

## Multi-Time Window Feature Extraction Technique for Anger Detection in Gait Data

Beom Kwon\*, Taegeun Oh\*\*

\*Assistant Professor, Dept. of Artificial Intelligence, Dongyang Mirae University, Seoul, Korea

\*\*Assistant Professor, Dept. of Electrical and Electronic, Dong Seoul University, Gyeonggi-do, Korea

### [Abstract]

In this paper, we propose a technique of multi-time window feature extraction for anger detection in gait data. In the previous gait-based emotion recognition methods, the pedestrian's stride, time taken for one stride, walking speed, and forward tilt angles of the neck and thorax are calculated. Then, minimum, mean, and maximum values are calculated for the entire interval to use them as features. However, each feature does not always change uniformly over the entire interval but sometimes changes locally. Therefore, we propose a multi-time window feature extraction technique that can extract both global and local features, from long-term to short-term. In addition, we also propose an ensemble model that consists of multiple classifiers. Each classifier is trained with features extracted from different multi-time windows. To verify the effectiveness of the proposed feature extraction technique and ensemble model, a public three-dimensional gait dataset was used. The simulation results demonstrate that the proposed ensemble model achieves the best performance compared to machine learning models trained with existing feature extraction techniques for four performance evaluation metrics.

▶ **Key words:** Anger Detection, Emotion Recognition, Ensemble Learning, Feature Extraction, Gait Data, Machine Learning

### [요 약]

본 논문에서는 보행자의 걸음걸이로부터 분노 감정 검출을 위한 다중 시간 윈도우 특징 추출 기술을 제안한다. 기존의 걸음걸이 기반 감정인식 기술에서는 보행자의 보폭, 한 보폭에 걸리는 시간, 보행 속도, 목과 흉부의 전방 기울기 각도(Forward Tilt Angle)를 계산하고, 전체 구간에 대해서 최솟값, 평균값, 최댓값을 계산해서 이를 특징으로 활용하였다. 하지만 이때 각 특징은 보행 전체 구간에 걸쳐 항상 균일하게 변화가 발생하는 것이 아니라, 때로는 지역적으로 변화가 발생한다. 이에 본 연구에서는 장기부터 중기 그리고 단기까지 즉, 전역적인 특징과 지역적인 특징을 모두 추출할 수 있는 다중 시간 윈도우 특징 추출(Multi-Time Window Feature Extraction) 기술을 제안한다. 또한, 제안하는 특징 추출 기술을 통해 각 구간에서 추출된 특징들을 효과적으로 학습할 수 있는 앙상블 모델을 제안한다. 제안하는 앙상블 모델(Ensemble Model)은 복수의 분류기로 구성되며, 각 분류기는 서로 다른 다중 시간 윈도우에서 추출된 특징으로 학습된다. 제안하는 특징 추출 기술과 앙상블 모델의 효과를 검증하기 위해 일반인에게 공개된 3차원 걸음걸이 데이터 세트를 사용하여 시험 평가를 수행했다. 그 결과, 4가지 성능 평가지표에 대해서 제안하는 앙상블 모델이 기존의 특징 추출 기술로 학습된 머신러닝(Machine Learning) 모델들과 비교하여 최고의 성능을 달성하는 것을 입증하였다.

▶ **주제어:** 분노 검출, 감정인식, 앙상블 학습, 특징 추출, 걸음걸이 데이터, 머신러닝

- First Author: Beom Kwon, Corresponding Author: Taegeun Oh
- \*Beom Kwon (bkwon@dongyang.ac.kr), Dept. of Artificial Intelligence, Dongyang Mirae University
- \*\*Taegeun Oh (tgoh@du.ac.kr), Dept. of Electrical and Electronic, Dong Seoul University
- Received: 2023. 04. 03, Revised: 2023. 04. 21, Accepted: 2023. 04. 24.

## I. Introduction

최근, 간헐적 폭발 장애(Intermittent Explosive Disorder)에 의한 인한 기물 파손, 방화, 살인, 폭언, 폭행 등 사건 사고가 잇달고 있다[1-3]. 일반인들에게는 간헐적 폭발 장애라는 이름보다는 분노조절장애라는 이름으로 더 잘 알려져 있으며, 분노와 관련된 감정 조절을 정상적으로 할 수 없어 공격 충동이 억제되지 않아 공격성이 행동으로 표현되는 일종의 정신질환이다. 특히 총기 소지가 합법인 미국과 같은 국가에서는 분노조절장애를 앓고 있는 사람이 총기를 공원, 학교와 같이 사람들이 모여 있는 장소에서 무차별 난사하여 다수의 인명 피해가 발생하는 상황으로까지 이어지고 있어, 분노조절장애는 더 이상 개인만의 문제가 아닌 사회적인 문제로 다뤄지고 있다[4-6].

최근, 영상 감시 시스템(Video Surveillance System)은 컴퓨터비전(Computer Vision) 기술과 인공지능 기술을 활용하여, 카메라로부터 획득한 영상 데이터에서 자동으로 사람을 검출할 수 있는 지능형 영상 감시 시스템으로 발전하고 있다[7]. 최근에는 단순히 사람을 검출하는 데에 그치지 않고 사람의 행동을 인식하는 수준까지 이르러, 그 활용도가 매우 높아졌다. 예를 들어, 분노조절장애를 앓고 있는 사람의 공격적인 행동을 인식하는 데에도 활용될 수 있으며, 교도소, 대규모 집회 및 시위장, 스포츠 경기장, 정신 병원과 같은 다양한 형태의 폭력과 타인을 향한 공격 충동이 일어나기 쉬운 환경에서도 활용될 수 있다[8-10].

하지만 행동 인식 기술은 사람이 행동한(행동하는) 것을 인식하는 기술이지, 사람의 미래 행동을 예측하는 기술은 아니다[11]. 따라서 행동 인식 기술을 탑재한 지능형 영상 감시 시스템은 공격적인 행동을 인식하여 사후 조치를 할 때 활용될 수 있지만, 사전에 공격적인 행동을 예측하고 이를 방지하는 데 활용하기에는 어려움이 있다.

본 논문에서는 이 어려움을 극복하기 위한 한 가지 방법으로, 사람의 분노 감정을 인식하는 방법에 관한 연구를 수행하였다. 예컨대 분노의 감정을 가진 사람이 타인을 향해 공격적인 행동을 하기 전에, 분노 감정을 검출하여 이를 보안 요원에게 알린다면 보안 요원이 조기에 적절한 조치를 할 수 있을 것이다.

분노 감정 검출과 관련된 기존 연구들에서는 목소리, 몸의 자세 및 움직임, 표정 등의 다양한 데이터가 사용됐다. 목소리 기반 분노 검출 기술의 경우 음성 감지 센서가 필요한데, 검출 성능이 센서 주변에서 발생하는 소음에 의해 영향을 받는다. 가령 애초에 소음 신호의 크기가 클 수도 있고, 소음 신호의 크기가 작더라도 센서와 사람 사이의

거리가 멀어 녹음되는 음성 신호의 크기가 소음 신호의 크기와 비슷할 수도 있다. 이 경우 음성 신호와 주변 소음 신호를 분간하기 어려워지고, 그 결과 높은 검출 성능을 기대할 수 없게 된다[12].

표정 기반 기술에서는 이미지 센서가 사용된다. 이 기술의 경우 검출 성능이 사람의 얼굴과 이미지 센서 사이의 거리, 촬영 각도에 영향을 받는다. 사람의 표정은 안면 근육의 움직임을 통해 표현되는데, 근육의 움직임을 정확하게 포착하기 위해서는 얼굴과 가까운 위치에서 촬영이 이루어져야 한다. 또한 얼굴의 정면이 아닌 측면이 촬영되거나 마스크를 착용하고 있으면, 얼굴의 일부분이 가려져 감정 검출에 활용할 수 있는 표정 정보가 제한되고 이 경우 높은 검출 성능을 기대할 수 없게 된다[13].

