

# 멀티 2D 라이다 센서와 딥러닝을 활용한 실시간 모션인식

\*김민희, \*조청문, \*김준모, \*김선호, \*윤지훈  
\*서울과학기술대학교

\*{kimmu1018, jcm5900, 0904junmong, hokim3031}@seoultech.ac.kr

## Real-Time Motion Recognition Using Multi-2D LiDAR Sensors and Deep Learning

\*Kim Min Ui, \*Cho Cheong Mun, \*Kim Jun Mo, \*Kim Seon Ho, and Ji-Hoon Yun  
\*Seoul National University of Science and Technology (SeoulTech).

### 요약

모션 인식 기술은 인간의 움직임을 디지털 데이터로 변환하여 활용된다. 기존에 카메라 센서 기반 모션 인식은 주변 환경에 취약하고 사생활 침해 및 개인정보 유출의 우려가 있었으며, 3D 라이다 센서는 높은 가격으로 홈환경에 부적합하다. 본 연구는 이를 해결하기 위해 멀티 2D 라이다 센서를 활용한 모션 인식 시스템을 제안한다. 제안된 Transformer+CNN 모델은 9가지 모션을 실시간 96.27%의 정확도와 0.13초의 지연 시간으로 실시간 인식하며, 적응형 ROI 필터링을 적용해 환경 변화에도 성능을 유지한다. 또한, Unity와 연동하여 실시간 모션 인식 게임을 구현하였다.

### 1. 서론

모션인식 기술은 카메라 센서를 기반으로 활발히 연구되었으나, 인간의 신체 및 얼굴 데이터를 수집하면서 사생활 침해와 개인정보 유출의 위험이 있다.[1] 특히, AI 기술이 고도화되면서 개인 초상권 보호의 중요성이 더욱 강조되면서, 라이다 센서는 개인정보 보호가 요구되는 환경에서 유망한 대안으로 주목받고 있다.[2] 라이다 센서는 다양한 조도 환경에서도 일관된 성능을 제공하고 넓은 범위와 먼 거리의 정밀 측정이 가능하지만, 3D 라이다 센서는 높은 가격으로 가정 환경을 포함한 다양한 모션인식 활용에 적용되기 어렵다. 본 연구에서는 저렴한 복수 개의 2D 라이다 센서를 결합해 카메라와 유사한 수직 측정각을 확보하고, 낮은 수직 해상도 문제점을 Transformer와 CNN을 결합 모델로 해결하고자 한다. 제안된 모델은 시간적, 공간적 정보를 효과적으로 학습하여 실시간 예측 정확도 96.27 및 시간 딜레이 0.13초의 성능을 달성하여 사용자에게 불편함 없는 실시간 사용 경험을 제공한다.

II장에서 시스템 구성도, ROI(Region of Interest) 필터 및 설계한 딥러닝 모델, 모델 실험 내용을 소개한다. III장에서는 결론 및 향후 연구 방향을 제시한다.

## II. 본론

### II.1 시스템 구성도

시스템은 크게 라이다 모듈, 모션인식 모델, 그리고 유니티로 구성된다. 사용자가 특정 모션을 취할 때, 라이다 센싱이 활성화되며, USB 통신을 통해 라이다 모듈에서 데이터를 처리하고 Wi-Fi를 통해 데이터를 PC로 전송한다. 이후, ROI 필터링과 데이터 증강을

적용하여 모델 훈련 및 실시간 테스트를 진행한다. 이 과정을 거쳐 모션 인식이 완료되고 인식 결과는 유니티를 사용한 모션인식 게임에 전달된다. 유니티와의 연동 과정에서는 QT 사용자 인터페이스를 통해 데이터 수집부터 실시간 검증까지의 모든 단계를 통합하여 관리한다.

