

휠체어 간섭 환경에서 보치아 선수의 자세추정을 위한 CNN-Transformer 하이브리드 모델 제안

정소현, 한요섭
숭실대학교

jshyun0527@gmail.com, yoseob.han@ssu.ac.kr

A CNN-Transformer Hybrid Model for Pose Estimation of Boccia Athletes in Wheelchair-Interference Environments

Sohyun Jeong, Yoseob Han
Soongsil University

요약

스포츠 분야에서의 영상분석 기술은 경기 중 판정은 물론, 자세교정 등 훈련보조에 활용되고 있다. 이를 위해 선수의 자세추정 기술이 주요하게 활용되고 있으나, 기존 연구는 주로 비장애인을 대상으로 하여 휠체어 등 보조기구와 선수의 신체 일부가 중첩되고, 관절 위치의 비정형성 등의 이유로 장애인 스포츠에 적용할 경우 성능 저하가 발생한다.^[1,2] 본 논문에서는 이러한 문제를 해결하기 위해 Transformer 기반 자세추정 방법인 PCT^[3]를 통해 계층적 신체구조 관계와 포즈 사전정보를 추론하고, CNN 기반 AlphaPose^[4]의 히트맵 예측 단계에 반영하여 휠체어 간섭 및 가립 환경에서도 안정적인 추정을 가능하게 하는 하이브리드 구조를 제안한다. 제안한 모델은 AlphaPose 단일 모델 대비 4.28%, PCT 단일 모델 대비 16.27% 성능 향상을 확인하였다.

I. 서론

스포츠 분야의 영상분석 기술은 육안으로 확인하는 단순 비디오 판독을 벗어나 구기종목에서 공의 실시간 추적, 경기규칙과 선수의 움직임을 바탕으로 경기상황을 분석하여 심판판정 보조, 자세교정 등 훈련상황의 보조 목적으로 영상분석 기술이 도입되고 있다. 이러한 영상분석 기술에서 개별선수의 움직임을 추적하는 것이 중요하고, 자세추정 기술이 주요하게 활용되고 있다. 한편, 이러한 자세추정 기술은 비장애인 스포츠만 아니라, 장애인 스포츠에 도입될 경우, 영상분석 결과에 대한 시각화를 통해 보다 효과적인 의사소통을 가능하게 하고, 훈련효과를 극대화하는 데 기여할 수 있다. 그러나, 기존 연구는 주로 비장애인을 대상으로 하여 장애인 스포츠에 적용할 경우, 휠체어 등 운동보조기구와 선수의 신체가 일부 중첩되고, 관절위치의 비정형성으로 인하여 히트맵 기반 모델에서 키포인트가 불안정하게 추정되는 문제가 발생한다.^[1,2] 본 연구는 CNN 기반의 히트맵 회귀 방식인 AlphaPose^[4]와 Transformer 기반 PCT^[3]의 구조적 자세추정을 결합한 하이브리드 모델을 제안한다.

II. 본론

A. 관련연구

자세추정은 딥러닝 기술의 발전과 함께 높은 정확도를 달성하고 있고, 히트맵 기반 방법, 토큰 및 그래프 기반 방법 등이 제안되었다. 키포인트 예측에 사용되는 CNN 기반 히트맵 회귀방식은 공간적 지역성을 활용하여 표준적인 접근법으로 자리잡아^[5], Alpha-Pose 등을 통해

그 성능이 입증되었다. 그러나 이러한 CNN 기반 방법은 휠체어, 보조기구 등에 의해 관절 주변의 시각적 단서가 왜곡되거나 가려지면서 탐지된 키포인트가 객체영역을 벗어나는 등 왜곡에 취약하고, 키포인트의 절대위치보다는 상대적 관계와 구조적 제약을 고려한 추정이 요구된다. Graph-PCNN^[6], HKE-GCN^[7]등은 히트맵으로 추정 후 그래프기반 정제를 수행하지만 키포인트 수준 좌표에 대한 보정에 머무르는 한계가 있다. C2FNet^[8], CFN^[9]등의 경우 히트맵, 오프셋과 같이 Dense map 형태로 제공되는 사전정보를 활용하여 예측 정확도를 향상시키는 구조를 제안하였으나, 사전정보가 Dense map에 국한되어 전역적 인체구조의 제약을 반영하는데 한계가 있다. 최근, Transformer 기반 모델은 이러한 한계에 대한 보완 가능성을 제시하고 있다. PCT^[3]의 경우, 인체를 독립적인 관절이 아닌 토큰관계로 표현하여 신체 일부가 가려진 상황에서도 좋은 성능을 보인다.