최근, 인간 자세 추정(Human Pose Estimation) 기술의 발전으로 사람의 모션 데이터(Human Motion Data)의 획득이 더욱 빠르고 간단해졌다[14]. 그 결과 모션 데이터는 현재 다양한 응용 분야에서 활용되고 있다. 대표적인 응용 분야에는 사람 식별, 성별 분류, 애니메이션 생성, 체질량지수 분류, 행동 인식 등이 있다[15-20]. 기존의 목소리 및 표정 기반 감정 검출 기술과 비교하여, 모션 데이터 기반 기술은 상대적으로 멀리 떨어진 거리에서도 수행될 수 있으며, 전방뿐만 아니라 다양한 각도에서 촬영된 영상 데이터에 대해서도 수행될 수 있다는 이점이 있다.

본 논문에서 연구한 분노 감정 검출 기술은 모션 데이터 기반 기술 범주에 속하며, 다양한 모션 중에서도 걸음걸이에 초점을 맞추어 연구를 수행하였다. 걸음걸이에 초점을 맞춘 이유는 걸음걸이가 사람이 수행하는 가장 보편적인 동작이며, 또한 일상생활에서 가장 많이 관찰되는 동작이기 때문이다. 본 연구의 주요 기여 포인트는 다음과 같다.

- 다중 시간 윈도우 특징 추출(Multi-time Window Feature Extraction) 기술을 제안한다. 기존 연구들에서는 걸음걸이 데이터의 시작과 끝 사이에서 최솟값, 평균값, 최댓값 계산을 통해 추출된 특징들(Features)을 분노 검출에 활용하는데, 본 연구에서는 다중 시간 윈도우를 사용하여 여러 다른 시간대에서 추출한 특징들을 활용하여 검출 정확도를 개선하였다.

- 3차원 걸음걸이 데이터(3D Gait Data)에서 사람의 분노 감정을 검출할 수 있는 앙상블 모델(Ensemble Model)을 제안한다. 제안하는 모델은 다중 시간 윈도우에서 계산된 특징들을 활용하여 복수의 머신러닝 분류기(Machine Learning Classifiers)를 각각 독립적으로 학습시킨다. 이후 각 분류기에서 얻은 검출 결과를 종합하여 최종 결과를 도출한다. 일반인에게 공개된 걸음걸이 데이터 세트를 활

용하여 제안하는 모델의 성능 평가를 수행한 결과, 제안 모델은 기존 기법들보다 우수한 성능을 달성하였다.

본 논문의 구성은 다음과 같다. II장에는 걸음걸이 데이터 기반의 감정인식 기술과 관련된 기존 연구들에 대한 검토내용이 기술되어 있다. III장에는 본 연구의 동기가 기술되어 있다. IV장에서는 본 논문에서 제안하는 분노 감정 검출 방법이 기술되어 있다. V장에서는 제안 방법의 성능 평가를 위해 사용된 데이터 세트, 실험 환경 및 성능 평가 결과에 대해 기술되어 있다. VI장에서는 본 논문의 결론 및 향후 과제에 대해 기술되어 있다.

## II. Related Works

걸음걸이로부터 사람의 감정을 인식하는 연구는 지난 수십 년간 꾸준히 진행됐다. 예를 들어, Montepare et al.은 감정 상태에 따른 걸음걸이 차이를 관찰하기 위한 실험을 수행하였다[21]. 이 실험에 참여한 지원자는 5명이었으며, 각 지원자는 네 가지 감정(분노, 행복, 자부심, 슬픔) 상태로 견도록 요청받았다. 실험 결과, 사람의 감정 상태에 따라 팔 스윙(Swing), 보폭, 보행 속력이 다양하게 변화된다는 것이 관찰됐다. Crane과 Gross는 감정 상태와 보행 속력 사이의 관계를 조사하는 데 초점을 맞추었다[22]. 이를 위해, Crane과 Gross는 42명의 대학생(여성 22명, 남성 20명)을 대상으로 다섯 가지 감정 상태(분노, 만족, 기쁨, 중립, 슬픔)에 따른 걸음걸이 차이를 비교하는 실험을 수행했다. 실험 결과에 따르면, 사람의 감정은 보행 속력에 영향을 미쳤다. 예를 들어, 기쁨의 경우 가장 빠른 보행 속력이 관찰되었다. 또한, 분노의 경우 슬픔의 경우보다 보행 속력이 더 빨랐다.

Karg는 걸음걸이 기반 감정인식 분야에서 많은 연구를 수행했다[23-26]. 예를 들어, [23]에서 Karg et al.은 일반적인 상태에서의 보행과 피로한 상태에서의 보행을 구분하는 방법을 연구했다. 연구를 위해서 Karg et al.은 VICON이라 불리는 모션 캡처 시스템(Motion Capture System)을 사용하여 14명의 남성 참가자의 걸음걸이 데이터를 촬영 및 기록했다. 그리고 주성분 분석(Principal Component Analysis, PCA)을 사용하여 데이터로부터 주성분을 추출하고, 그중 처음 네 개의 주성분을 아이겐 포즈(Eigen Pose)라고 명명하였다. 그리고 아이겐 포즈를 특징으로 활용하여 선형 판별 분석(Linear Discriminant Analysis), 나이브 베이즈(Naive Bayes), k-최근접 이웃(k-Nearest Neighbors, k-NN) 및 서포트 벡터 머신(Support Vector Machine)을 포함한 네 개의 머신러닝

모델을 학습시켰다. 학습이 완료된 모델들을 평가한 결과, 나이브 베이즈 모델이 정확도 96%를 달성하였고, 다른 모델들은 모두 정확도 100%를 달성했다. [24]에서 Karg et al.은 [23]에서 제안한 방법을 확장하여 걸음걸이로부터 사람의 네 가지 감정(분노, 기쁨, 중립, 슬픔)을 인식하는 연구를 수행했다. 그 결과, 나이브 베이즈 모델이 72%의 최고 분류 정확도를 달성하였다. [25]에서 Karg et al.은 프레임(Frame)마다 어깨의 외전 각도(Abduction Angle)와 목과 흉부의 전방 기울기 각도(Forward Tilt Angle)를 계산했다. 그다음 각 각도의 최솟값, 평균값 및 최댓값을 특징으로 하여, 보행자의 감정 상태를 인식하는 데 활용했다. [26]에서는 보행자의 감정인식을 위한 두 단계 분류 접근법이 제안됐다. 첫 번째 단계에서는 [25]에서 제안된 특징을 사용하여 보행자의 신원을 추정한다. 그다음, 두 번째 단계에서는 추정된 신원 정보를 바탕으로 사용자별 맞춤형으로 학습된 분류기(Classifier)를 선택한다. 그다음 선택된 분류기를 사용하여 보행자의 감정을 추정한다.

앞서 [21-26]의 연구들에서 걸음걸이 기반 감정인식을 위해 활용된 특징들은 모두 시간 영역(Time Domain)에서 추출됐다. 일부 연구들에서는 주파수 영역(Frequency Domain)에서 추출한 특징을 활용한 감정인식 연구를 수행했다. 예를 들어, [27]에서 Li et al.은 6개 관절(좌우 무릎, 발목, 손목)의 시계열 데이터에 대해 이산 푸리에 변환(Discrete Fourier Transform, DFT)을 수행하고, 처음 20개 계수의 진폭과 위상을 주파수 영역의 특징으로 선택했다. [28]에서 Xue et al.은 이산 코사인 변환(Discrete Cosine Transform, DCT)을 사용하여 주파수 영역에서 특징을 추출하는 방법을 고안했다. Xue et al.은 Microsoft의 Kinect 센서를 사용하여 걸음걸이 데이터를 64프레임 길이로 기록했는데, Kinect 센서는 모션을 25개 관절의 3D 좌표값 형태로 표현한다. Xue et al.은 총 75개의 좌표값(=25개 관절×3D 좌표값/관절)에 대해서 DCT를 수행했다. 그리고 그 결과 각 걸음걸이 데이터에서 총 4,800개의 주파수 영역의 특징(= 64프레임×75개 좌표값/프레임)이 추출되었으며, 특징 모두 보행자 감정을 인식하는 데 활용됐다.

