

실생활 환경에서 다중 IR-UWB 레이더를 이용한 Transformer 기반 아동 인원수 추정 연구

송영우, 류경필, 이경우*
연세대학교

yeongwoo.song@yonsei.ac.kr, kp.ryoo@yonsei.ac.kr, *kyoungwoo.lee@yonsei.ac.kr

Transformer-based Children Counting in Real-World using Multiple IR-UWB Radars

Song Yeong Woo, Ryoo Kyung Phil, Lee Kyoung Woo*
Yonsei University

요약

현실 환경에서 센서 기술이 발전함에 따라 실시간으로 사람의 유무를 판단하고 사람의 수를 판단하는 기술이 중요해지고 있다. 특히, 프라이버시 보호가 요구되는 환경에서 카메라를 대체할 수 있는 비식별화 센서 기술의 수요가 증가하고 있다. 하지만 많은 연구에서 아동을 대상으로 인원을 계수한 연구는 부족한 상황이다. 본 연구에서는 장애물이 많은 현실 세계에서 IR-UWB 를 이용하여 아동의 수를 Transformer 모델을 이용하여 0 명/1 명/2 명 이상 혹은 0~4 명으로 계수하는 모델을 구현하였다.

I. 서 론

우리나라의 보육교사 법적 1 인당 아동 수는 20 명으로 OECD 평균을 웃돌아[1] 안전사고에 즉각적인 조치가 어려운 상황이다. 특히 교사가 인지하기 어려운 화장실이나 탈의실과 같은 사각지대에서 일어난 사고는 교사가 인지하기 더욱 어렵다. 따라서 사각지대에 있는 아동 인원수 정보는 교사의 아동에 대한 상황인식을 높이는 필수 정보 중 하나이다. 교사가 이를 활용하면 업무강도를 줄이고 안전사고에 대해 보다 적극적인 조치를 할 수 있다.

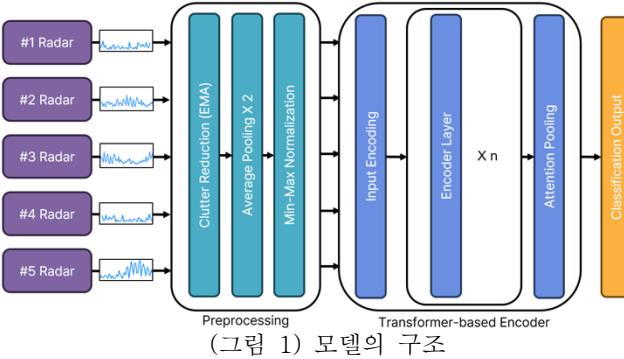
인원수를 추정하는 연구는 꾸준히 진행되었고 카메라, LIDAR 등 다양한 센서들을 활용해서 발전됐다. 이들은 인원수를 즉각적으로 파악할 수 있는 장점이 있으나, 카메라를 통한 비전 시스템은 프라이버시 문제에 자유롭지 못하고 LIDAR 는 차폐에 취약하여 구조물이 있는 공간에서는 활용이 제한된다[2]. 반면에 IR-UWB 레이더는 비식별화된 신호를 사용하면서 장애물 투과에 뛰어나, 장애물과 칸막이가 있는 공간에서도 프라이버시 문제없이 인원수를 추정할 수 있다.

본 연구에서는 사각지대에 있는 아동 인원수 정보를 추정하기 위해 다중 IR-UWB 레이더를 사용한 Transformer 기반의 딥러닝 모델을 제안한다. 이를 위해 실제 어린이집 화장실에서 측정된 데이터를 활용해 아동 인원수를 추정하는 모델을 학습 및 평가하였다.