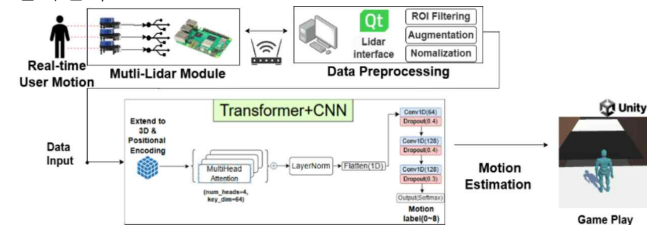


그림1. 시스템 구성도

### II.2 모션 데이터셋 구성

데이터셋 구성을 위해, 세 개의 2D 라이다 센서를 이용하여 사용자의 머리, 허리, 다리를 2D로 센싱하고, 각도에 따른 포인트 클라우드와 모션 라벨을 파일로 저장한다. 최종적으로 9가지 모션을 실시간 인식용 모션으로 선정하였고, 이에 대한 데이터셋을 수집하였다.

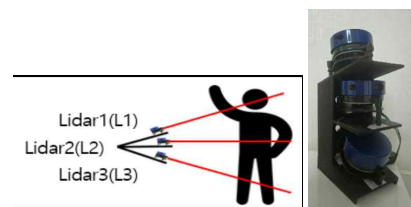


그림2. 멀티 2D 라이다 센싱 모듈 구성 및 구현 모습

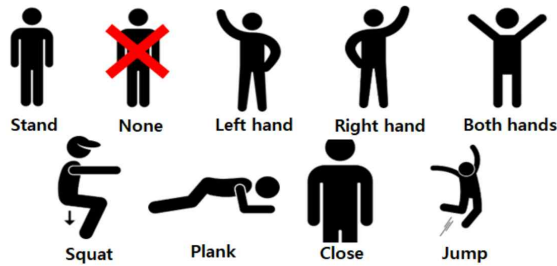


그림3. 9가지 실시간 인식용 모션

### II.3 ROI 필터링

데이터를 가공하지 않고 학습을 진행하였을 때 주변 환경까지 학습되어 모션인식 성능이 저하되는 문제가 발생했다. 이를 해결하기 위해 불필요한 환경적 요소를 제거하고 순수한 모션 데이터만을 남기기 위한 적응형 ROI 필터를 적용했다. 그림3은 필터 적용 전후의 Stand 모션 데이터셋에서 저장된 색조를 보여주며, 필터 적용으로 환경적 요소인 빨간 박스 부분이 제거된 것을 보여준다.

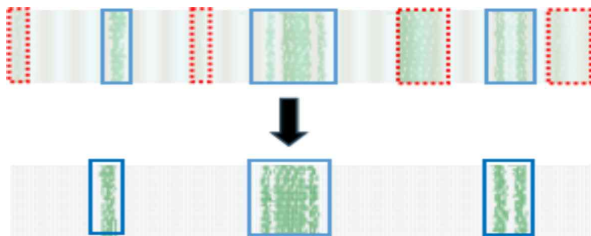


그림4. Stand 자세 데이터셋에 ROI 필터링 전후 비교

ROI 필터링이 모델 정확도에 미치는 영향을 파악하기 위해 최적화된 모델 중 CNN을 사용하여 Stand 자세의 정확도 향상을 측정한 결과, 88%에서 95%로 정확도 향상을 보였고 이로써 ROI 필터링의 효과를 확인하였다.

### II.4 Transformer+CNN 모션인식 모델

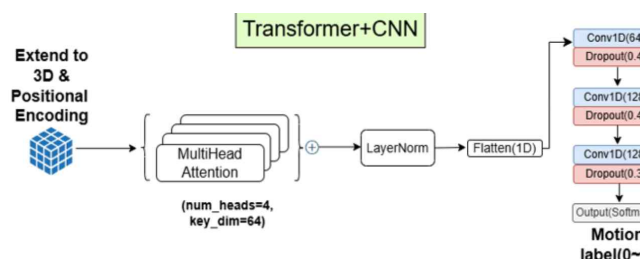


그림5. Transformer+CNN 구조

Transformer+CNN 구조는 메모리와 파라미터 용량을 유지하면서도 높은 실시간 정확도와 최소 시간 지연을 제공하도록 설계되었다. Attention 메커니즘과 1D CNN 레이어를 결합하여 입력 데이터의 행 간 특징 관계를

선택적으로 집중하고 장기 의존성과 로컬 특징을 효과적으로 처리함으로써 불필요한 메모리 사용을 줄였다. 이러한 하이브리드 모델[3]을 사용하여 9개 모션 클래스를 정밀히 구분하며, 높은 정확도와 최소 시간 지연을 달성했다.