최근 몇 년간 딥러닝(Deep Learning)이 수많은 분야에서 기존의 성능을 뛰어넘는 최고의 성능을 달성하면서, 연구자들로부터 큰 관심을 받고 있다. 딥러닝에서는 최적의 성능을 달성하는 데 필요한 특징이 대규모 데이터로부터 자동으로 학습된다. 일부 연구 그룹은 딥러닝 기술을 사용하여 걸음걸이 데이터에서 사람의 감정을 인식하는 연구를 수행했다. 예를 들어, Zhuang et al.은 사람의 감정과 관련된 특징이 자동으로 학습되는 G-GCSN(Global Graph Convolution Shrinkage Network)을 제안했다

[29]. Narayanan et al.은 합성곱 신경망(Convolutional Neural Network, CNN) 기반 감정인식 방법을 제안했다 [30]. 이 방법에서는, 미리 정의된 시간 동안의 3D 포즈 데이터를 이미지로 변환하고, 이 이미지를 ProxEmo라 불리는 CNN 아키텍처(Architecture)에 입력한다. 시험 평가에서 ProxEmo는 입력된 이미지에 대해서 보행자의 감정을 82.4%의 정확도로 인식했다. [31]에서 Bhattacharya et al.은 보행자의 감정을 인식하는 STEP(Spatial Temporal Graph Convolutional Networks for Emotion Perception)이라 불리는 아키텍처를 제안했다. Bhattacharya et al.은 STEP의 성능을 평가하기 위해, 5,227개의 걸음걸이 데이터(실제 데이터: 4,227개, 합성 데이터: 1,000개)로 구성된, E-Gait라 불리는 걸음걸이 데이터베이스(Database)를 구축했다. E-Gait를 이용한 시험 평가에서 STEP은 88%의 분류 정확도를 달성하여, 기존 방법들보다 더 높은 정확도를 보였다.

### III. Motivation

본 연구에서는 3D 걸음걸이 데이터에서 사람의 분노 감정을 검출하는 방법과 검출 정확도를 높이는 방법을 조사했다. 이를 위해, 기존 연구들에서 제안된 특징 추출 과정을 면밀하게 검토하고, 각 특징이 분노 검출 정확도에 미치는 영향을 측정했다. [21,22,25,26]의 기존 연구들에서 3D 걸음걸이 데이터에서 보폭(요소1), 한 보폭에 걸리는 시간(요소2), 보행 속도(요소3), 목의 전방 기울기 각도(요소4), 흉부의 전방 기울기 각도(요소5)를 계산했다. 그다음 걸음걸이 데이터의 시작과 끝 사이에서, 요소별로 최솟값, 평균값, 최댓값을 찾고 이를 특징으로 활용했다.

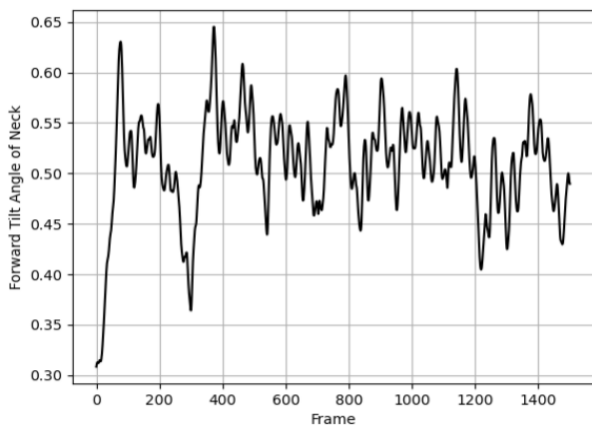


Fig. 1. Forward Tilt Angle of Neck at Calculated Each Frame

그림 1은 [32]에서 Ma et al.가 소개한 BML(Body Movement Library)이라 불리는 데이터 세트에서, 네 번째 지원자의 화난 상태 걸음걸이 데이터에서 시간에 따른 목의 전방 기울기 각도 값의 변화를 보여준다. 이 그림에서 볼 수 있듯이, 목의 전방 기울기 각도는 프레임에 따라 즉 시간에 따라 그 값이 변한다. 그리고 분산이 상대적으로 크거나 작은 특정 구간들이 있는데, 본 연구에서는 이 구간들에 주목했다. 기존 연구들에서는 상기 구간들을 무시하고, 전체 구간에 대해서 최솟값, 평균값, 최댓값을 찾고 이를 특징으로 활용하여 분노 감정을 검출하였다.

본 연구에서는 상기 특정 구간들이 사람의 분노 감정을 검출하는 데 활용될 수 있으며, 이를 활용하였을 때 검출 정확도가 개선되는 것을 확인했다. 이를 토대로, 본 연구에서는 전체 구간에 대해 찾아진 특징만을 사용하는 것이 아닌, 여러 시간대에 걸쳐 지역화된 구간에서 찾아진 특징을 함께 사용하는 방법을 제안한다. 그리고 서로 다른 구간에서 찾아진 특징들을 분류기 학습에 활용하여, 걸음걸이 데이터로부터 분노 감정을 검출하는 앙상블 모델을 제안한다.

### IV. Proposed Method

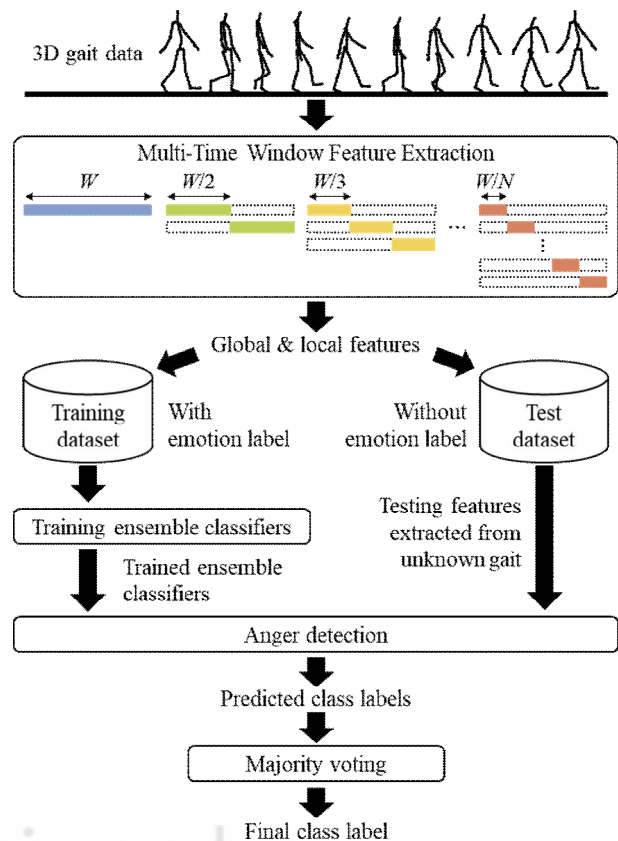


Fig. 2. Schematic Overview of the Proposed Method

그림 2는 걸음걸이로부터 분노 감정을 검출하기 위해 본 논문에서 제안하는 방법의 전체 개략도이다. 제안하는 방법은 그림 2에서와 같이 3D 걸음걸이 데이터를 입력으로 사용한다. 3D 걸음걸이 데이터는 시계열 데이터(Time Series Data)로, 매 프레임에서의 순간 자세는  $J$ 개의 관절로 구성된 3D 인간 스켈레톤(Human Skeleton) 모델로 표현된다. 여기서  $J$ 의 값은 모션 캡처 시스템 종류에 따라 다를 수 있다. 예를 들어, Microsoft Kinect 센서 버전 1의 경우  $J=20$ , 버전 2의 경우  $J=25$ , Azure Kinect 센서의 경우  $J=32$ 이다. 그리고 본 논문에서 사용한 BML 데이터 세트에서 제공하는 스켈레톤 모델의 경우 그림 3에서와 같이  $J=15$ 이다.