II. 본론

IR-UWB 레이더를 활용하여 인원수를 추정하는 연구는 꾸준히 진행됐으며, 이들은 대부분 넓고 트인 공간에서 단일 레이더를 활용하였다[3, 4, 5]. 이는 IR-UWB 레이더 신호가 투과하는 성질을 가졌지만, 좁고

구조물이 있는 공간에서는 사람으로부터의 반사 신호가 다중 경로로 수신되는 Multipath 가 일부 발생하는 단점이 있기 때문이다. 이러한 Multipath 의 문제를 보완하고자 본 연구의 선행 연구인 Park et al.[6]의 연구에서는 여러 대의 IR-UWB 레이더를 사용한 딥러닝 모델을 사용하였다. 하지만 CNN 구조의 딥러닝 모델에서 다중 IR-UWB 레이더 신호를 활용하기 위해 레이더 데이터를 1 차원으로 통합하였고, 이 과정에서 사용된 레이더의 순서에 의존적인 문제와 레이더 간 경계에 대한 부적절한 학습의 한계를 지니고 있다. 따라서 본 연구에서는 제안하는 모델은 레이더 간의 상관관계를 통해 인원수 추정을 위해 Transformer 의 Encoder 구조를 활용하였다[7]. 이는 프레임마다 각 레이더 신호를 하나씩 Input sequence 로 입력하는 것으로 Encoder 의 Multi-head Attention 이 레이더끼리의 연산으로부터 인원수에 대한 Representation 학습하도록 하였다. 이후 Attention Pooling 및 MLP layer 를 통해 인원수에 대해 classification 을하도록 하였다. 기존 Transformer 와는 달리 레이더 순서 정보가 포함되지 않은 순수한 신호 정보만을 학습하고자 Positional Encoding 을 적용하지 않았다. 자세한 Transformer 구조는 그림 1 과 같다. 제안하는 모델 역시 기존의 IR-UWB 레이더를 활용한 인원수 추정 연구들과 마찬가지로 레이더 신호에 대해 전처리 과정을 거쳐 활용하였다. 전처리 과정은 3 단계로, 배경 잡음 제거, 차원 축소, 정규화했다. 대표적인 배경 잡음 제거 기법인 EMA(Exponential Moving Average)기법[8]을 통해 움직이는 아동의 반사 신호를 추출하였다. 이후 레이더 신호의 거리 차원을 Average pooling layer 를 두 번 거쳐서 평균값을 활용하여 단발성 잡음을 제거하고 Attention 연산의 효율을 위해



차원을 축소하였다. 이후 Min-Max 정규화를 통해 데이터값의 범위를 0-1로 재정렬하였다. 전처리 단계는 각 레이더에 모두 적용되어 각각이 Encoder에 Input 되었다.

본 실험에서는 선행 연구[6]의 실험 데이터를 전부인 어린이집 화장실에 설치된 5 대의 IR-UWB 레이더 데이터와 인원수 정보를 활용하여, 제안하는 아동 인원수 추정 모델에 대해 두 가지 실험을 하였다. 첫 번째 실험은 0 명/1 명/2 명 이상의 3 Class 를 분류하는 사전 실험이고, 두 번째 실험은 실전 성능을 확인하기 위한 0 명부터 4 명까지의 5 Class 를 분류하는 실험을 진행하였다. 실험마다 선행 연구의 결과와 비교하여 제안하는 모델을 평가하였다.

학습 데이터 세트는 제주 명지 어린이집에서 31 명의 5 살 아동을 대상으로 23 년 5 월 8 일부터 6 월 9 일까지 얻은 데이터를 기반으로 하였다. 아동 인원수별 데이터의 비율이 1 명 데이터 쪽에 쏠려있어, 균등한 학습을 위해 가장 데이터양이 적은 4 명 데이터 수에 맞춰 Undersampling 하여 실험하였다. 총 실험에는 116,535 개의 데이터를 사용하였으며, 8:1:1로 데이터 세트를 나누어 학습 및 검증, 평가하였다. 평가 데이터 세트에서의 Accuracy(전체/class 별)와 F1 score 를 통해 학습된 모델의 성능을 평가하였다.

실험에 사용된 모델의 주요 파라미터 및 학습 하이퍼파라미터는 다음과 같다. Input dimension: 64, Multi-head 수: 8 개, Encoder layer 수: 24 개, Adam Optimizer, learning rate: 5e-4, epoch: 150 회, dropout : 0.1, FFN dimension : 512.

