스윙 모션 사전 지식을 활용한 정확한 야구 선수 포즈 보정

오승현[1], 김희원[2]
[1]숭실대학교 글로벌미디어학부 학부생
[2]숭실대학교 글로벌미디어학부 교수

ojames@soongsil.ac.kr, hwkim@ssu.ac.kr

Motion Prior-Guided Refinement for Accurate Baseball Player Pose Estimation

Seunghyun Oh[1], Heewon Kim[2]
[1]Global School of Media, Soongsil University
[2]Global School of Media, Soongsil University

요 약

현대 야구에서 타자의 스윙 패턴 분석은 상대 투수가 투구 전략을 수립하는데 상당히 중요하다. 이미지 기반의 인간 포즈 추정(HPE)은 대규모 스윙 패턴 분석을 자동화할 수 있다. 그러나 기존의 HPE 방법은 빠르고 가려진 신체 움직임으로 인해 복잡한 스윙 모션을 정확하게 추정하는 데 어려움이 있다. 이러한 문제를 극복하기 위해 스윙 모션에 대한 사전 정보를 활용하여 야구 선수의 포즈를 보정하는 방법(BPPC)을 제안한다. BPPC는 동작 인식, 오프셋 학습, 3D에서 2D 프로젝션 및 동작 인지 손실 함수를 통해 스윙 모션에 대한 사전 정보를 반영하여 기성 HPE 모델 결과를 보정한다. 실험에 따르면 BPPC는 벤치마크 데이터셋에서 기성 HPE 모델의 2D 키포인트 정확도를 정량적 및 정성적으로 향상시키고, 특히 신뢰도 점수가 낮고 부정확한 키포인트를 크게 보정했다.

1. 서론

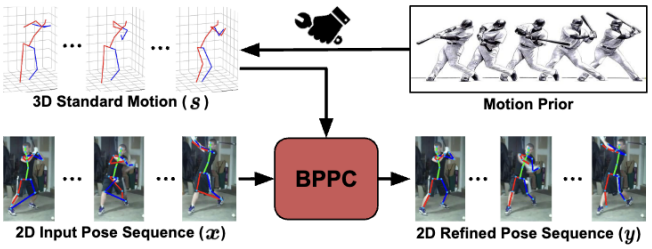
인간 포즈 추정(Human Pose Estimation, HPE)은 컴퓨터 비전 분야에서 중요한 응용 분야 중 하나로, 이미지 또는 동영상에서 사람의 위치와 자세를 추정하는 기술이다. 최근 합성곱 신경망이 빠르게 발전하며 인상적인 HPE 성능을 제공해왔다[1, 2, 3]. 그러나, 기존 모델들은 야구 선수의 스윙과 같이 움직임이 빠르고 심각한 가림이 있는 경우에 잘 동작하지 않는다는 문제를 가진다.

이를 해결하기 위해, 우리는 야구 선수의 스윙 포즈를 보정하는 방법(Baseball Player Pose Corrector, BPPC)을 제안한다. BPPC는 학습 데이터 없이 간단하고 효과적으로 기존 HPE 모델의 추정 결과를 보정한다. BPPC는 3차원으로 모델링 된 야구 스윙에 대한 표준 모션을 임의의 타자 스윙 동작에 투사한다. 투사된 표준 동작과 HPE 결과를 동작의 신뢰도를 고려하여 최적화한다.

실험에서 프로 야구 데이터셋과 단일 사람 포즈 추정 표준 벤치마크에서 다양한 HPE 모델에 대해 정량적 성능 개선과 정성적 성능 개선을 보였다.

본 연구의 기여는 다음과 같다:

- 야구 스윙에 대한 기존 HPE 방법의 추정 결과를 보정하는 간단하고 효과적인 방법을 제안한다.



(그림 1) 제안된 야구 스윙 보정 방법(BPPC) 개요. 인간 포즈 추정(HPE) 모델의 2D 키포인트 결과에 사전에 준비된 3차원 표준 스윙 모션을 적용해 결과를 보정한다.

