3-D 자세 추정 알고리즘 기반 고자유도 인터랙션 분류 시스템 연구

곽민지[2], 조현진[1], 이재윤[2], 최종원[1,2]*

[1]중앙대학교 첨단영상대학원 영상학과, [2]중앙대학교 일반대학원 AI학과

k_minz@vilab.cau.ac.kr, jincho@vilab.cau.ac.kr, leejaeyoon@vilab.cau.ac.kr, *choijw@cau.ac.kr

Research on a Few-shot Interaction Classification System Based on a 3-D Pose Estimation

Minji Kwak, Cho Hyun Jin, JaeYoon Lee, Jongwon Choi*

Chung-Ang Univ., GSAIM, Dept. of Adv. Imaging
Chung-Ang Univ., Graduate School, Dept. of AI

요 약

본 연구는 고자유도 인터랙션 기술을 개발하고, 이를 통해 복잡한 동작을 높은 정확도로 분류하는 시스템을 구축하는 것을 목표로 한다. 기존 행동 분류 시스템은 입력된 영상에서 학습 전 정해진 행동에 대해서만 분류할 수 있기 때문에, 사용자가 새로운 행동을 추가하고 싶을 경우 새롭게 모델을 학습해야 하는 번거로움이 있다. 이러한 문제점을 해결하기 위하여 제안하는 시스템은 3D 자세 추정 알고리즘을 통해 영상 내의 3D 자세를 취득하고, 사용자가 희망하는 자세와의 거리를 계산해 행동 분류를 수행하도록 설계되었다. 이 연구는 또한 복잡한 동작을 분류하기 위해 신체 부위를 나누어 관련된 관절들의 정보만을 활용하는 새로운 접근 방식을 제시한다. 제안하는 알고리즘을 5명 이상의 인원이 행동을 취하는 영상 기반으로 검증하였으며, 이를 통해 안정적인 성능을 보이는 것을 확인한다.

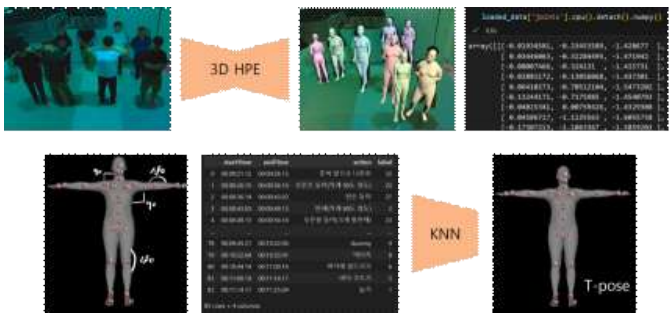

그림 1. 제안하는 시스템 파이프라인 개요도

Ⅰ. 서 론

최근 인터랙션 중심의 미디어아트 작품 개발과 전시가 활성화됨에 따라 다양한 종류의 인터랙션을 영상 기기로 인식하는 기술의 활용도가 증가하고 있다 [1,2]. 이러한 인터랙션의 경우, 미디어 아트에서 요구되는 행동의 종류가 제한적이기 때문에 이미지를 입력받으면 사용자가 목표로 하는 행동을 구분하도록 하는 행동 분류 알고리즘을 사용한다. 최근 딥러닝의 기술 발전으로 이러한 행동 분류 알고리즘의 성능이 크게 향상되었으며 [3], 이로 인해 저렴한 카메라 장비만으로도 동작 가능한 인터랙션 미디어 아트 작품이 개발되고 있다.

하지만, 기존 행동 분류 알고리즘은 사용자가 미리 정해둔 자세 혹은 행동의 인터랙션만 분류할 수 있기 때문에, 새로운 자세나 행동 인터랙션을 적용하기 위해서는 새로운 모델의 학습이 필요하다. 미디어 아트의 설치 환경이나 대상자의 변경에 따라 자세와 행동 인터랙션은 변경될 수 있으며, 사용자의 필요에 따라서도 손쉽게 바꿀 수 있는 기능이 필요하지만, 기존의 알고리즘은 이를 위해 많은 시간과 비용이 소요된다.

본 연구에서는 이러한 문제를 해결하기 위해, 적은 개수의 자세와 행도 인터랙션 데이터 만으로 새로운 인터랙션을 분류할 수 있는 시스템을 개발한다. 개발된 시스템에는 데이터 전처리 파이프라인을 구축하여 체험자의 위치 정보, 신체 정보 등을 수집하고, 이를 활용하여 동작 분류 모델을 개발한다. 우선 체험자의 3차원 신체 자세를 추정하고, 이를 바탕으로 정답 3차원 신체 자세와의 거리를 계산해 가장 가까운 인터랙션을 출력하는 동작 분류 알고리즘을 사용한다. 정확도를 향상시키기 위해 본 연구에서는 체험자의 신체 부위를 나누어 각 부위에 해당하는 동작을 분류하는 새로운 접근 방식을 제시한다. 제안된 시스템을 실제 미디어아트 작품을 위해 촬영된 동영상에서 테스트를 수행하였으며, 적은 개수의 정답 데이터 만으로도 안정적인 성능을 보이고 있음을 확인하였다.