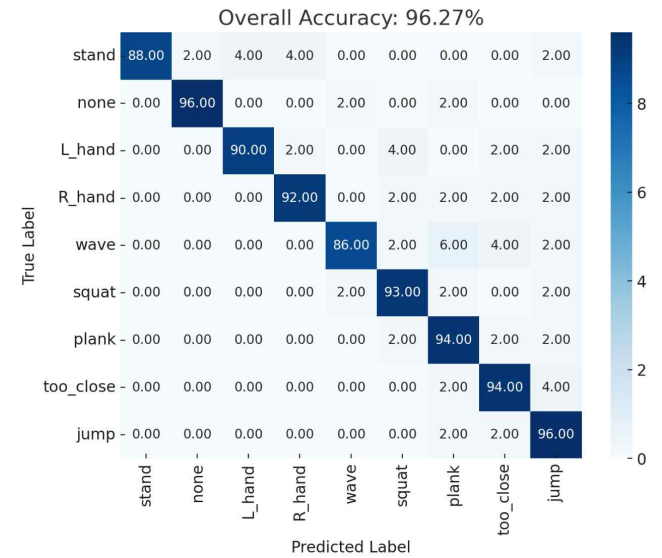


그림6. 9개 모션 클래스에 대한 혼동 행렬

그림6은 Transformer+CNN의 정확한 모션 예측 성능을 보여준다. 각 모션 별 정밀도 및 재현율을 따져보아도 각각 최대 오차 8% (L\_hand, wave), 12% (stand)를 기록할 뿐 모든 모션에 대한 평균 지표 98%를 상회한다.

### II.5 정적 모델 실험

모델 학습 단계에서는 실시간 검증 전, 모델의 성능과 신뢰성 확보를 위해 6개의 AI 모델에 학습을 진행했다. 실험 데이터는 총 4명의 데이터로 구성되고, 3명의 데이터를 통해 학습하고 다른 1명의 데이터로 예측했다. 학습에 사용된 6개 모델 및 정확도는 표1과 같다.

	Accuracy	Precision	Recall	F1-score
MLP	80.84%	78.76%	78.05%	78.40%
LSTM	92.17%	89.94%	89.23%	89.58%
CNN	95.47%	93.54%	93.52%	93.53%
Trans	95.33%	92.15%	92.04%	92.09%
Trans+MLP	94.00%	92.34%	91.84%	92.09%
Trans+CNN	98.88%	98.30%	98.31%	98.31%

표1. 6가지 정적 모델의 성능지표 비교

95% 이상의 정확도 및 90% 이상의 정밀도, 재현율, F1-score를 모델 선정 기준[4]으로 정하였고 이에 따라 CNN, Transformer, Transformer+CNN을 실시간 검증을 위한 모델로 선정했다.

## II.6 실시간 모델 실험

위 3개 모델들의 정확도와 시간 딜레이를 측정하여 실시간 모델 성능을 비교했다. 본 실험에서는 모델 출력을 실시간으로 추적하여 누적 프레임을 평가하는 실시간 테스트 방식을 고안했다. 그림7은 누적 프레임 평가법을 시행하는 모습을 보여준다.