- 대규모 이미지 학습 기반 HPE 방법이 내재하고 있는 예측 부정확성을 사전 정보 기반 최적화 기법으로 개선한다.
- 사전에 준비된 3차원 표준 모션을 임의의 2차원 스윙 모션에 투사하는 최적화 방법을 제안한다.
- 투사된 표준 모션과 기존 HPE 결과를 비교하여 개선된 2D 키포인트 추정 방법을 제안한다 .
- 일반인과 프로 선수의 스윙을 포함한 벤치마크 데이터셋에서 최신 HPE 모델의 정량적 성능과 정성적 성능을 개선한다.

2. 제안 방법

BPPC는 표준 모션(3D Standard Motion)이라고 명명한 스윙 모션에 대한 사전 지식(Motion Prior)을 활용하여 기존 HPE 모델에 의해 추정된 야구 스윙에 대한 2D 키포인트를 보정한다(그림 1 참조).

BPPC는 $\boldsymbol{x} \in \mathbb{R}^{\hat{F}\times K\times 2}$로 표시된 시험 비디오로부터 얻어진 2D 포즈 추정 결과와 $\boldsymbol{s} \in \mathbb{R}^{F\times K\times 3}$로 표시된 표준 모션을 나타내는 3D 키포인트를 입력으로 사용한다. $\boldsymbol{K}$는 키 포인트의 수를 나타내고 $\hat{\boldsymbol{F}}$와 $\boldsymbol{F}$는 각각 $\boldsymbol{x}$와 $\boldsymbol{s}$에 대한 프레임 수를 나타낸다. 추정된 2D 포즈($\boldsymbol{x}$)는 스윙 모션 전후의 상태를 포함하며 모션 블러 및 가려짐으로 인해 타격 포인트를 감지하는 데 자주 어려움을 겪는다. 반면 3D 표준 모션($\boldsymbol{s}$)은 고속 카메라에 의해 촬영된 이미지를 활용하여 인간 주석자가 스윙 모션의 시작과 끝을 선택하고 3차원 키포인트를 제작한다.

BPPC는 보정된 2D 포즈 $y \in R^{\hat{F}\times K\times 2}$를 학습 데이터 없이 최적화한다. 최적화 목표는 다음과 같다:

$$\min_{\boldsymbol{y}} \mathcal{L}_{data}(\boldsymbol{x}, \boldsymbol{y}) + \lambda \mathcal{L}_{prior}(\boldsymbol{s}, \boldsymbol{y}),$$

여기서 데이터 항($\mathcal{L}_{data}$)은 시험 포즈 시퀀스($\boldsymbol{x}$)를 활용하고 사전 정보 항($\mathcal{L}_{prior}$)은 표준 모션($\boldsymbol{s}$)을 사용한다. 기울기 하강을 통해 $\mathcal{L}_{prior}(\boldsymbol{s}, \boldsymbol{y})$를 최적화하기 위해 $\boldsymbol{s}$와 $\boldsymbol{x}$를 정렬하는 미분 가능한 4D(3D 공간 + 시간) 투영 방법을 도입한다.

3. 실험 결과

데이터셋. 우리는 MLB-YouTube 데이터셋[4]과 Penn Action 데이터셋[5]을 활용하여 BPPC를 평가한다. MLB-YouTube 데이터셋은 유튜브에서 시청할 수 있는 2017년도 MLB 포스트시즌 야구 경기 20개로 구성된 대규모 데이터셋으로 42시간 이상 길이의 영상을 제공한다. Penn Action 데이터셋에는 15가지 액션의 2,326개 비디오 시퀀스와 각 시퀀스에 대한 사람의 공동 주석이 포함되어 있다. 이들 중 스윙에 대한 160개의 비디오 시퀀스를 포함한 Penn Action Swing 데이터셋을 준비하여 BPPC의 성능 평가를 위해 사용한다.

평가 지표. 추정된 키포인트의 정량적 평가를 위해 Percentage of Correct Keypoints at Head scale(PCKh)를 사용했다. 이는 특정 임곗값 이하의 오차를 가진 키포인트를 올바르게 감지된 것으로 간주한다. 우리는 추정된 키포인트가 정답 키포인트로부터 머리 크기의 50% 이내에 위치할 때 정확한 추정으로 판단하였다.