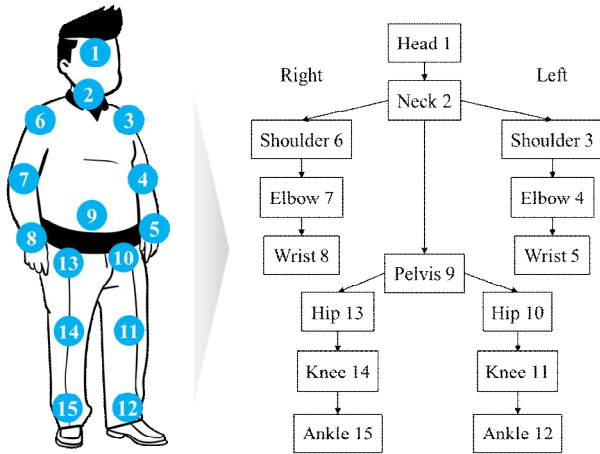


Fig. 3. 3D Human Skeleton Model Consisting of 15 Joints

입력된 3D 걸음걸이 데이터의 전체 프레임에 대해 최솟값, 평균값, 최댓값을 계산하는, 즉 전역적인(Global) 특징만을 추출하여 사용하는 기존 연구들과는 달리, 본 연구에서는 여러 시간대에 걸쳐 지역적인(Local) 특징도 함께 추출하는 다중 시간 윈도우 특징 추출 과정을 수행한다. 제안하는 다중 시간 윈도우 특징 추출 과정의 설명을 위해, 입력 데이터의 총 프레임 길이를  $W$ 라고 한다면 제안하는 특징 추출 과정에서는 길이가  $W$ 인 시간 윈도우가 1개, 길이가  $W/2$ 인 시간 윈도우가 2개, 길이가  $W/3$ 인 시간 윈도우가 3개, ..., 길이가  $W/N$ 인 시간 윈도우가  $N$ 개, 즉 총  $N(N+1)/2$ 개의 다중 시간 윈도우가 사용된다.

제안 방법에서는 상기 다중 시간 윈도우가 걸쳐진 영역에 대해서 특징 추출 과정이 독립적으로 수행되는데, 본 연구에서는 1) 보폭, 2) 한 보폭에 걸리는 시간, 3) 보행 속력, 4) 목의 전방 기울기 각도, 5) 흉부의 전방 기울기 각도를 계산해 분노 검출을 위한 특징들로 활용한다.

제안 방법의 특징 추출 과정 설명을 위해,  $j$ 번째 관절의  $t$ 번째 프레임에서의 좌표를  $P_j[t] = (x_j[t], y_j[t], z_j[t])$ 라

하자. 여기서  $x_j[t]$ ,  $y_j[t]$ ,  $z_j[t]$ 는 각각  $j$ 번째 관절의  $t$ 번째 프레임에서의  $x$ 축,  $y$ 축,  $z$ 축 좌표값이다.

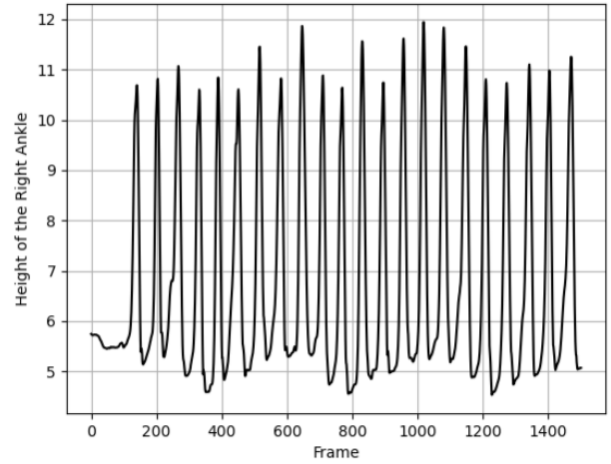


Fig. 4. Trajectory of  $z_{15}[t]$  During Walking

그림 4는 오른쪽 발목(Right Ankle)의  $z$ 축 좌표값 즉,  $z_{15}[t]$ 의 시간에 따른 변화를 보여준다. 여기서 극소점(Local Minimum Point)은 오른쪽 발목이 보행 중에 지면에 닿는 시점을 나타내며, 왼쪽 발이 한 발자국 움직일 때까지 한동안 그 값이 유지된다. 그리고 왼쪽 발의 움직임이 종료된 후에, 오른쪽 발이 지면에서 떨어지면서 오른쪽 발목의 높이( $=z_{15}[t]$ )가 증가한다. 그리고 극대점(Local Maximum Point)에 도달한 이후에 오른쪽 발이 다시 지면을 향해 내려가기 시작하면  $z_{15}[t]$  값이 감소하고, 발이 지면에 닿게 되면 극소점이 생성된다. 이 과정은 보행 중에 반복적으로 일어나며, 결과적으로 그림 4에서와 같이 극소점과 극대점이 번갈아 가며 반복적으로 생성되는 형태로 표현된다.

본 논문에서는  $z_{15}[t]$  값을 기반으로 보폭을 계산한다. 지면에 오른쪽 발목이 닿는 시점 즉, 극소점을 보행의 시작으로 간주하고, 그 후에 일어나는 오른쪽 발목의 지면 접촉 즉, 다음 극소점을 보행의 종료로 간주한다.  $i$ 번째 극소점을  $t_i$ 이라 하고,  $(i+1)$ 번째 극소점을  $t_{i+1}$ 이라 하면  $i$ 번째 보폭( $=Stride_i$ )은 다음과 같이 계산할 수 있다.

$$Stride_i = \sqrt{(x_{15}[t_{i+1}] - x_{15}[t_i])^2 + (y_{15}[t_{i+1}] - y_{15}[t_i])^2}. \quad (1)$$

$i$ 번째 보폭에 대해, 보폭이 시작되고 종료되기까지 걸리는 시간( $=Time_i$ )은 다음과 같이 계산할 수 있다.

$$Time_i = t_{i+1} - t_i. \quad (2)$$

그리고  $i$ 번째 보폭에 대해, 평균 속력( $=Speed_i$ )은 수식 (1)과 (2)를 이용하여 다음과 같이 계산할 수 있다.

$$Speed_i = Stride_i / Time_i. \quad (3)$$

제안 방법에서는 입력된 3D 걸음걸이 데이터의 전체 프레임 중에 다중 시간 윈도우가 걸쳐진 각 영역에 대해서 독립적으로, 수식 (1)-(3)을 이용하여  $Stride_i$ ,  $Time_i$ ,  $Speed_i$ 을 계산한다. 해당 영역에서 보폭이  $I$ 번 있었다면,  $I$ 개의 계산된 값들로부터 최솟값, 평균값, 최댓값을 구하여 특징으로 활용하며, 수식으로 표현하면 다음과 같다.

