

2D 자세 추정 결과를 활용한 규칙 기반 동작 분류 기법

강신재, 한동석*

경북대학교 전자전기공학부

kangsj129@knu.ac.kr, *dshan@knu.ac.kr

Rule-Based Human Action Classification Using 2D Pose Estimation

Shin Jae Kang, Dong Seog Han*

School of Electronic and Electrical Engineering, Kyungpook National Univ.

요약

본 논문에서는 포즈 추정 결과를 활용하여 사람의 동작을 분류하는 규칙 기반 분류기를 제안한다. 자세 추정 모델로부터 추출한 2차원 관절 정보들을 기반으로, 관절 간의 상대 위치, 움직임, 각도 등의 수식적 조건을 정의하여 ‘서기’, ‘걷기’, ‘앉기’, ‘넘어짐’ 총 네 가지 동작을 프레임 단위로 판단한다. 전체 시스템은 데이터 로더, 좌표계 정의, 수식 기반 레이블링, 분류 및 시각화의 흐름으로 구성되며, 각 단계는 단순하지만 명확한 조건에 기반하여 동작을 결정한다. 제안된 시스템은 학습 데이터 없이 동작을 실시간으로 분류할 수 있으며, 특히 차량 내 승객 모니터링, 작업자 행동 감지 등 반복적이고 명확한 동작이 요구되는 환경에서의 적용 가능성이 높다.

I. 서 론

자세 추정 기술은 영상 속 인간의 움직임을 이해하기 위한 핵심 기술 중 하나이다 [1], [2]. 자세 추정은 사람의 신체를 구성하는 주요 관절의 위치를 2차원 또는 3차원 공간상에서 추정함으로써, 이후 동작 분류나 행동 분석 등 고차원적 인식 작업을 위한 기본 정보를 제공한다. 일반적으로 포즈 추정 이후의 행동 인식 단계는 시간에 따른 관절의 변화를 분석하는 과정을 포함하며, 최근에는 이 과정을 심층 신경망 기반의 분류기로 처리하는 연구가 활발히 이루어지고 있다 [3], [4].

그러나 모든 상황에서 딥러닝 기반 분류기를 적용하는 것이 항상 최적의 선택은 아니다. 특히, 행동의 종류가 비교적 제한적이고 명확하게 정의될 수 있는 상황에서는, 오히려 복잡한 학습 기반 모델보다 단순하고 직관적인 규칙 기반 분류기가 더 적합할 수 있다. 본 논문에서는 관절 간 상대 위치, 높이 관계, 움직임의 변화량 등 정량적으로 정의 가능한 조건을 활용하여, 사람의 대표적인 네 가지 동작(서기, 걷기, 앉기, 넘어짐)을 분류하는 규칙 기반 분류기를 설계하였다.

제안하는 분류기는 포즈 추정 결과로부터 수식화 가능한 특징을 추출하고, 이를 조건문 형태로 해석하여 동작을 실시간으로 판단한다. 이러한 방식은 모델 해석 가능성이 높고 데이터 의존성이 낮으며, 시스템 전체의 경량화를 도모할 수 있다는 점에서 실용성이 크다. 특히, 차량 내부 안전 모니터링, 고령자 낙상 감지, 산업 현장의 작업 자세 판별 등과 같이 사전에 정의된 동작을 빠르게 탐지해야 하는 응용 분야에서 활용 가능성이 높다.

II. 규칙 기반 분류기

본 연구에서는 포즈 추정 결과를 입력으로 받아 사람의 동작을 분류하는 규칙 기반 분류기를 제안한다. 제안된 시스템은 다섯 단계로 구성되며, 전체적인 처리 흐름은 그림 1과 같다.

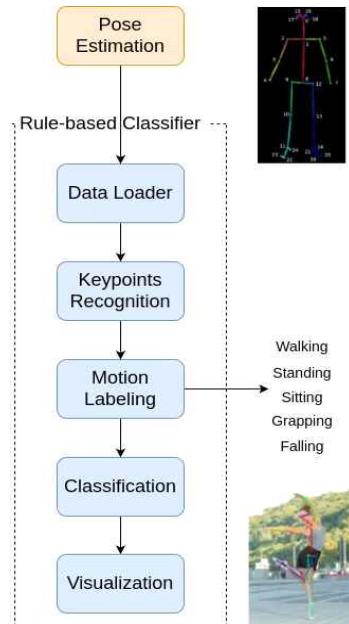


그림 1. 규칙 기반 분류기의 전체적인 처리 흐름

첫 번째 단계인 데이터 로더는 대표적인 자세 추정 모델 중 하나인 OpenPose를 통해 생성된 JSON 포맷의 프레임 단위 관절 데이터를 불러온다 [5]. 이전 프레임의 데이터와 함께 관리하여 시간에 따른 자세 변화 정보를 추출할 수 있도록 한다. 다음 단계인 관절 인지 부분에서는 그림 2에 보이는 것과 같이 각 관절의 위치를 중심으로 수직 축(y 축)을 정의하고, y 축에 수직인 모든 축을 x 축으로 정의하여, 관절 간 상대적인 위치 관계를 계산하는 큰 틀이 된다.

