

딥러닝을 이용한 다양한 보행 상황에서의 스켈레톤 영상 기반 성별 분류 연구

김다은*, 정아현**, 정승민**, 하니파**, 민세동* **

순천향대학교 의료IT공학과*, 순천향대학교 소프트웨어융합학과**

Gender Classification from Skeleton Videos in Various Walking Scenarios Using Deep Learning

Daeun Kim*, A Hyun Jung**, Seungmin Jeong**, Fajrin Hanifah Rahmi**, Se Dong Min*, **

Department of Medical IT Engineering, SoonChunHyang University*,

Department of Software Convergence, SoonChunHyang University**

Abstract - 보행은 인간 활동의 핵심적인 행동으로, 성격, 감정, 성별 등과 같은 인간의 특성을 반영한다. 특히 보행 시 성별 분석은 HCI(인간-컴퓨터 상호 작용), 보안 등 다양한 분야에서 중요하게 활용될 수 있지만, 현재 다양한 보행 상황을 반영한 연구는 부족하다. 본 연구에서는 PsyMo Dataset을 활용하여 4 가지 보행 상황(전체, 스마트폰 사용, 신체 일부 가려짐, 속도 변화)에서 보행 시 스켈레톤 영상을 추출하고, 이를 ResNet-18 모델을 사용하여 분석하였다.

인공지능 모델을 통해 보행의 1주기마다 2D 스켈레톤 이미지를 분석하여 성별을 분류하였다. 분석 결과는 정확도, 정밀도, 재현율, F1 점수를 통해 확인하였으며, 7가지 보행 상황의 통합 데이터셋에서 각각 80.1%, 80.4%, 80.1%, 80%의 성능을 보였다. 또한, 신체 일부가 가려진 경우, 속도가 다른 경우, 스마트폰을 사용한 경우에는 각각 83.9%, 77.2%, 80.6%의 정확도를 보였다. 이를 통해 대부분의 보행 상황에서 80% 이상의 일관된 성능을 유지하며, 실생활에서 일반적으로 발생하는 4가지 보행 상황에서의 성별 분류 가능성을 확인할 수 있었다. 향후 더 다양한 상황에 대한 보행 데이터를 확보하여 신뢰성 높은 성별 분류 모델을 구축하고, 모델의 성능개선을 통해 성별 분류 정확도를 향상시켜 범용성이 높은 실생활 보행 기반 성별 분류 연구에 기여할 수 있을 것으로 기대한다.

1. 서 론

보행은 일상생활 중 가장 흔히 발생하는 행동으로, 신체의 여러 관절과 근육의 움직임으로 구성되어 개인 식별 수단으로 활용되는 핵심적인 활동이다[1]. 최근 보행 분석을 통해 성격, 성별, 감정 등 인간의 특성을 분류하는 연구가 활발히 진행되고 있으며[1][2][3], 특히 성별 분석은 HCI(인간-컴퓨터 상호 작용), 보안, 맞춤형 광고 제공, 실종자 추정 등에서 중요하게 활용되고 있다[1][4].

초기의 성별 분류 연구는 얼굴 특징, 머리카락, 의복을 기반으로 하지만[5], 기술 발전에 따라 보행 시 관절의 움직임을 통한 성별 분류 연구가 진행되고 있다. Kinect 센서를 사용한 연구에 따르면, 보행 시 신체의 일부만 인식되는 상황에서도 9개의 관절을 사용하여 97%의 정확도로 성별을 분류할 수 있었다[1]. 그러나 이러한 접근은 높은 분류 정확도를 보이지만 다양한 보행 상황을 반영하지 못하고, 측정 장비가 필요하다는 한계점이 있다. 일반적인 성별 분류 연구는 정적 자세, 일정한 속도, 특정 질병 등의 한정된 상황에 집중한다[3][6]. 이러한 접근은 다양한 보행 상황을 반영하지 않기 때문에 일상생활에서 보행 시 성별 분류가 가능하다고 보기 어렵다. 특히 보행 중 스마트폰 사용이 전 세계적으로 증가하고 있지만, 이 상황에서 성별을 분류한 기초 연구는 부족하다[2].