$$Stride_{\min} = \min(Stride_1, \dots, Stride_I), \quad (4)$$

$$Stride_{\text{mean}} = \text{mean}(Stride_1, \dots, Stride_I), \quad (5)$$

$$Stride_{\max} = \max(Stride_1, \dots, Stride_I), \quad (6)$$

$$Time_{\min} = \min(Time_1, \dots, Time_I), \quad (7)$$

$$Time_{\text{mean}} = \text{mean}(Time_1, \dots, Time_I), \quad (8)$$

$$Time_{\max} = \max(Time_1, \dots, Time_I), \quad (9)$$

$$Speed_{\min} = \min(Speed_1, \dots, Speed_I), \quad (10)$$

$$Speed_{\text{mean}} = \text{mean}(Speed_1, \dots, Speed_I), \quad (11)$$

$$Speed_{\max} = \max(Speed_1, \dots, Speed_I). \quad (12)$$

또한, 제안 방법에서는 목의 전방 기울기 각도를 프레임마다 계산한다. 이를 위해서 먼저 Median Plane을 정의해야 하는데, 본 연구에서는 Median Plane을 법선 벡터가  $(x_{13}[t] - x_{10}[t], y_{13}[t] - y_{10}[t], 0)$ 이고 골반(Pelvis) 좌표  $P_9[t]$ 를 지나는 평면으로 정의하였으며, 수식 표현의 간결성을 위해 프레임 인덱스  $[t]$ 를 생략하고 정리하면 다음과 같다.

$$(x_{13} - x_{10})(x - x_9) + (y_{13} - y_{10})(y - y_9) = 0. \quad (13)$$

머리(Head) 좌표  $P_1 = (x_1, y_1, z_1)$ 과 목(Neck) 좌표  $P_2 = (x_2, y_2, z_2)$ 에서 Median Plane에 내린 수선의 발을 각각  $P_1' = (x_1', y_1', z_1')$ 와  $P_2' = (x_2', y_2', z_2')$ 라 하자. 그러면  $x_1'$ ,  $y_1'$ ,  $x_2'$ ,  $y_2'$ 은 다음과 같이 구할 수 있다.

$$x_1' = k_1(x_{13} - x_{10}) + x_1, \quad (14)$$

$$y_1' = k_1(y_{13} - y_{10}) + y_1, \quad (15)$$

$$x_2' = k_2(x_{13} - x_{10}) + x_2, \quad (16)$$

$$y_2' = k_2(y_{13} - y_{10}) + y_2. \quad (17)$$

수식 (14)-(15)에서  $k_1$ 은 다음과 같이 구할 수 있다.

$$k_1 = \frac{(x_9 - x_1)(x_{13} - x_{10}) + (y_9 - y_1)(y_{13} - y_{10})}{(x_{13} - x_{10})^2 + (y_{13} - y_{10})^2}. \quad (18)$$

수식 (16)-(17)에서  $k_2$ 은 다음과 같이 구할 수 있다.

$$k_2 = \frac{(x_9 - x_2)(x_{13} - x_{10}) + (y_9 - y_2)(y_{13} - y_{10})}{(x_{13} - x_{10})^2 + (y_{13} - y_{10})^2}. \quad (19)$$

목의 전방 기울기 각도  $\alpha$ 는 다음과 같이 구할 수 있다.

$$\alpha = \arcsin\left(\frac{\sqrt{(x_2' - x_1')^2 + (y_2' - y_1')^2}}{\sqrt{(x_2' - x_1')^2 + (y_2' - y_1')^2 + (z_2' - z_1')^2}}\right). \quad (20)$$

제안 방법에서는 다중 시간 윈도우가 걸쳐진 영역에 대해서 수식 (20)을 이용하여 매 프레임에서  $\alpha$  값을 계산한다. 다중 시간 윈도우의 시작 프레임을  $t_{start}$ 라 하고, 종료 프레임을  $t_{end}$ 라 하면 구간  $[t_{start}, t_{end}]$ 에서 계산된  $\alpha$  값 중에서 최솟값, 평균값, 최댓값을 구하여 특징으로 활용하며, 수식으로 표현하면 다음과 같다.

$$\alpha_{\min} = \min(\alpha[t]) \quad \forall t \in [t_{start}, t_{end}], \quad (21)$$

$$\alpha_{\text{mean}} = \text{mean}(\alpha[t]) \quad \forall t \in [t_{start}, t_{end}], \quad (22)$$

$$\alpha_{\max} = \max(\alpha[t]) \quad \forall t \in [t_{start}, t_{end}]. \quad (23)$$

제안 방법에서는 또한 흉부의 전방 기울기 각도  $\beta$ 를 프레임마다 계산하며, 수식 (16), (17), (19)를 이용하여 다음과 같이 구할 수 있다.

$$\beta = \arcsin\left(\frac{\sqrt{(x_9 - x_2')^2 + (y_9 - y_2')^2}}{\sqrt{(x_9 - x_2')^2 + (y_9 - y_2')^2 + (z_9 - z_2')^2}}\right). \quad (24)$$

제안 방법에서는 다중 시간 윈도우가 걸쳐진 구간  $[t_{start}, t_{end}]$ 에 대해서 수식 (24)를 이용하여 매 프레임에서  $\beta$  값을 계산한다. 그리고 계산된  $\beta$  값 중에서 최솟값, 평균값, 최댓값을 구하여 특징으로 활용하며, 수식으로 표현하면 다음과 같다.

$$\beta_{\min} = \min(\beta[t]) \quad \forall t \in [t_{start}, t_{end}], \quad (25)$$

$$\beta_{\text{mean}} = \text{mean}(\beta[t]) \quad \forall t \in [t_{start}, t_{end}], \quad (26)$$

$$\beta_{\max} = \max(\beta[t]) \quad \forall t \in [t_{start}, t_{end}]. \quad (27)$$

수식 (4)-(12), (21)-(23), (25)-(27)를 사용하여 각 특징을 추출한 후, 이를 결합하여(Concatenate) 특징 벡터로 만든다. 그리고 그 결과, 특징 벡터의 차원은 15가 된다.

제안 방법에서 앙상블 모델은 특징 추출 과정에서 사용된 다중 시간 윈도우의 총 개수  $N(N+1)/2$ 와 동일한 개수의 분류기로 구성된다. 학습 및 시험 단계에서, 각 다중 시간 윈도우 영역에서 추출된 특징 벡터는 해당 영역에 대응되는 분류기의 입력값으로 사용된다. 기존 연구들에서는 3D 걸음걸이 데이터의 전체 프레임에 대해 최솟값, 평균값, 최댓값을 계산하여, 즉 전역적인 특징만을 추출하여 머신러닝 모델을 학습시켰기 때문에, 머신러닝 모델은 지역적인 특징을 학습할 수 없었다. 반면에 제안하는 방법은 장기부터 중기 그리고 단기까지, 다중 시간 윈도우를 통해 전역적인 특징뿐만 아니라 지역적인 특징을 동시에 추출하여 머신러닝 모델을 학습시키기 때문에 기존 방법들과 비교하였을 때, 분노 감정 검출에 있어 우수한 성능을 달성하는 것이 가능하다.

제안하는 앙상블 모델을 구성하고 있는 분류기의 인덱스를  $c \in \{1, \dots, N(N+1)/2\}$ 라고 하자. 그리고  $c$ 번째 분류기의 분노 감정 검출 결과를  $r_c \in \{0, 1\}$ 라고 하자.

그러면 앙상블 모델 내  $N(N+1)/2$ 개의 분류기들의 결과는 다음과 같이 벡터  $R$ 로 표현할 수 있다.

$$R = [r_1, \dots, r_{N(N+1)/2}]. \quad (28)$$

그리고 수식 (28)의  $R$ 에 대해, 다수 투표법(Majority Voting)을 적용하여 제안하는 앙상블 모델의 최종 검출 결과값을 도출한다.