본 논문에서는 PCT의 관절 간 상호 의존성을 반영하여 가려짐 등의 상황에서 보다 안정적인 결과를 얻는 이점을 고려하여, PCT를 통해 계층적 신체구조 관계와 포즈 사전정보를 추론하고 AlphaPose의 키포인트 예측단계에 반영하는 방법을 제안한다.

B. 제안 방법

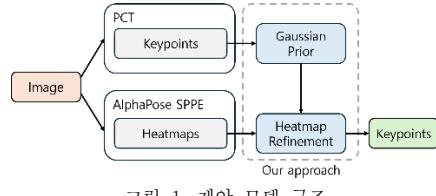


그림 1. 제안 모델 구조

본 논문에서는 CNN-Transformer 하이브리드 구조를 제안한다. 제안방법은 기존 자세추정 모델의 구조를 변경하지 않고, 히트맵 수준에서 보정을 수행하는 방식으로, PCT로부터 예측된 키포인트를 가우시안 분포 형태의 사전정보(prior)로 변환하여 AlphaPose로부터 생성된 히트맵을 보정하는 방식으로 예측을 수행한다. 다음 수식(1)은 2D 가우시안 함수를 바탕으로, PCT로 예측된 관절 j 에 대한 공간적 사전정보로 변환하는 방법을 나타낸다. 이때, (x, y) 는 관절 j 의 위치를, (u, v) 는 히트맵을 나타내며, σ 는 사전정보의 공간적 확산 정도를 제어한다. 생성된 사전정보는 히트맵과 스케일 정합을 위해 최대값으로 정규화 되고, 신뢰도가 낮은 키포인트는 보정과정에서 제외되어 오류를 방지한다.

$$G_j(u, v) = \exp\left(-\frac{(u - x_j)^2 + (v - y_j)^2}{2\sigma^2}\right) \quad (1)$$

다음으로 사전정보를 이용하여 관절예측에 따른 히트맵의 보정을 실시한다. 다음 수식(2)은 AlphaPose SPPE로부터 예측된 각 관절에 대한 히트맵 H_j 에 대하여 사전정보 G_j 를 선형결합하여 보정된 히트맵 H'_j 을 생성함을 나타낸다.

$$H'_j = (1 - w_j) H_j + w_j G_j \quad (2)$$

여기서 w_j 는 PCT 키포인트의 신뢰도에 기반한 가중치로, AlphaPose의 예측구조를 유지하면서 PCT가 제공하는 단서를 확률적으로 반영할 수 있다. 보정된 히트맵으로부터 최종 키포인트의 결정은 AlphaPose의 변환방식을 그대로 사용하여 추출하였다.

C. 실험

제안방법과 기존방법간 비교평가를 위해 유튜브를 통해 공개된 패럴림픽 영상 등을 수집하였고, 보치아 선수의 투구장면을 중심으로 카메라 포즈 및 휠체어 간섭, 투구동작으로 인한 가려짐을 포함하도록 구성하여 모델훈련과 검증에 활용하였다.