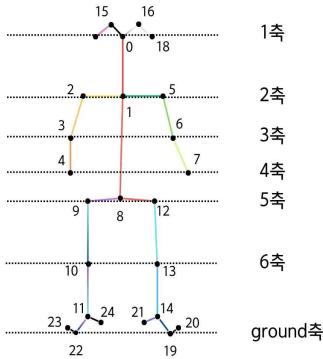


그림 2. COCO 기반 사람 관절의 번호와 관절 좌표계 정의

세 번째 단계인 행동 레이블링 모듈은 추출된 관절 정보를 수학적 조건에 따라 해석하여 동작 레이블을 할당하는 과정이다. 이 단계에서는 관절 간 거리, 높이, 이동량, 각도 등의 수치적 기준을 활용하여 총 네 가지 동작 클래스(서기, 걷기, 앓기, 넘어짐) 중 하나로 분류된다. 표 1에서 조건문으로 풀어 설명한 분류 기준을 확인할 수 있다. 네 번째 단계인 행동 분류 단계에서는 레이블링 된 정보를 바탕으로 최종 동작을 결정한다. 마지막 단계인 시각화 모듈은 입력 영상에 해당 프레임의 관절 정보와 분류된 동작을 시각적으로 중첩하여 결과를 출력한다.

제안 시스템은 복잡한 학습 기반 구조 없이도 간단한 수식적 조건으로 그림 4와 같이 안정적인 동작 인식이 가능함을 보이며, 실시간 응답성과 해석 가능성 측면에서 높은 활용 가능성을 가진다. 특히, 인지해야 할 동작의 범위가 명확하고 반복적인 환경에서는 본 시스템이 효과적으로 작동할 수 있다.

표 1. 자세 추정에 따른 행동 분류 기준 요약

클래스	행동 분류 기준 요약
서기 (엄격)	무릎이 엉덩이보다 아래에 있고, 발은 바닥 가까이에 있으며, 어깨가 엉덩이보다 위에 있음
서기 (완화)	서 있는 형태지만 조금 흐트러져 있어도, 무릎 < 발, 어깨 > 엉덩이 조건이라면 허용
앉기	엉덩이가 무릎보다 약간 위거나 비슷하고, 발이 바닥에 가까우며, 어깨는 엉덩이보다 위에 있음 엉덩이와 무릎 높이가 비슷하고, 어깨가 엉덩이보다 위에 있음
넘어짐	어깨가 엉덩이보다 아래에 있고, 머리와 손이 모두 아래에 위치한 자세
걷기	양발이 서로 멀리 떨어져 있고, 프레임 간 발 위치 변화가 크며, 양 발 높이가 비슷함
감지가 안될 때	주요 관절(머리, 어깨, 골반 등)의 신뢰도가 매우 낮아서 사람이 인식되지 않음
불확실한 경우	위 조건들 중 어느 것도 명확히 만족하지 않아서 분류가 어려운 경우

III. 결론

본 논문에서는 자세 추정 결과로부터 수식 기반의 조건을 활용하여 사람의 동작을 실시간으로 분류할 수 있는 규칙 기반 동작 분류기를 설계하고 구현하였다. 제안된 분류기는 별도의 학습 과정 없이 관절 간의 기하학적 관계만으로 분류가 가능하여, 해석 가능성과 경량성 측면에서 새로운 판점 을 제시한다. 총 네 가지 대표 동작에 대해 명확한 기준을 정의하였으며, 그림 3에서 추정된 자세에 따른 동작을 안정적으로 분류할 수 있음을

확인할 수 있다.

그러나 제안 방식은 동작의 정의가 불분명하거나 복합적인 행동이 동시에 발생하는 상황에서는 분류가 어려울 수 있다는 한계가 존재한다. 따라서 향후 연구에는 본 분류기를 보완하는 딥러닝 기반 분류기와의 하이브리드 구조로 확장하거나, 규칙 조건을 자동 생성 또는 강화 학습 기반으로 최적화하는 방향이 고려될 수 있다. 또한 실시간 시스템 적용을 위한 최적화, 다양한 카메라 환경에서의 일반화 평가, 자세 추정 기반 행동 인지 모델의 레이블링 생성 등 역시 중요한 후속 과제로, 해당 기술의 발전 가능성이 높음을 시사한다.



그림 3. 자세 추정 기반 행동 인지 모델 시각화

ACKNOWLEDGMENT

본 연구는 산업통상자원부와 한국산업기술진흥원의 지역혁신클러스터육성(R&D, P0025274) 사업의 지원을 받아 수행된 연구결과임. This research was financially supported by the Ministry of Trade, Industry and Energy, Korea, under the “Regional Innovation Cluster Development Program(R&D, P0025274)” supervised by the Korea Institute for Advancement of Technology (KIAT).

참고 문헌

- [1] A. Toshev and C. Szegedy, DeepPose: Human Pose Estimation via Deep Neural Networks, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1653 - 1660.
- [2] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7291 - 7299.
- [3] M. Ronald, A. Poulose, and D. S. Han, “iSPLInception: An Inception-ResNet Deep Learning Architecture for Human Activity Recognition,” IEEE Access, vol. 9, pp. 68985 - 69001, 2021, doi: 10.1109/ACCESS.2021.3078184.
- [4] L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, and L. Van Gool, Temporal Segment Networks: Towards Good Practices for Deep Action Recognition, in European Conference on Computer Vision (ECCV), 2016, pp. 20 - 36.
- [5] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, in CVPR, 2017.