## V. Experimental Results

### 1. Dataset and Evaluation Protocol

본 연구에서는 [32]에서 소개된 BML이라 불리는 3D 걸음걸이 데이터 세트를 이용하여 제안 방법의 성능 평가를 수행하였다. BML에는 30명(남성: 15명, 여성: 15명)의 지원자로부터 수집된 3D 걸음걸이 데이터가 있다. 지원자 30명의 평균 나이는 22세이고, 가장 어린 지원자는 17세, 가장 나이가 많은 지원자는 29세이다. 지원자마다 네 가지 감정(분노, 행복, 중립, 슬픔) 상태의 3D 걸음걸이 데이터가 제공되며, 각 감정 상태에 대해 걸음걸이 동작은 두 번씩(시계방향으로 한 번, 반시계 방향으로 한 번) 기록되었다. 그 결과, 지원자마다 8개(=4가지 감정×2번 반복)의 3D 걸음걸이 데이터가 제공되고, BML 내 3D 걸음걸이 데이터의 개수는 총 240개이다. 하지만, 데이터 세트 내에서 이름이 "chr"이라고 명명된 7번째 지원자의 경우, 데이터 시각화를 통해 확인해 본 결과 중립 감정 상태에서 시계방향으로 걷는 동작이 올바르게 기록되어 있지 않았다. 따라서 본 연구에서는 상기 비정상 데이터를 제외한 239개의 3D 걸음걸이 데이터를 활용하였다. 그리고 분노 걸음걸이 데이터 60개에 대해서는 레이블(Label) 값을 1로 지정하였으며, 나머지 걸음걸이 데이터에 대해서는 레이블 값을 0으로 지정하였다.

제안하는 앙상블 모델의 성능을 평가에 앞서 전처리(Preprocessing) 과정으로, 추출한 특징들에 대해서 표준화(Standardization)를 통한 데이터 정규화를 수행하였다. 그 이후, 제안하는 앙상블 모델의 일반화된 성능을 얻기 위해서, 5-겹 교차 검증(5-Fold Cross Validation)을 사용하였다. 5-겹 교차 검증에는 전체 데이터의 80%를 사용하였고, 나머지 20%의 데이터는 5-겹 교차 검증으로부터 얻은 모델을 평가하는 데 사용하였다. 성능 평가지표로는 True Positive Rate(TPR), Positive Predictive Value(PPV),  $F_1$  Score, 정확도(Accuracy, ACC)를 사용하였다.

### 2. Results

제안하는 다중 시간 윈도우 특징 추출 기술과 앙상블 모델의 성능 평가를 위해서 본 연구에서는 Logistic Regression, k-NN, Naive Bayes, C-Support Vector Machine, Decision Tree의 5가지 머신러닝 알고리즘을 이용하였다. 그리고 각 알고리즘에 대해서, 파이썬 프로그래밍(Python Programming) 언어용 머신러닝 라이브러리(Library) 사이킷런(scikit-learn)을 활용하여 모델을 생성하였다. 그리고 각 모델의 최적 하이퍼파라미터 구성(Optimal Hyperparameter Configuration)을 찾기 위해서, HalvingGridSearchCV() 함수를 사용하여 하이퍼파라미터 최적화를 수행하였다.

Table 1. Ensemble Models with Different Numbers of Classifiers

Types of Multi-Time Windows	Total Number of Classifiers
$W, W/2$	3
$W/2, W/3$	5
$W, W/2, W/4$	7
$W, W/3, W/5$	9
$W, W/2, W/3, W/5$	11

제안하는 앙상블 모델을 개발하는 데 있어서, 어떤 종류의 분류기를 몇 개나 사용해야 가장 좋은 분노 검출 성능을 달성할 수 있는지, 알아내는 것은 매우 중요한 문제였다. 이 문제의 답을 얻기 위해서, 표 1에서와 같이 3, 5, 7, 9, 11개의 분류기로 구성된 앙상블 모델들을 설계했다. 여기서, 분류기의 개수를 홀수로 설정한 이유는 다음과 같다. IV장에서 설명한 바와 같이 제안하는 앙상블 모델은 각 분류기의 분류 결과(1: 분노 상태, 0: 분노 상태가 아니다)를 취합하고, 다수 투표법을 바탕으로 최종 검출 결과값을 도출한다. 그런데 이때, 만약 분류기의 개수가 짝수로 구성된다면 수식 (28)의  $R$  내에 0과 1의 개수가 같을 때 즉, 동점이 발생했을 때 다수 투표법만으로는 최종 분류 결과값을 도출하지 못하는 문제가 발생할 수 있다. 이에 본 연구에서는 제안하는 앙상블 모델을 구성하는 분류기의 개수를 홀수로 설정하였다.

Table 2. Performance Comparisons Between Ensemble Models

Case	TPR	PPV	$F_1$	ACC
$E(LR,3)$	0.500	0.375	0.429	0.667
$E(LR,5)$	0.500	0.462	0.480	0.729
$E(LR,7)$	0.583	0.467	0.519	0.729
$E(LR,9)$	0.667	0.500	0.571	0.750
$E(LR,11)$	0.333	0.364	0.348	0.688
$E(KNN,3)$	0.417	0.455	0.435	0.729
$E(KNN,5)$	0.417	0.714	0.526	0.812
$E(KNN,7)$	0.500	0.667	0.571	0.812
$E(KNN,9)$	0.417	0.625	0.500	0.792
$E(KNN,11)$	0.417	0.714	0.526	0.812
$E(NB,3)$	0.583	0.438	0.500	0.708
$E(NB,5)$	0.583	0.438	0.500	0.708
$E(NB,7)$	0.583	0.438	0.500	0.708
$E(NB,9)$	0.583	0.412	0.483	0.688
$E(NB,11)$	0.667	0.444	0.533	0.708
$E(SVM,3)$	0.500	0.500	0.500	0.750
$E(SVM,5)$	<b>0.583</b>	<b>0.875</b>	<b>0.700</b>	<b>0.875</b>
$E(SVM,7)$	0.500	0.600	0.545	0.792
$E(SVM,9)$	0.417	0.625	0.500	0.792
$E(SVM,11)$	0.417	0.833	0.556	0.833
$E(DT,3)$	0.333	0.444	0.381	0.729
$E(DT,5)$	0.417	0.455	0.435	0.729
$E(DT,7)$	0.500	0.750	0.600	0.833
$E(DT,9)$	0.500	0.750	0.600	0.833
$E(DT,11)$	0.250	0.750	0.375	0.792

표 2는 분류기 종류와 사용된 분류기의 개수에 따른 제안하는 앙상블 모델의 성능 결과를 보여준다. 표현의 간결성을 위해서 설계한 앙상블 모델을  $E(\text{분류기 종류, 사용된 분류기 개수})$  형태로 표기하였다. 여기서 분류기 종류 자리에 적혀있는 기호에 관해 설명하는 다음과 같다.  $LR$ 은 Logistic Regression을 지칭하며,  $KNN$ 은 k-NN을 지칭한다.  $NB$ 는 Naive Bayes을 지칭하고,  $SVM$ 은 C-Support Vector Machine을 그리고  $DT$ 는 Decision Tree를 지칭한다. 사용된 분류기 개수 자리에 적혀있는 숫자 값은 앞서 표 1에서 설명한 내용과 같다. 표 2에서와 같이 TPR, PPV,  $F_1$ , ACC의 4가지 성능 평가지표에 대해서 가장 좋은 성능을 보인 앙상블 모델은  $E(SVM, 5)$ 였다. 표 2의 실험 결과를 바탕으로, 본 연구에서는 5개의 C-Support Vector Machine 분류기들로 구성된 앙상블 모델을 최적의 분노 감정 검출 모델로 선택하였다.

