]
메모리 사용량	약 4.3 GB	약 2.6 GB
학습 파라미터 수	208,712	147,687
종합 정확도	66.21 %	94.86 %

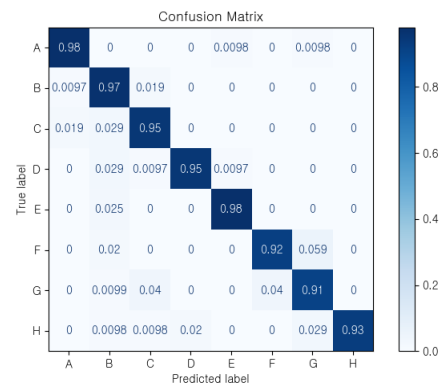


그림 2. 손동작 인식 모델의 분류 정확도 오차 행렬

## III. 결론

본 논문에서는 포인트 클라우드를 2 차원 평면에 투영하여 활용함으로써 시스템 복잡도를 감소시킨 mmWave 레이더 기반 손동작 인식 시스템을 제안하였다. 3D 복셀 데이터를 2D 로 차원축소하여 데이터 처리에 대한 부담을 줄이면서도 포인트 클라우드의 주요 특징을 추출하는 데 성공하였음을 실험을 통해 검증하였다. 본 연구는 실시간 응용에 적합한 저비용 시스템 구현을 목표로 수행되었으며 실제 실시간 환경에서의 구현 및 검증은 추후 연구에서 이루어질 예정이다. 또한 모바일 환경 응용을 위한 추가적인 경량화 및 성능 개선 방법에 대한 연구도 함께 진행할 예정이다.

## ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.

(No. 2021R1F1A10639451221082086810102)

## 참고 문헌

- [1] A. D. Singh, et al, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems.*, pp. 51-56, Oct. 2019.
- [2] G. Lee, J. Kim, "Improving Human Activity Recognition for Sparse Radar Point Clouds: A Graph Neural Network Model with Pre-Trained 3D Human-Joint Coordinates," *Applied Sciences.*, vol. 12, p. 2168, Feb. 2022.