

# WiFi CSI 를 이용한 잠재적 위치 추론

오궁요, 홍지연, 유재현\*

성신여자대학교 AI 융합학부 학부생,  
\*성신여자대학교 AI 융합학부 교수

[20231383@sungshin.ac.kr](mailto:20231383@sungshin.ac.kr), [20241283@sungshin.ac.kr](mailto:20241283@sungshin.ac.kr), \*[jhyoo@sungshin.ac.kr](mailto:jhyoo@sungshin.ac.kr)

## Position Estimation Using WiFi CSI

Oh Geung Yo, Hong Ji Yeon, Yoo Jae Hyun\*  
School of AI convergence, Sungshin Women's University

### 요 약

본 논문은 WiFi CSI(Channel State Information) 신호를 활용하여 사람의 skeleton 좌표와 위치 구간을 동시에 추정하는 멀티태스킹 모델을 제안한다. 해당 모델은 단일 송수신기로 수집한 CSI 신호를 encoder 로 압축하여 4 개의 네트워크에 전달해 각기 다른 태스크를 수행한다. 실험 결과, 전체적인 성능에는 한계가 있었으나 일부 데이터셋의 특정 구간에서 skeleton 좌표가 적절한 형태로 재현되었으며, 구간 분류에서도 높은 recall 이 나타나는 구간이 확인되었다. 이를 통해 기존의 센서 방식의 대안으로 단일 송수신기의 CSI 신호가 구간 구분에 활용 가능한 특징을 포함하고 있음을 시사하며, 향후 데이터셋의 확장과 일반화를 통한 발전 가능성을 제시한다.

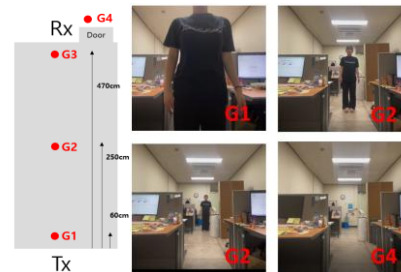
### I. 서 론

최근 컴퓨터 비전 분야의 발전으로 RGB 카메라, LiDAR, 레이더 센서를 활용한 사람 자세 추정 연구가 활발히 이루어지고 있다. 그러나 카메라 기반 방식은 프라이버시 침해 우려가 있으며, LiDAR 와 레이더 기반 기술은 장비 비용과 전력 소모가 크다는 한계가 있다. 이러한 문제를 보완하기 위해, 최근에는 WiFi 신호에서 얻을 수 있는 CSI(Channel State Information, 채널 상태 정보)를 활용한 위치 추정 연구가 주목받고 있다. 기존 연구들은 주로 다수의 송수신기를 사용하여 풍부한 정보를 확보함으로써 성능을 높였지만, 본 연구는 단일 송수신기만으로도 효과적인 자세 추정과 위치 추정이 가능함을 보이고자 한다. 또한 기존 연구들은 주로 동일한 데이터셋을 훈련과 테스트로 분할하여 성능을 평가하였기 때문에, 실제 환경에서의 일반화 성능을 충분히 확인하기에는 한계가 있었다. 이에 본 연구는 학습에 포함되지 않은 별도의 데이터를 이용해 모델의 일반화 가능성을 평가하였다. 따라서 본 연구는 WiFi CSI 신호를 기반으로 실내 위치 구간을 추정하고, 동시에 사람의 skeleton 좌표를 회귀적으로 예측해 시각화할 수 있는 멀티태스킹 모델을 제안한다.

### II. 본 론

#### 2.1. CSI 데이터 수집/전처리 및 분석

본 실험에서 CSI 데이터는 송신기(Tx) 1 대와 수신기(Rx) 1 대를 이용하여 수집하였다. 송신기는 ESP32-S3 보드를, 수신기는 ipTIME A2004SR 을 사용하였다. 송수신 간 통신은 ESP-NOW 프로토콜을 기반으로 수행되었으며, ESP-IDF 환경에서 서버를 구축하여 데이터를 수집하고 CSV 파일 형식으로 저장하였다. 실험은 총 4 개의 구간에서 진행되었으며, 그림 1 과 같이 송신기로부터의 거리에 따라 근거리(G1), 중간 거리(G2), 원거리(G3)로 나누고, 사람이 존재하지