Table 3. Performance Comparisons Between the Proposed Ensemble Model and Five Benchmark Methods

Case	TPR	PPV	$F_1$	ACC
$LR$	0.500	0.400	0.444	0.688
$KNN$	0.417	0.417	0.417	0.708
$NB$	0.583	0.467	0.519	0.729
$SVM$	0.417	0.417	0.417	0.708
$DT$	0.333	0.364	0.348	0.688
<i>Proposed</i>	<b>0.583</b>	<b>0.875</b>	<b>0.700</b>	<b>0.875</b>

표 3은 제안하는 앙상블 모델( $Proposed$ )과 기존 연구들에서 제안된 특징 추출 방법으로 학습된 머신러닝 모델들의 TPR, PPV,  $F_1$ , ACC 성능 비교 결과를 보여준다. 표 3에서 볼 수 있듯이, 제안하는 앙상블 모델은 모든 성능 평가지표에 대해서 가장 좋은 성능을 달성하였다. 가장 좋은 성능을 달성할 수 있었던 이유는, 제안하는 앙상블 모델은 장기부터 중기 그리고 단기까지 다중 시간 윈도우를 통해 추출된 전역적인 특징과 지역적인 특징을 모두 활용하여 학습되었기 때문이다. 반면에 기존 방법으로 학습된 머신러닝 모델들의 경우, 전역적인 특징만으로 학습되었기 때문에 제안하는 앙상블 모델에 비해 낮은 성능을 보였다. 기존 방법 기반의 머신러닝 모델 중에서 가장 좋은 성능을 보인 모델은  $NB$  즉, Naive Bayes 분류기였다. 그리고 가장 낮은 성능을 보인 모델을  $DT$  즉, Decision Tree 분류기였다.

## VI. Conclusions

감정 상태는 보행자의 보폭, 한 보폭에 걸리는 시간, 보행 속도, 목과 흉부의 전방 기울기 각도에 영향을 준다고 알려져 있다. 이에 기존 연구들에서는 상기 특징들을 3D 걸음걸이 데이터에서 추출하여, 분노 감정을 검출하는 데 활용했다. 이때 각 특징은 걸음걸이 데이터 전체 프레임에 걸쳐 항상 균일하게 변화가 발생하는 것이 아니라, 때로는 지역적으로 변화가 발생한다. 이에 본 연구에서는 장기부터 중기 그리고 단기까지 즉, 전역적인 특징과 지역적인 특징을 모두 추출할 수 있는 다중 시간 윈도우 특징 추출 기술을 제안했다. 또한, 제안하는 특징 추출 기술을 통해 추출된 전역 & 지역적인 특징들을 효과적으로 학습하여, 보행자의 감정 상태를 분류할 수 있는 앙상블 모델을 제안했다. 제안하는 앙상블 모델을 구성하고 있는 각 분류기는 서로 다른 다중 시간 윈도우에서 추출된 특징으로 학습된다.



BML에서 제공하는 3D 걸음걸이 데이터 세트에서 실험을 수행한 결과, 제안하는 앙상블 모델은 5개의 C-Support Vector Machine 분류기로 구성되었을 때 최적의 성능을 달성했다. 그리고 기존의 특징 추출 기술로 학습된 머신러닝 모델들과 성능 비교를 수행한 결과, 제안하는 앙상블 모델이 4가지 성능 평가지표 모두에 대해서 최고의 성능을 달성하여 그 우수성을 입증하였다. 본 연구에서는 보행자의 3D 관절 좌표가 모두 완벽하게 제공되는 상황을 가정하고 있다. 하지만, 실제 상황에서는 신체 일부가 가려지는 경우가 발생할 수 있다. 만약 일부 관절이 장애물 등에 의해 가려져 있다면, 본 논문에서 사용한 특징(보폭, 한 보폭에 걸리는 시간, 보행 속도, 목과 흉부의 전방 기울기 각도)을 추출하는 것이 불가능할 수 있다. 그리고 이 경우, 특징 벡터는 Not a Number(NaN) 값이 포함될 수 있으며, 가려짐이 심해질수록 벡터 내 NaN 값의 수가 증가할 수 있다. 이러한 벡터를 분노 감정 검출에 사용하면 앙상블 모델의 성능에 좋지 않은 영향을 미칠 수 있다. 따라서 향후 연구에서는 일부 신체 부위가 가려진 보행자에 대해서도 안정적으로 분노 감정을 검출할 수 있도록, 본 연구에서 제안한 방법을 보완 및 확장할 계획이다.

## ACKNOWLEDGEMENT

This paper is financially supported by Dongyang Mirae University Research Grant in 2022.

## REFERENCES

- [1] N. K. Ciesinski, D. A. Drabick, and M. S. McCloskey, "A Latent Class Analysis of Intermittent Explosive Disorder Symptoms," *Journal of Affective Disorders*, Vol. 302, pp. 367-375, April 2022. DOI: 10.1016/j.jad.2022.01.099
- [2] M. S. McCloskey, E. Y. Chen, T. M. Olino, and E. F. Coccaro, "Cognitive-Behavioral Versus Supportive Psychotherapy for Intermittent Explosive Disorder: A Randomized Controlled Trial," *Behavior Therapy*, Vol. 53, No. 6, pp. 1133-1146, November 2022. DOI: 10.1016/j.beth.2022.05.001
- [3] J. R. Fanning, M. Coleman, R. Lee, and E. F. Coccaro, "Subtypes of Aggression in Intermittent Explosive Disorder," *Journal of Psychiatric Research*, Vol. 109, pp. 164-172, February 2019. DOI: 10.1016/j.jpsychires.2018.10.013
- [4] J. Tseng, M. Nuño, A. V. Lewis, M. Srour, D. R. Margulies, and R. F. Alban, "Firearm Legislation, Gun Violence, and Mortality in Children and Young Adults: A Retrospective Cohort Study of 27,566 Children in the USA," *International Journal of Surgery*, Vol. 57, pp. 30-34, September 2018, doi: 10.1016/j.ijsu.2018.07.010.
- [5] C. Pritchard, M. Parish, and R. J. Williams, "International Comparison of Civilian Violent Deaths: A Public Health Approach to Reduce Gun-Related Deaths in US Youth," *Public Health*, Vol. 180, pp. 109-113, March 2020. DOI: 10.1016/j.puhe.2019.11.003
- [6] A. D. Tasci, and S. Sönmez, "Lenient Gun Laws, Perceived Risk of Gun Violence, and Attitude Towards a Destination," *Journal of Destination Marketing & Management*, Vol. 13, pp. 24-38, September 2019. DOI: 10.1016/j.jdmm.2019.03.007
- [7] B. Kwon, and T. Kim, "Toward an Online Continual Learning Architecture for Intrusion Detection of Video Surveillance," *IEEE Access*, Vol. 10, pp. 89732-89744, August 2022. DOI: 10.1109/ACCESS.2022.3201139
- [8] P. C. Ribeiro, R. Audigier, and Q. C. Pham, "RIMOC, a Feature to Discriminate Unstructured Motions: Application to Violence Detection for Video-Surveillance," *Computer Vision and Image Understanding*, Vol. 144, pp. 121-143, March 2016. DOI: 10.1016/j.cviu.2015.11.001
- [9] G. Sreenu, and S. Durai, "Intelligent Video Surveillance: A Review Through Deep Learning Techniques for Crowd Analysis," *Journal of Big Data*, Vol. 6, No. 1, pp. 1-27, June 2019. DOI: 10.1186/s40537-019-0212-5
- [10] A. B. Mabrouk, and E. Zagrouba, "Abnormal Behavior Recognition for Intelligent Video Surveillance Systems: A Review," *Expert Systems with Applications*, Vol. 91, pp. 480-491, January 2018. DOI: 10.1016/j.eswa.2017.09.029
- [11] B. Kwon, J. Kim, K. Lee, Y. K. Lee, S. Park, and S. Lee, "Implementation of a Virtual Training Simulator Based on 360° Multi-View Human Action Recognition," *IEEE Access*, Vol. 5, pp. 12496-12511, July 2017. DOI: 10.1109/ACCESS.2017.2723039
- [12] Y. B. Singh, and S. Goel, "A Systematic Literature Review of Speech Emotion Recognition Approaches," *Neurocomputing*, Vol. 492, pp. 245-263, July 2022. DOI: 10.1016/j.neucom.2022.04.028
- [13] F. Z. Canal, T. R. Müller, J. C. Matias, G. G. Scotton, A. R. de Sa Junior, E. Pozzebon, and A. C. Sobieranski, "A Survey on Facial Emotion Recognition Techniques: A State-of-the-Art Literature Review," *Information Sciences*, Vol. 582, pp. 593-617, January 2022. doi: 10.1016/j.ins.2021.10.005
- [14] K. Lee, W. Kim, and S. Lee, "From Human Pose Similarity Metric to 3D Human Pose Estimator: Temporal Propagating LSTM Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 2, February 2023. DOI: 10.1109/TPAMI.2022.3164344