표 1 모델 성능 비교

| Method | AP | AP ⁵⁰ | AP ⁷⁵ |
|--------------------------|-------|------------------|------------------|
| AlphaPose ^[4] | 59.23 | 97.92 | 57.27 |
| PCT ^[3] | 47.24 | 89.12 | 32.86 |
| Our | 63.51 | 99.01 | 62.07 |

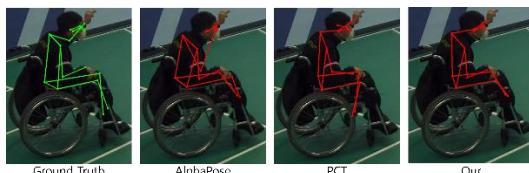


그림 2. 휠체어에 가려진 관절에 대한 예측 비교

실험결과, 그림 2는 제안방법과 기존 방법을 적용한 결과를 나타낸다. 히트맵 보정되지 않은 기존 AlphaPose는 투구동작 중 카메라 앵글에 의해 가려진 좌측 팔꿈치의 키포인트가 예측범위를 벗어났으며, PCT 방식은 이에 대한 예측은 잘 수행하였으나, 구부려져 겹쳐진 두 다리의 관절부의 예측에 어려움을 보였다. 반면, 제안방법은 팔꿈치와 다리 관절을 비교적 정확히 추정함을 보였다. 정량적 평가 또한 표 1과 같이 제안방법이 기존 AlphaPose 대비 4.28%, PCT 대비

16.27% 향상되어, 장애인 스포츠와 같은 비정형적 자세와 가려짐이 발생하는 환경에 보다 안정적인 자세추정이 가능함을 확인하였다.

III. 결론

본 논문에서는 비장애인을 대상으로 훈련된 기존 방법을 장애인 스포츠에 적용할 때의 자세추정 문제를 해결하기 위하여 기존 히트맵 기반 CNN 방식과 Transformer 기반 자세추정 방식을 결합한 하이브리드 방법을 제안하였다. 이를 위한 제안방법은 PCT^[3]와 AlphaPose^[4]를 바탕으로 구조적 포즈 사전 정보를 활용하여 가우시안 분포로 보정된 히트맵을 통한 키포인트 예측을 수행함으로써, (1) 간섭 및 관절 가림 상황에서도 안정적인 자세 추정을 가능하게 하고, (2) 비정형 포즈에 대한 일반화 성능 개선할 수 있다. 향후 연구에서는 투구동작만 아니라 다양한 동작에 적용하여 제안 방법의 유효성을 검증하고, 일반화 성능을 개선하여 보조기구를 사용하는 다양한 장애인 스포츠 영상에 적용하고자 한다.

참 고 문 헌

- [1] W. Huang et al., "WheelPose: Data synthesis techniques to improve pose estimation performance on wheelchair users," in Proc. CHI Conf. Human Factors in Computing Systems, Honolulu, HI, USA, May 2024, pp. 1-25.
- [2] H.-J. Bae et al., "Semantic occlusion augmentation for effective human pose estimation," The Transactions of the Korea Information Processing Society, vol. 11, no. 12, pp. 517-524, 2022.
- [3] Z. Geng et al., "Human pose as compositional tokens," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, June 2023, pp. 660-671.
- [4] H.-S. Fang et al., "AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 6, pp. 7157-7173, June 2023.
- [5] C. Zheng et al., "Deep learning-based human pose estimation: A survey," ACM Computing Surveys, vol. 56, no. 1, pp. 1-37, 2023.
- [6] J. Wang et al., "Graph-PCNN: Two-stage human pose estimation with graph pose refinement," in Proc. European Conf. Computer Vision (ECCV), Glasgow, UK, Aug. 2020, pp. 492-508.
- [7] H. Xia et al., "HKE-GCN: Heatmaps-guided keypoints encoder and graph convolutional network for human pose estimation," in Proc. Int. Joint Conf. Neural Networks (IJCNN), Padua, Italy, July 2022, pp. 1-8.
- [8] J. Sun et al., "C2FNet: Coarse-to-fine keypoint localization network for monocular 6D object pose estimation," in Proc. China Automation Congress (CAC), Chongqing, China, Nov. 2023, pp. 5306-5311.
- [9] S. Huang et al., "A coarse-fine network for keypoint localization," in Proc. IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, Oct. 2017, pp. 3047-3056.