[그림 1] CSI 신호 데이터 수집 환경

않는 구간을 G4 로 정의하였다. 측정 시간 10 분 동안 3 번 이동하며 각 4 분 55 초의 CSI 데이터를 수집하였고, 데이터는 CSV 파일로 서버에 저장되었다. 구간 이동 시 발생하는 시간(5 초)는 전처리 과정을 통해 제외하고, 학습에 활용하기 위해 구간별로 CSI 데이터의 모든 패킷에 label 정보를 추가하였다. 수집한 CSI 데이터는 진폭(amplitude)과 위상(phase) 정보를 담고 있는 총 128 개의 서브캐리어로 구성되어 있으며, 이 중 시작 지점에서 반복적으로 나타나는 22 개의 0 값 구간은 유의미한 채널 응답을 나타내지 않는다고 판단하여 제거하였다. 따라서 총 106 개의 서브캐리어를 학습에 활용하였으며, 초당 약 3~4 패킷의 속도로 수집하였다. test 데이터는 학습 데이터와는 별도로 여러 세트 수집하였으며, 동일한 환경과 방식으로 구간별 1 분 55 초 동안의 CSI 데이터를 확보하였다. 수집 결과, CSI 데이터는 사람의 자세 변화나 주변 기기의 간섭 등 환경적 요인에 따라 민감하게 변동하는 특성이 관찰되었다.

#### 2.2 skeleton 데이터 수집 및 전처리

본 실험에서는 4 개 구간에 대한 영상을 촬영한 후, OpenPose 라이브러리를 활용하여 각 구간의 skeleton

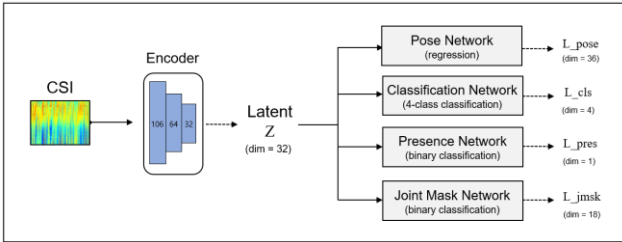


[그림 2] skeleton 관절 좌표 추출 예시(G1/G3)

관절 좌표를 JSON 형식으로 추출하였다. 관절 개수는 OpenPose 의 COCO 18 모델 규칙에 따라 총 18 개로 정의하였으며, 각 프레임은 x, y 좌표로 구성된 (18, 2) 형태로 저장하였다. 사람이 없는 G4 구간은 모든 좌표를 (0, 0)으로 통일하였다. 이후 데이터셋 수 통일을 위해 프레임 개수를 조정하는 과정을 거친 후, 일괄적으로 NPY 배열로 변환하여 학습 데이터로 사용하였다.

### 2.3 학습 방법

본 연구에서는 CSI 데이터를 입력으로 skeleton 좌표와 구간(class)을 동시에 학습하는 멀티태스킹 신경망(Multi-task Neural Network)을 사용한다. 전체 데이터셋은 학습(80%)과 검증(20%)으로 분할하였으며, 동일 환경에서 별도로 수집한 test 데이터를 실험에 활용하였다.



[그림 3] 멀티태스킹 모델 구조

그림 3 과 같이 해당 모델은 Encoder 와 4 개의 출력 Head 로 구성된다. Encoder 에서 106 차원의 CSI 데이터를 입력으로 받아 32 차원의 latent vector 로 압축한 후, 각 네트워크의 입력으로 사용된다. 4 개의 네트워크는 각각 Pose Network, Classification Network, Presence Network, Joint Mask Network 이며, 각 Network 들은 서로 다른 태스크를 수행한다. Pose Network 는 G1~G4 구간에 따른 skeleton 좌표(18 joints, 36 차원)를 MSE loss 를 사용하여 회귀 방식으로 추정한다. Classification Network 는 CSI 입력에 해당하는 구간(G1~G4) label 을 학습에 사용해 Cross-Entropy(CE, 교차 엔트로피) loss 를 최소화한다. Presence Network 는 프레임에 사람이 존재하는지 여부를 sigmoid 출력을 통해 학습하며, Binary Cross Entropy(BCE, 이진 교차 엔트로피) loss 를 사용한다. 마지막으로 Joint Mask Network 는 화면에 가려지는 관절 좌표 처리를 담당하며, 각 관절의 가시 여부를 예측하여 skeleton 추정 시 불확실한 관절을 구분하며, BCE loss 를 최소화한다. 손실 함수는 다음 4 개의 항으로 구성된다.