Ⅱ. 본론

○ 시스템 파이프라인 설명

본 시스템 파이프라인은 3차원 신체 자세 추정과 인터랙션 분류의 두가지 단계로 나뉘며, 이는 그림 1에서 보이는 개요와 같다.

- 3차원 신체 자세 추정

우선, Bedlam-cliff [4] 를 활용하여 영상 내 체험자들의 신체 데이터 취득한다. 비디오를 입력으로 넣어 이미지 프레임으로 쪼갠 후, 프레임마다 사람의 정보를 취득하여, 월드 좌표계(world-pose)와 SMPL (Skinned Multi-Person Linear model) [5] 신체 쉐잎(shape), SMPL 신체 포즈(shape)을 저장한다.

앞서 얻은 신체 데이터 중 카메라 파라미터를 적용하여 얻은 각 체험자의 골반 픽셀 좌표와 3D 신체 모델인 SMPL의 파라미터 중 각 관절에 대한 각도를 의미하는 포즈 파라미터 추출한다. 이때 포즈 파라미터 중 첫 번째 값은 해당 카메라 각도에서 SMPL의 전체 회전 각도를 의미하므로 다양한 카메라 각도의 영상으로부터 일관된 신체 정보를 얻기 위하여 제외하고, 얼굴과 손가락 관절 등에서도 노이즈가 많이 발생하므로 역시 제외하여 총 23개의 주요 관절만을 활용한다. SMPL은 키네마틱 트리를 기반으로 하는 관절의 상대적인 각도를 제공하기 때문에 다양한 카메라 각

도의 영상에서도 일관된 관절 각도 데이터를 획득할 수 있다.

- 인터랙션 분류

신체의 전체 관절에 대한 각도를 모두 활용하여 K-Nearest Neighbor 분류 알고리즘을 학습시켰을 때, 성능이 좋지 않은 복잡한 동작을 분류한다. 복잡한 동작에 대해 해당 동작과 관련이 있는 관절들의 정보만을 활용하기 위하여, 신체 부위를 팔과 몸통 윗부분, 다리와 몸통 아랫부분 그리고 머리로 나누어 관절 데이터 분류한다.

분류한 관절에 대한 데이터만을 학습 데이터로 활용하여 K-Nearest Neighbor 분류 알고리즘을 학습한다. 따라서, 새로운 영상에 대하여 특정 신체 부위의 관절 각도가 일치하면 해당 동작을 수행한 것으로 인식하며, 그 결과 상체와 팔의 관절들을 활용하여 '머리 위에 하트'와 같은 다관절을 활용한 1개의 복잡한 동작 분류할 수 있다.

행동 분류를 위한 KNN(K-Nearest Neighbor) 알고리즘의 효율성을 증진하기 위해, 본 연구에서는 신체 자세 추정 모델로부터 예측된 신체 관절 데이터 중 이상치를 제거했다. 신체 자세 추정 모델은 카메라 각도나 조명 등 환경적 요인의 영향을 받아 이상치가 발생할 수 있으며, 이러한 이상치는 KNN 알고리즘의 분류 정확도를 저하시킨다.

이 문제에 대응하기 위해, 본 연구에서는 Isolation Forest [6] 알고리즘을 적용한다. Isolation Forest는 비지도 학습 방식의 앙상블 기반 이상 탐지 알고리즘으로, 다수의 결정 트리를 종합적으로 활용한다. 이 알고리즘은 데이터 포인트를 분할하여 모든 데이터 포인트를 고립시키는데, 이때 각 데이터 포인트를 고립시키는 데 필요한 분할 수를 기준으로 이상치를 판별한다. 정상 데이터는 많은 분할을 거쳐 정상 데이터로 결정되는 반면, 이상치는 적은 분할로도 빠르게 고립되어 이상치로 결정된다.

Isolation Forest를 적용하기 위해서 먼저 데이터 중 일부를 비복원 추출하고, 그 중 몇 개의 데이터를 일정 비율만큼 무작위로 선택한다. 그 다음, 선택된 데이터들의 최솟값과 최댓값을 추출하여 해당 범위 안에서 무작위로 분할 지점을 선택하고, 무작위로 선택된 분할 지점을 기반으로 데이터를 분할한다. 분할된 그룹에 대해서는 재귀적으로 위의 과정을 반복한다. 이를 통해 결정 트리가 계속 분할되어 각각의 경로가 생성되고, 정상 데이터는 트리의 말단에 위치하고 이상치는 분할이 많이 되지 않아 트리의 시작점에 가까운 곳에 위치하게 된다. 이 방식을 적용하여, 다양한 포즈에서 추출된 SMPL 관절 각도 데이터에 Isolation Forest 알고리즘을 학습시켜, 이상치를 효과적으로 식별했다. 이후, 이상치로 분류된 관절 각도 값들은 데이터셋에서 제거했다.