본 연구에서는 Kinect 센서 등의 장비를 사용하지 않고, 인공지능 기반으로 추출된 관절 좌표 이미지를 사용하여 성별을 분류한다. 또한, 일상생활에서 발생하는 4가지 상황(전체, 스마트폰 사용, 신체 일부 가려짐, 속도 변화)에서의 분류 정확도를 비교하고자 한다.

2. 본 론

2.1 데이터셋

본 연구에서는 심리적 특성과 보행 분석을 위해 구축된 PsyMo 데이터셋을 사용한다. 데이터는 <표 1>의 7가지 보행 방법에 대해 6가지 카메라 각도로 수집되었으며, 총 312명(남 199명, 여 113명)의 보행 영상 기반 실루엣, 스켈레톤, SMPL 데이터가 포함되어 있다[7].

<표 1> 7가지 보행 방법

| 약어 | 의미 | 약어 | 의미 |
|-----|--------------------|-----|-----------------------|
| NM | Normal walking | WSF | Walking speed fast |
| CL | Changing clothing | TXT | Walking while texting |
| BG | Carry bag | PH | Talking on the phone |
| WSS | Walking speed slow | | |

<표 1>의 7가지 보행 방법을 비슷한 특징을 기준으로 4가지 상황으로 나누어 데이터셋을 구성하였다.

보행 영상 기반 데이터 중 2D 스켈레톤 데이터와 성별 데이터를 사용하였으며, 성별 분류에 적합한 측면(90°) 각도에서 촬영된 데이터를 사용하였다. 2D 스켈레톤 데이터는 <그림 1>과 같이 COCO pose 형식의 17개의 관절의 x, y 좌표로 구성되어 있으며, 한 사람당 평균 5-6개의 보행 주기가 포함되어 있다.

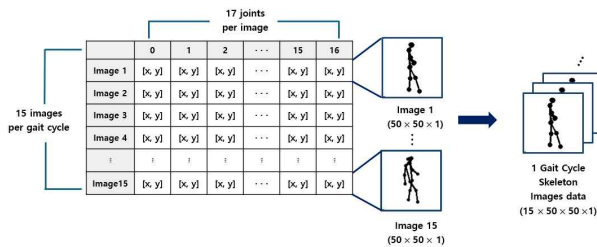
| Index | Key point | Index | Key point |
|-------|-------------------------|-------|----------------------|
| 0 | 코 (Nose) | 9 | 왼쪽 손목 (Left Wrist) |
| 1 | 왼쪽 눈 (Left Eye) | 10 | 오른쪽 손목 (Right Wrist) |
| 2 | 오른쪽 눈 (Right Eye) | 11 | 왼쪽 엉덩이 (Left Hip) |
| 3 | 왼쪽 귀 (Left Ear) | 12 | 오른쪽 엉덩이 (Right Hip) |
| 4 | 오른쪽 귀 (Right Ear) | 13 | 왼쪽 무릎 (Left Knee) |
| 5 | 왼쪽 어깨 (Left Shoulder) | 14 | 오른쪽 무릎 (Right Knee) |
| 6 | 오른쪽 어깨 (Right Shoulder) | 15 | 왼쪽 발목 (Left Ankle) |
| 7 | 왼쪽 팔꿈치 (Left Elbow) | 16 | 오른쪽 발목 (Right Ankle) |
| 8 | 오른쪽 팔꿈치 (Right Elbow) | | |

<그림 1> COCO pose 형식

2.2 데이터 전처리

2D 스켈레톤 데이터에서 17개 관절의 x, y 좌표를 추출한 후, COCO pose 형식에 따라 관절 간 선을 연결하여 뼈대를 시각화하였다. 시각화된 이미지를 50×50×1 이미지로 변환하고, 안정적인 모델 학습을 위해 최대-최소 정규화를 적용하였다. 그 후, 전체 이미지 데이터를 보행 1주기마다 나누어, 2차원 스켈레톤 이미지가 시간 축(보행 1주기)에 따라 늘어선 형태의 영상 데이터로 구성하였다. 이미지 개수를 일치시키기 위해 한 피험자당 3개의 보행 1주기에 해당하는 데이터를 사용하였다.