$$L = W_{pose}L_{pose} + W_{cls}L_{cls} + W_{pres}L_{pres} + W_{jmsk}L_{jmsk}$$

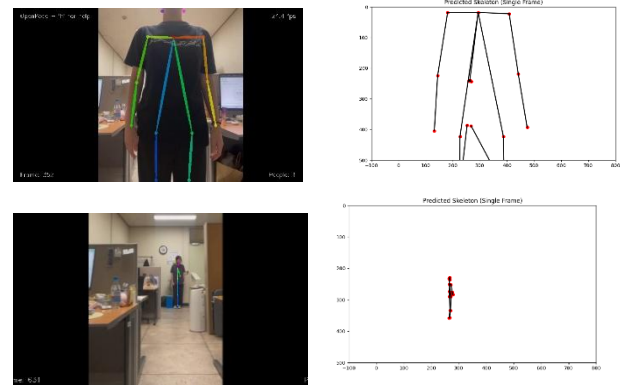
가중치는 여러 차례 학습을 반복하면서 validation 성능을 기준으로 실험적으로 조정하였다. 최종적으로

$$W_{pose} = 1.0, W_{cls} = 1.2, W_{pres} = 0.7, W_{jmsk} = 0.3$$

으로 설정하였으며, 이는 분류 정확도를 강조하면서 presence 및 joint mask 학습은 상대적으로 완화하는 방향으로 설계한 결과이다.

### 2.4 실험 결과 및 논의

본 연구에서는 앞서 언급한 여러 개의 test 데이터셋을 별도로 수집하여 모델을 평가하였다. 구간 분류 성능은 recall 을 주요 평가 지표로 활용하여 정량적으로 검증하였으며, skeleton 좌표 예측은 그림 4 와 같이 시각화를 통해 분석하였다. 실험 결과, accuracy 기준의 전체 성능은 높지 않았다. 이러한 결과는 초당 CSI 패킷 수의 부족, 자세 및 주변 환경 변화와 같은 요인으로 인해 데이터셋 간 편차가 발생하며 일반화 성능이 제한적이었음을 보여준다. 그러나 구간 분류 성능을 recall 지표로 분석한 결과, 일부 데이터셋에서 특정 구간에 대한 높은 성능이 확인되었다. 예를 들어, 한 데이터셋에서는 G1 구간의 recall 이 93.5% 였으며, 또 다른 데이터셋에서는 G4 구간의 recall 이 98.7%로 나타났다. 이는 서로 다른 데이터셋에서도 특정 구간을 안정적으로 분류함을 보여주며, CSI 신호가 구간 구분에 활용 가능한 특징을 내포하고 있음을 시사한다.



[그림 4] skeleton 예측 좌표 시각화 예시(위: G1, 아래: G3)

### III. 결론

본 연구는 기존 카메라·LiDAR·레이더 기반 방식의 대안으로, WiFi CSI 신호만을 활용하여 사람의 위치 구간과 skeleton 좌표를 동시에 추정하는 멀티태스킹 모델을 제안하였다. 데이터셋의 편차와 환경적 다양성 부족으로 인해 일반화 성능은 제한적이었으나, 특정 구간에서 의미 있는 분류 성능이 확인되었다는 점에서 CSI 신호가 구간 구분에 활용 가능한 특징을 포함함을 보여주었다. 따라서 본 연구는 단일 송수신기의 CSI 데이터만으로 유의미한 분류 성능을 확보할 수 있음을 시사하며, 향후 데이터 확장과 환경 다양화를 통한 일반화 성능 개선 연구의 기반을 마련한다.

### ACKNOWLEDGMENT

이 논문은 경찰청이 지원한 ‘사회적 약자 보호 강화 기술 개발(www.kipot.or.kr)’의 지원을 받아 수행된 연구결과입니다.

[과제명: 저전력 복합측위, 근접탐색 기술 기반 범죄피해 안전조치 대상자 위치추적 통합관제 플랫폼 개발/과제번호: RS-2023-00236101]

### 참 고 문 헌

[1] Deng, J.; Chen, K.; Jing, P.; Dong, G.; Yang, M.; Zhu, A.; Li, Y. CSI-Channel Spatial Decomposition for WiFi-Based Human Pose Estimation. *Electronics* 2025, 14, 756.