정량적 비교. 표 1은 신뢰도에 따른 키포인트 정확도를 나타낸다. 신뢰도가 낮을수록 키포인트 추정 성능이 떨어지는 경향이 있다. BPPC는 낮은 신뢰도에서 뚜렷한 성능 향상을 보인다. DARK (HRNet-W32)와 비교했을 때, BPPC는 신뢰도 0.5 미만에서 7.4%, 신뢰도 0.9 이상에서 0.2%의 성능이 향상된다. 이는 BPPC가 기존 모델이 헷갈리는 결과를 훌륭히 보완하는 것을 의미한다.

정성적 비교. 그림 2는 MLB-YouTube 데이터셋에서 HRNet-W48[2]의 결과(a)와 이를 BPPC로 보정한 이미지 결과(b)를 보여준다. 빨간 동그라미로 표시된 영역을 보면, HRNet-W48은 흐리게 보이는 야구 방망이를 선수의 팔로 인식한 반면, HRNet-W48+BPPC는 앞으로 뻗어진 선수의 팔을 정확히 추정하였다.

<표 1> Penn Action Swing 데이터셋에서 키포인트 신뢰도에 따른 HPE 모델[1, 2, 3]의 성능 향상 결과

Methods	PCKh@0.5					
	[0, .5)	[.5, .6)	[.6, .7)	[.7, .8)	[.8, .9)	[.9, 1]
Simple (ResNet-50)	54.8	71.4	78.6	90.7	97.1	94.6
+ BPPC (Ours)	**59.3**	**72.4**	**80.3**	**90.9**	**97.2**	94.6
Simple (ResNet-101)	56.3	59.0	74.8	90.8	95.9	100.0
+ BPPC (Ours)	**60.4**	**60.5**	**75.6**	**91.0**	**96.1**	100.0
Simple (ResNet-152)	57.4	62.2	68.0	91.8	97.8	98.1
+ BPPC (Ours)	**64.7**	**64.3**	**69.6**	**92.1**	97.8	98.1
HRNet-W32	49.1	63.4	79.7	94.9	97.3	98.0
+ BPPC (Ours)	**51.2**	**65.2**	**81.2**	**95.0**	**97.4**	**98.1**
HRNet-W48	39.1	60.9	70.0	89.3	96.9	98.1
+ BPPC (Ours)	**44.2**	**63.8**	**71.1**	**89.7**	**97.0**	**98.2**
DARK (HRNet-W32)	48.7	53.6	73.5	88.7	97.0	98.7
+ BPPC (Ours)	**53.6**	**56.7**	**74.1**	**89.0**	**97.1**	**98.9**
DARK (HRNet-W48)	41.4	66.4	71.8	89.8	97.7	99.4
+ BPPC (Ours)	**48.8**	**67.6**	**73.8**	**90.3**	**97.9**	99.4

(a) HRNet-W48 (b) HRNet-W48 + BPPC (Ours)

(그림 2) BPPC 적용 모델의 시각적인 결과 비교

4. 결론

본 연구는 표준 동작에 대한 사전 정보를 활용하여 야구 스윙에 대한 키포인트 추정 결과를 보정하는 방법을 제안한다. 제안 방법은 기존의 학습 기반 방식의 흔들리거나 폐색이 큰 이미지에서 부정확한 예측 결과를 사전 정보 기반 최적화 기법을 사용하여 보정한다. 실험 결과는 제안 방법이 다양한 기존 모델들의 성능을 실제 스윙 영상에서 향상시키는 것을 보였으며, 이러한 방법론이 스포츠 과학에 활용되길 기대한다.

참고문헌

[1] Xiao, Bin et al., "Simple Baselines for Human Pose Estimation and Tracking", ECCV 2018.

[2] Sun, Ke et al., "Deep High-Resolution Representation Learning for Human Pose Estimation", CVPR 2019.

[3] Zhang, Feng et al., "Distribution-Aware Coordinate Representation for Human Pose Estimation", CVPR 2020.

[4] AJ Piergiovanni and Michael S. Ryoo, "Fine-Grained Activity Recognition in Baseball Videos", CVPRW 2018.

[5] Zhang, Weiyu et al., "From Actemes to Action: A Strongly-Supervised Representation for Detailed Action Understanding", ICCV 2013.