정제된 데이터를 사용함으로써 KNN 기반 행동 분류 알고리즘의 전반적인 정확도가 향상되었으며, 이를 통해 저조도 환경과 같이 변동성이 큰 환경에서도 정확한 신체 관절 각도를 얻을 수 있었다. 교차 검증을 사용하여 다양한 K값에 대한 K-Nearest Neighbor 분류 알고리즘의 성능을 평가하고, 각 행동 분류에 대한 정확도가 최대가 되는 K값 선택하였다. 또한, 시간 축에 따른 안정적인 분류를 위하여 해당 동작이 연속으로 3개 프레임 동안 지속되면 해당 동작을 수행하고 있는 것으로 인식하였다.

○ 시스템 파이프라인 검증

체험자의 행동을 분류하기 위한 학습 데이터를 확보하기 위하여 6가지 카메라 각도에서 동시에 촬영된 영상 데이터 활용하였다. 체험자 6명의 개인 인터렉션에 대한 총 14개의 동작(준비 앞으로 나란히, 오른손 들어, 왼손 들어, 작은 만세, 오른팔 들어, 왼팔 들어, 큰 만세, 머리위에 하트, 양 뺨에 손, 오른 뺨에 손, 왼 뺨에 손, 오른팔 오른쪽으로 펴기, 왼팔 왼쪽으로 펴기, 양팔 양쪽으로 펴기)으로 분류하였으며, 해당 영상들을 이미지 프레임으로 나눈 후, 각 프레임마다 행동에 대한 라벨을 지정하여 라벨을 정확하게 출력시키는 정확도를 계산하였다. 이 때, 동작이 변경되거나 행동이 명확하지 않은 시간대의 프레임은 제외하였다.

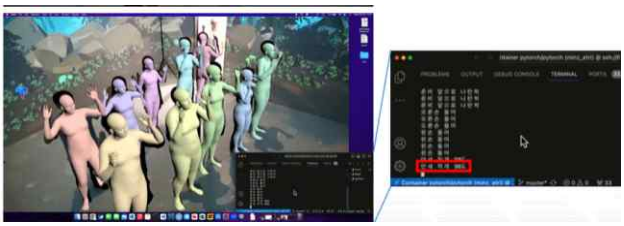

그림 2. 정성적인 결과 예시 그림

동작	Acc	동작	Acc	동작	Acc
준비	94.4	왼팔	92.8	왼뺨	75.0
오른손	84.7	큰만세	91.9	오른펴기	92.7
왼손	83.2	하트	85.3	왼펴기	94.0
작은만세	82.4	양뺨	76.2	양펴기	93.2
오른팔	92.4	오른뺨	71.5	평균	84.6

표 1 동작 분류 정량적 결과

그 결과는 그림 2에서 보이는 것과 같이 정성적으로 우수한 성능을 보이고 있다. 이는 단순한 웹캠 카메라만을 활용하여 얻은 결과임에도 불구하고 새로운 인터랙션에 대한 우수한 대응성을 보여주고 있다. 동일 동영상에 대해 정확도를 추출한 결과는 표 1과 같다. 모든 동작 분류에 대해 최소 71.5% 이상의 정확도를 보이고 있으며, 평균 84.6%의 정확도를 보여 준수한 성능을 보이고 있다.

Ⅲ. 결론

본 논문은 새로운 자세와 행동의 인터랙션에 대응하기 위한 고자유도 인터랙션 시스템을 제안한다. 기존 알고리즘은 고정된 인터랙션에만 대응이 가능했다면, 제안된 시스템에서는 새로운 자세의 출현에도 손쉽게 대응 가능한 장점이 있다. 제안하는 시스템을 아직 자체 동영상에서만 실험을 수행하여 일부 검증하였으나, 향후 연구에서는 다양한 환경과 데이터셋에 대해서 검증을 수행하여 알고리즘의 범용성을 확인할 예정이다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국 콘텐츠진흥원의 2023년도 문화기술 연구개발 사업(과제명 : 저조도 조명환경 극복을 위한 다중 투사 공간 활용 고자유도 대규모 사용자 상호작용 기술개발, 과제번호 : RS-2023-00222280, 기여율 : 90%)과 2023 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-01341, 인공지능대학원 지원(중앙대학교))

참 고 문 헌

[1] Y. Gong, "Application of virtual reality teaching method and artificial intelligence technology in digital media art creation," Ecological Informatics, 2021.

[2] N. Anatrasirichai & D. Bull, "Artificial intelligence in the creative industries: a review," Artificial Intelligence Review, 2021

[3] P. Pareek & A. Thakkar, "A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications," Artificial Intelligence Review, 2020

[4] M. J. Black, P. Patel, J. Tesch & J. Yang, "BEDLAM: A Synthetic Dataset of Bodies Exhibiting Detailed Lifelike Animated Motion." CVPR. 2023.

[5] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll & M. J. Black, "SMPL: A Skinned Multi-Person Linear Model," ACM Trans. on Graphics, 2015

[6] F. T. Liu, K. M. Ting & Z. Zhou, "Isolation Forest," IEEE ICDM, 2008.