성별 데이터는 '남성(M)'은 0, '여성(F)'은 1로 변환해 분류 모델의 훈련 및 예측에 활용하였고, 데이터의 불균형 문제를 방지하기 위해 무작위로 남성 데이터 86개를 제거하여 남성(113명)과 여성(113명)의 수를 일치시켰다.



〈그림 2〉 보행 1주기의 관절 좌표에 대한 이미지 변환

또한, 걷기 방법에 따른 성별 분류 정확도를 파악하기 위해 <표 2>와 같이 데이터를 4가지의 상황으로 나누어 분석하였다. 데이터 전처리 과정에서 이미지 개수가 적은 피험자는 제외하였으며, 총 308명의 보행 영상을 4가지 상황별로 모델의 입력 데이터로 사용하였다. 입력 데이터셋의 형태는 <그림 2>와 같다.

〈표 2〉 4가지 상황별 데이터셋

| 상황 | 포함된 데이터 | 입력 데이터 수 |
|-------------------|-------------------------------|----------|
| 전체 데이터셋 (A) | NM, CL, BG, WSS, WSF, TXT, PH | 5,664개 |
| 신체 일부가 가려진 상황 (B) | NM, CL, BG | 2,427개 |
| 보행 속도가 다른 상황 (C) | NM, WSS, WSF | 2,426개 |
| 휴대폰을 사용하는 상황 (D) | NM, TXT, PH | 2,417개 |

2.3 딥러닝 모델 성능평가

보행 1주기의 보행 영상 데이터를 사용하여 4가지 보행 상황에서 성별을 분류하기 위해, 이미지 분석에 뛰어난 성능을 보이는 CNN(Convolutional Neural Network) 기반의 ResNet 모델을 사용한다. 특히, 3D ResNet 모델은 3D 컨볼루션 레이어와 배치 정규화, ReLU 활성화 함수로 구성된 Residual Block을 포함하고 있어 공간적 및 시간적 특징을 동시에 학습할 수 있어 영상 데이터를 분석하기에 적합하다. 본 연구에서는 3D ResNet-18 모델을 사용하였으며, ResNet-50 이상의 더 깊은 모델은 학습 시간만 길어질 뿐 성능 향상에 도움을 주지 않았다[8].

3D ResNet-18 모델의 입력은 (15, 50, 50, 1) 형태로, 각 피험자의 1주기 보행 시 스켈레톤 이미지 데이터를 사용하였다. 모델의 activation 함수는 ReLU, output activation 함수는 sigmoid를 사용하였고, epoch은 100으로 설정해 학습시켰다. 또한, 모델의 안정적인 학습을 위해 학습 과정에서 학습률 감소 방법을 사용하였다. 보행 1주기 2D 스켈레톤 이미지를 무작위로 섞어 8:1:1 비율로 학습 데이터, 검증 데이터, 테스트 데이터로 나누어 사용하였으며, 각 데이터셋의 입력 데이터 수는 <표 2>에 나타나 있다.

2.4 결과 및 고찰

4가지 보행 상황에서의 보행 1주기의 보행 영상 데이터를 입력으로 사용한 ResNet-18의 성별 분류 결과는 <표 2>와 같다. 모델의 성별 분류 성능평가는 정확도(accuracy), 정밀도(precision), 재현율(recall), F1 점수(F1 score)를 사용한다.