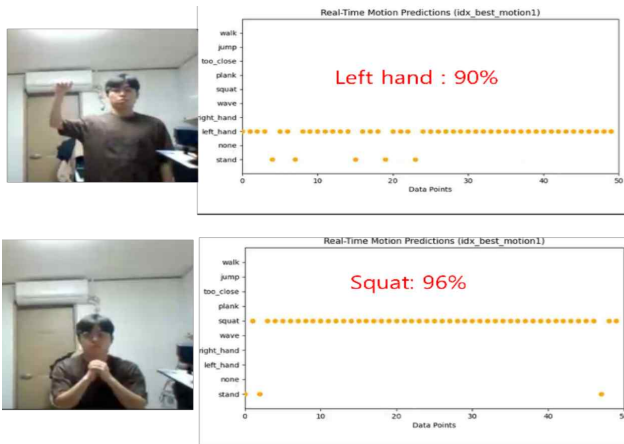


그림7. 누적 프레임 평가를 수행하는 모습과 모션을 실시간으로 예측하는 시계열 플롯

시계열 플롯에서 예측과 다른 이상치들을 오류 값으로 측정하고 이를 실시간 정확도에 반영한다. 총 9개의 모션에 대해서 평균 96%의 실시간 정확도를 확인하였다 이러한 평가 방법을 바탕으로, Transformer+CNN 모델의 실시간 성능이 가장 뛰어난 것을 확인했다 그림8은 3가지 실시간 모델의 정확도와 시간 딜레이를 보여주며, 그래프 중간의 점선은 정확도와 시간 딜레이 목표 스펙[5]을 나타낸다.

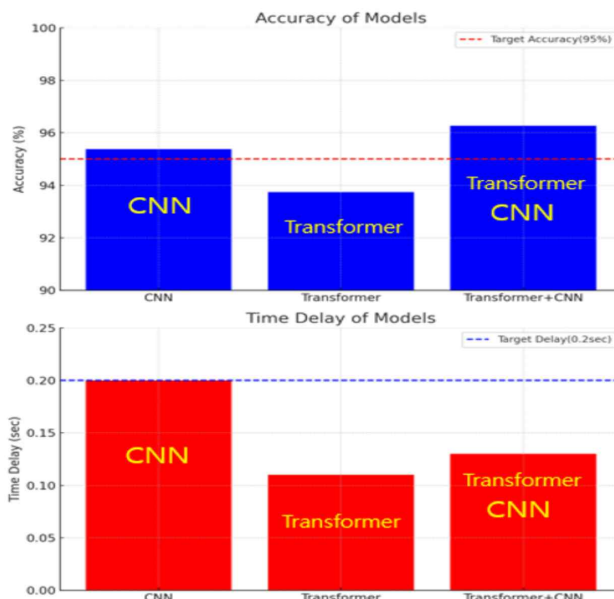


그림8. 실시간 검증 모델의 정확도 및 딜레이

Transformer+CNN 모델은 96.27%의 가장 높은 정확도 및 0.13sec의 타겟 범위 내의 시간 딜레이 스펙을 보인다. 이것으로 실시간 성능 우수성을 입증했다

## III. 결론

본 연구에서는 멀티 2D 라이다 센서를 활용하여 새로운 액티비티 컨트롤러를 개발하였다. 이로써 9가지의 높은 정확도를 가진 모션 인식이 가능한 멀티 라이다 시스템을 구현했다. 라이다 모듈의 통합 관리와 무선 통신 기능 및 모델부에서 적응형 ROI 필터링과 최적화된 Transformer+ CNN 모델 설계를 통해 실시간 정확도 96.27% 및 시간 딜레이 0.13초를 달성했다. 또한, 유니티를 라이다 모듈과 연동하여 모션 인식 게임 플레이가 가능한 생동감 넘치는 애니메이션을 구현했다.

## 참 고 문 헌

- [1] Javed, Mohammed Hamza. "Internet of Things Hacking: Ethical Hacking of a Smart Camera." EngRxiv, pp. 13-14, 2023. (<https://doi.org/10.31224/3079>)
- [2] Gómez, Juan, Olivier Aycard, and Junaid Baber. "Efficient Detection and Tracking of Human Using 3D LiDAR Sensor." Sensors, vol. 23, no. 10, p. 4720, 2023. (<https://doi.org/10.3390/s23104720>)
- [3] Park, Yeonju, et al., "Behavior Recognition Using Multi Sensor Data Based on Transformer," Proceedings of the Korean Institute of Electrical Engineers Conference, pp. 351-355, Oct. 2021.
- [4] Ordóñez, Francisco Javier, and Daniel Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," Sensors, vol. 16, no. 1, p. 115, Jan. 2016. (<https://doi.org/10.3390/s16010115>)
- [5] MILLER, Robert B. "Response time in man-computer conversational transactions," Proceedings of the fall joint computer conference, pp. 267-277, Dec.1968. (<https://doi.org/10.1145/1476589.1476628>)