첫 번째 실험의 결과는 전체 accuracy 93.56%, F1 score 0.97 의 성능을 나타냈다. 선행 연구의 전체 accuracy 가 96.00%, F1 score 0.99 으로 유사하지만 다소 저하된 성능을 보인다고 할 수 있다. 각 class 별 인원수 예측 결과는 표 1 의 혼동행렬과 같다.

두 번째 실험의 결과는 전체 accuracy 82.73%, F1 score 0.83 의 성능을 나타냈다. 선행 연구의 전체 accuracy 가 74.40%, F1 score 0.73 으로, 본 연구에서 제안하는 모델이 전체적인 성능에서 크게 향상된 것을 확인하였다. 특히 표 2 의 각 class 별 예측 결과의 혼동행렬은 3 명 분류에서 77.86%의 정확도를 보여 이전 연구 대비 44.61% 향상되는 것을 확인할 수 있다.

III. 결론

보육교사의 관리·감독이 힘든 사각지대 영역에서 아동 인원수 정보는 아동의 안전을 위해 활용될 수 있다. 본 연구는 어린이집 화장실에서의 아동 인원수를 CNN 으로 추정한 선행 연구를 토대로 Transformer 기반의 구조를 사용함으로 82.73%의 정확도를 얻는 모델을 학습하였다. 이를 통해 실제 어린이집 안전 모니터링 시스템에서 활용될 수 있는 기반을 제시하였다.

(표 1) 실험 1 의 혼동행렬 결과

예측 Class 실제 Class \	0 명	1 명	2 명 이상
0 명	12,854	41	152
1 명	886	11,685	476
2 명 이상	208	757	12,082

(표 2) 실험 2 의 혼동행렬 결과

예측 Class 실제 Class \	0 명	1 명	2 명	3 명	4 명
0 명	2,253	5	16	15	42
1 명	166	1,907	175	66	17
2 명	46	273	1,606	285	120
3 명	8	43	232	1,815	233
4 명	7	12	88	164	2,060

ACKNOWLEDGMENT

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행한 연구임(No.2021-0-00155, 유아상황 및 행동분석기반 안전보육환경 문제해결기술개발)

참 고 문 헌

- [1] Organisation for Economic Co-operation and Development (OECD), "Students per teaching staff, by type of institution (indicator)", <https://data-explorer.oecd.org/>. [Accessed: Oct. 10, 2025].
- [2] T. Suleymanov, M. Gadd, L. Kunze and P. Newman, "LiDAR Lateral Localisation Despite Challenging Occlusion from Traffic," *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, Portland, OR, USA, 2020, pp. 334-341, doi: 10.1109/PLANS46316.2020.9109911.
- [3] J. Choi, J. Kim, N. Jeong, K. Kim and S. Jin, "Accurate People Counting Based on Radar: Deep Learning Approach," *2020 IEEE Radar Conference (RadarConf20)*, Florence, Italy, 2020, pp. 1-5, doi: 10.1109/RadarConf2043947.2020.9266496.
- [4] J. Choi, J. Kim and K. Kim, "People Counting Using IR-UWB Radar Sensor in a Wide Area," in *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5806-5821, 1 April 2021, doi: 10.1109/IoT.2020.3032710.
- [5] C. Pham, V. Luong, D. Nguyen, H. Vu, and M. Le, "Convolutional neural network for people counting using UWB impulse radar," *Journal of Instrumentation*, Aug. 2021, doi: 10.1088/1748-0221/16/08/P08031.
- [6] A. Park, K. Ryoo, S. Lee, S. Park, W. Lee and K. Lee, "CNN-Based Children Counting in Real-World Using Multiple IR-UWB Radars," *2024 IEEE Radar Conference (RadarConf24)*, Denver, CO, USA, 2024, pp. 1-6, doi: 10.1109/RadarConf2458775.2024.10549702.
- [7] A. Vaswani et al. "Attention Is All You Need," arXiv.org, Jun. 12, 2017, doi: 10.48550/arXiv.1706.03762.
- [8] L. Yang, X. Lai, X. He, H. Zhu, Z. Kan, and X. Song, "A Newton-Cotes-based online acceleration signal fast processing approach to obtain displacement for digital twins," *Mechanical Systems and Signal Processing*, vol. 242, p. 116129, Jan 2025, doi: 10.1016/j.measurement.2024.116129.