- [15] B. Kwon, and S. Lee, "Human Skeleton Data Augmentation for Person Identification over Deep Neural Network," *Applied Sciences*, Vol. 10, No. 14, July 2020, pp. 1-22, DOI: 10.3390/app10144849
- [16] B. Kwon, and S. Lee, "Joint Swing Energy for Skeleton-Based Gender Classification," *IEEE Access*, Vol. 9, pp. 28334-28348, February 2021. DOI: 10.1109/ACCESS.2021.3058745
- [17] B. Kwon, J. Huh, K. Lee, and S. Lee, "Optimal Camera Point Selection Toward the Most Preferable View of 3D Human Pose," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 52, No. 1, pp. 533-553, January 2022. DOI: 10.1109/TSMC.2020.3004338
- [18] B. Kwon, and S. Lee, "Ensemble Learning for Skeleton-Based Body Mass Index Classification," *Applied Sciences*, Vol. 10, No. 21, pp. 1-23, November 2020. DOI: 10.3390/app10217812
- [19] B. Kwon, D. Kim, J. Kim, I. Lee, J. Kim, H. Oh, H. Kim, and S. Lee, "Implementation of Human Action Recognition System Using Multiple Kinect Sensors," in *Proc. 16th Pacific Rim Conference on Multimedia (PCM)*, pp. 334-343, Gwangju, South Korea, September 2015. DOI: 10.1007/978-3-319-24075-6\_32
- [20] B. Kwon, J. Kim, and S. Lee, "An Enhanced Multi-View Human Action Recognition System for Virtual Training Simulator," in *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit Conference (APSIPA ASC)*, pp. 1-4, Jeju, South Korea, December 2016. DOI: 10.1109/APSIPA.2016.7820895
- [21] J. M. Montepare, S. B. Goldstein, and A. Clausen, "The Identification of Emotions From Gait Information," *Journal of Nonverbal Behavior*, Vol. 11, No. 1, pp. 33-42, March 1987. DOI: 10.1007/BF00999605
- [22] E. Crane and M. Gross, "Motion Capture and Emotion: Affect Detection in Whole Body Movement," in *Proceedings of the 2nd International Conference on Affective Computing Intelligent Interaction (ACII)*, pp. 95-101, Lisbon, Portugal, September 2007. DOI: 10.1007/978-3-540-74889-2\_9
- [23] M. Karg, K. Kuhlentz, M. Buss, W. Seiberl, F. Tusker, M. Schmeelk, and A. Schwirtz, "Expression and Automatic Recognition of Exhaustion in Natural Walking," in *Proceedings of IADIS International Conference on Interfaces and Human Computer Interaction (IHCI)*, pp. 165-172, Amsterdam, Netherlands, July 2008.
- [24] M. Karg, R. Jenke, K. Kuhlentz, and M. Buss, "A Two-Fold PCA-Approach for Inter-Individual Recognition of Emotions in Natural Walking," in *Proceeding of the 6th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM)*, pp. 51-61, Leipzig, Germany, July 2009.
- [25] M. Karg, R. Jenke, W. Seiberl, K. Kuhlentz, A. Schwirtz, and M. Buss, "A Comparison of PCA, KPCA and LDA for Feature Extraction to Recognize Affect in Gait Kinematics," in *Proceedings of the 3rd International Conference on Affective Computing Intelligent Interaction and Workshops (ACIIW)*, pp. 1-6, Amsterdam, Netherlands, September 2009. DOI: 10.1109/ACII.2009.5349438
- [26] M. Karg, K. Kuhlentz, and M. Buss, "Recognition of Affect Based on Gait Patterns," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 40, No. 4, pp. 1050-1061, August 2010. DOI: 10.1109/TSMCB.2010.2044040
- [27] B. Li, C. Zhu, S. Li, and T. Zhu, "Identifying Emotions from Non-Contact Gaits Information Based on Microsoft Kinects," *IEEE Transactions on Affective Computing*, Vol. 9, No. 4, pp. 585-591, October-December 2018. DOI: 10.1109/TAFFC.2016.2637343
- [28] P. Xue, B. Li, N. Wang, and T. Zhu, "Emotion Recognition from Human Gait Features Based on DCT Transform," in *Proceedings of International Conference on Human Centered Computing (HCC)*, pp. 511-517, Čačak, Serbia, August 2019. DOI: 10.1007/978-3-030-37429-7\_51
- [29] Y. Zhuang, L. Lin, R. Tong, J. Liu, Y. Iwamoto, and Y. W. Chen, "G-GCSN: Global Graph Convolution Shrinkage Network for Emotion Perception from Gait," in *Proceedings of the 15th Asian Conference on Computer Vision (ACCV)*, pp. 46-57, Kyoto, Japan, November-December 2020. DOI: 10.1007/978-3-030-69756-3\_4
- [30] V. Narayanan, B. M. Manoghar, V. S. Dorbala, D. Manocha, and A. Bera, "ProxEmo: Gait-Based Emotion Learning and Multi-view Proxemic Fusion for Socially-Aware Robot Navigation," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8200-8207, Las Vegas, NV, USA, February 2021. DOI: 10.1109/IROS45743.2020.9340710
- [31] U. Bhattacharya, T. Mittal, R. Chandra, T. Randhavane, A. Bera, and D. Manocha, "STEP: Spatial Temporal Graph Convolutional Networks for Emotion Perception from Gaits," in *Proceedings of the AAAI Conference Artificial Intelligence*, pp. 1342-1350, New York, USA, April 2020. DOI: 10.1609/aaai.v34i02.5490
- [32] Y. Ma, H. M. Paterson, and F. E. Pollick, "A Motion Capture Library for the Study of Identity, Gender, and Emotion Perception from Biological Motion," *Behavior Research Methods*, Vol. 38, No. 1, pp. 134-141, February 2006. DOI: 10.3758/BF03192758

## Authors



Beom Kwon was born in Seoul, South Korea, in 1989. He received the B.S. degree in electrical and electronic engineering from Soongsil University, Seoul, in 2012, and the M.S. and Ph.D. degrees in electrical and

electronic engineering from Yonsei University, Seoul, in 2018. From March 2018 to September 2019, he was a Senior Researcher with the Agency for Defense Development (ADD), Daejeon, South Korea. From October 2019 to August 2021, he was a Staff Engineer with Samsung Electronics Company, Ltd., Suwon, Gyeonggi-do, South Korea. Since September 2021, he has been an assistant professor in the Department of Artificial Intelligence at Dongyang Mirae University, Seoul. His research interests include artificial intelligence and its applications.



Taegeun Oh received the B.S. and Ph.D. degrees in Electrical and Electronic Engineering from Yonsei University, South Korea, in 2007 and 2014, respectively. From 2014 to 2022, he was a Senior Researcher

with the Agency for Defense Development (ADD), Daejeon, Korea South, where he was involved in Avionics and Task Allocation for Unmanned Aerial Vehicles. He joined the faculty of the Department of Electrical and Electronic at Dong Seoul University, Seongnam, Gyeonggi-do, South Korea, in 2022, where he is currently an Assistant Professor. His research interests include image/video quality assessment, computer vision, autonomous driving and task allocation algorithms for multi-vehicle.