〈표 3〉 ResNet-18을 사용한 성별 분류 결과

| 상황 | 정확도(%) | 정밀도 | 재현율 | F1 점수 |
|------------------|--------|-------|-------|-------|
| 전체 데이터셋(A) | 80.1 | 0.804 | 0.801 | 0.8 |
| 신체 일부가 가려진 상황(B) | 83.9 | 0.839 | 0.839 | 0.839 |
| 보행 속도가 다른 상황(C) | 77.2 | 0.777 | 0.773 | 0.772 |
| 휴대폰을 사용하는 상황(D) | 80.6 | 0.807 | 0.806 | 0.806 |

결과 지표에 따르면, 모델은 안정적이고, 균형 잡힌 성능을 보인다. 신체 일부가 가려진 상황(B)에서 분류 정확도가 83.9%로 가장 높게 나타났으며, 이는 Kinect와 같은 관절 측정 장비가

없이도 추출된 관절 이미지만으로 모델이 보행의 주요 특징을 잘 파악하고 불필요한 정보를 배제하여 중요한 특징에 집중할 수 있기 때문으로 사료된다. 또한, 보행 속도가 다른 상황(C)에서는 77.2%로 가장 낮은 정확도를 보였는데, 이는 속도에 따라 보행 패턴이 달라지기 때문에 일관된 특징을 추출하는 데 어려움을 겪었기 때문으로 분석된다. 이러한 결과는 측정 장비 없이 수집된 관절 이미지를 통해 일상생활에서 발생하는 다양한 보행 상황에서 성별을 분류할 수 있음을 입증하였다.

3. 결 론

본 연구에서는 ResNet-18 모델을 사용하여 4가지 보행 상황에서의 영상 데이터로 성별을 분류하였으며, 대부분의 상황에서 80% 이상의 분류 정확도를 나타내면서 관절 측정 장비 없이 수집된 관절 이미지를 통해 일상생활에서 발생하는 보행 상황에서 성별 분류의 가능성을 확인하였다.

하지만 보행 상황은 본 연구에서 제시한 4가지 상황만으로 한정할 수 없기 때문에, 모든 보행 상황에서 성별 분류가 가능하다고 단정할 수 없다. 향후 더욱 다양한 상황과 조건에 대한 보행 데이터를 확보하여 폭넓고 신뢰성 높은 성별 분류 모델을 구축하고자 한다. 또한, 딥러닝 모델의 성능을 개선하여 다양한 보행 상황에서의 성별 분류 정확도를 향상시킴으로써 범용성이 높은 실생활 보행기반 성별 분류 연구에 기여할 것으로 기대된다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 2024년도 SW중심대학사업의 연구 결과로 수행되었음(2021-0-01399)
이 논문은 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2021R1I1A3059769)
본 논문은 한국연구재단 4단계 두뇌한국21사업(4단계 BK21사업)의 지원을 받아 작성되었음(과제번호:5199990914048)

[참 고 문 헌]

- [1] Azhar, Muhammad, et al, "A gait-based real-time gender classification system using whole body joints", Sensors, 22.23, 9113, 2022
- [2] Bovonsunthonchai, et al, "The impact of different mobile phone tasks on gait behaviour in healthy young adults." Journal of Transport & Health, 19, 100920, 2020
- [3] Xu, Shihao, et al. "Emotion recognition from gait analyses: Current research and future directions", IEEE Transactions on Computational Social Systems, 11.1, 363-377, 2022
- [4] W. Zhang, et al, "Gender and gaze gesture recognition for human-computer interaction", Computer Vision and Image Understanding, vol. 149, 32-50, 2016
- [5] V. Albiero, K. Zhang, M. C. King, K. W. Bowyer, "Gendered Differences in Face Recognition Accuracy Explained by Hairstyles, Makeup, and Facial Morphology," IEEE Transactions on Information Forensics and Security, vol. 17, 127-137, 2022
- [6] D. Guffanti, et al, "Non-Invasive Multi-Camera Gait Analysis System and its Application to Gender Classification", IEEE Access, vol 8, 95734-95746, 2020
- [7] Cosma, Adrian, et al, "PsyMo: A Dataset for Estimating Self-Reported Psychological Traits from Gait", IEEE/CVF Winter Conference on Applications of Computer Vision, p. 4603-4613, 2024.
- [8] He, Kaiming, et al, "Deep residual learning for image recognition", IEEE conference on computer vision and pattern recognition, p. 770-778, 2016