일반논문 (Regular Paper)
방송공학회논문지 제28권 제6호, 2023년 11월 (JBE Vol.28, No.6, November 2023)
https://doi.org/10.5909/JBE.2023.28.6.698
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

외관 변화를 고려한 적응적 다수 휴먼 추적 방법

전 도 현[a)], 장 주 용[a)‡]

Adaptive Multi-person Tracking Method considering Appearance Changes

Do Hyun Jeon[a)] and Ju Yong Chang[a)‡]

요 약

다수 휴먼 추적 모델은 일반적으로 검출기의 높은 성능을 바탕으로 검출 기반의 추적 방식을 따른다. 이러한 방식의 다수 휴먼 추적 모델은 휴먼 객체에 대한 검출 영역 내의 외관 변화에 따라 추적 성능이 크게 저하되는 문제점을 가진다. 이를 해결하기 위해 외관 변화를 고려한 2단계의 추적 방법이 제안된 바 있다. 본 논문에서는 기존 방법을 더욱 개선하여 높은 신뢰도의 검출 결과에서 모션 블러(motion blur)를 낮추고 외관 변화에 강인한 자세 복원 방법을 활용하여 특징을 추출하는 방법을 제안한다. 그리고 이러한 특징을 기반으로 프레임 간 휴먼 객체 유사도를 측정하고 최적의 정합을 결정한다. PoseTrack21 데이터셋에 대한 실험에 따르면 제안하는 방법은 기존 방법에 비해 0.82 향상된 HOTA와 정성적으로 개선된 결과를 산출한다. 이는 제안하는 방법이 외관 변화가 발생하는 상황에서의 휴먼 추적에 효과적임을 보여준다.

Abstract

In general, multi-person tracking models follow a detection-based tracking approach that relies on the high performance of the detector. The problem with this type of multi-person tracking model is that tracking performance degrades significantly with changes in appearance within the detection area for human subjects. To address this, a two-step tracking method has been proposed that considers appearance changes. In this paper, we propose a new method that improves on the existing method, reduces motion blur with high confidence detection results, and utilizes a pose reconstruction method that is robust to appearance changes to extract features. Based on these features, the similarity between human subjects in subsequent frames is measured and the best match is determined. Experiments on the PoseTrack21 dataset show that the proposed method improves HOTA by 0.82 and produces qualitatively better results compared to the existing method. This shows that the proposed method is effective for multi-person tracking in the presence of appearance changes.

Keyword : Multi-person tracking, 3D human reconstruction, Deep learning

그림 1. 역동적인 자세(1행)와 모션 블러(2행)가 발생하는 경우에 대해 PHALP+B를 사용한 추적 결과에서의 ID 스위칭 발생
Fig. 1. ID switching in tracking results using PHALP+B for dynamic pose (row 1) and motion blur (row 2)

Ⅰ. 서 론

다수 휴먼 추적(multi-person tracking)은 컴퓨터 비전의 주요 연구 분야 중 하나로서 입력 단안 비디오로부터 각 휴먼 객체에 대한 바운딩 박스(bounding box)와 추적 ID(identification)를 추정하는 것을 그 목표로 가진다. 딥러닝 기술의 발전과 함께 최근 다수 휴먼 추적 방법의 성능이 크게 향상되었고, 그 결과 감시 시스템(surveillance system), 스포츠 분석, 자율 주행 등에서 활발하게 사용되고 있다. 그러나 영상으로부터 다수 휴먼에 대한 추적을 성공적으로 수행하는 것은 여전히 어려운 일이며, 특히 검출 영역 내 휴먼 객체의 외관 변화로 인한 추적 성능 저하는 해결해야 할 문제들 중 하나이다.

단안 비디오를 입력으로 사용하여 추적을 수행하는 다수 휴먼 추적 모델은 역동적인(dynamic) 자세, 가리워짐(occlusion), 모션 블러(motion blur) 등으로 인한 휴먼 객체의 외관 변화(appearance changes)가 발생할 때 추적 성능이 저하되는 결과를 보인다. 이는, 검출에 의한 추적(tracking by detection) 구조의 모델의 경우, 이러한 외관 변화가 발생할 때, 추출된 특징이 해당 휴먼 객체의 특징을 제대로 반영하지 못하기 때문이다. 이러한 문제를 해결하기 위해 PHALP+B[1]에서는 외관 변화를 고려한 2단계의 다수 휴먼 추적 방법이 제안되었다. PHALP+B는 높은 신뢰도의 검출 결과들의 경우 외관 변화의 정도가 크지 않다는 가정 아래 검출된 휴먼 객체에 대한 3차원 정보, 즉, 자세(pose), 위치(location), 외관을 추출하여 추적에 활용한다. 하지만 높은 신뢰도의 검출 결과에서도 큰 외관 변화로 인해 추적 성능이 저하되는 경우가 발생할 수 있다. 예를 들어 그림 1의 1행과 같이 점프 중 낙하하는 선수가 역동적인 자세를 취하거나 2행과 같이 육상 선수들에 대해 가려짐과 높은 속도로 인한 모션 블러가 발생될 때, ID 스위칭(ID switching)이 나타난다.

본 논문에서 우리는 이러한 문제를 해결하기 위해 외관 변화를 고려하여 추적 단계를 나누는 PHALP+B 방법을 확장하여 높은 신뢰도의 검출 결과에 대하여 강화된 특징 추출 방법을 추적에 활용한다. 구체적으로 첫 번째 추적 단계에서 검출 영역에 대한 디블러(deblur)와 트랜스포머(transformer) 기반의 3차원 휴먼 복원 방법을 활용하여 외관 변화에 강인하게 외관과 자세에 대한 휴먼 객체 특징을 추출

a) 광운대학교 전자통신공학과(Department of Electronics and Communications Engineering, Kwangwoon University)
‡ Corresponding Author : 장주용(Ju Yong Chang)
E-mail:: juyong.chang@gmail.com
Tel: +82-2-940-5136
ORCID: https://orcid.org/0000-0003-3710-7314
※ This work was partly supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00348, Development of A Cloud-based Video Surveillance System for Unmanned Store Environments using Integrated 2D/3D Video Analysis, 90%) and the Excellent researcher support project of Kwangwoon University in 2023 (10%).
· Manuscript September 1, 2023 ; Revised October 25, 2023; Accepted, October 25, 2023.

한다.

우리는 제안하는 추적 방법을 정량적, 정성적으로 평가한다. 그리고 평가 결과로부터 제안하는 추적 방법이 역동적인 자세, 가리워짐, 모션 블러와 같은 휴먼 객체의 외관 변화에 대해 강인한 성능을 나타냄을 보인다.

본 논문의 구성은 다음과 같다. II장에서는 제안하는 다수 휴먼 추적 모델에 관련된 기존 연구들이 소개된다. III장에서는 제안하는 방법에 대한 자세한 설명이 추적 과정에 따라 제시된다. IV장에서는 제안하는 방법의 정량적, 정성적 결과 및 분석이 기술되며, 마지막으로 V장에서는 결론이 주어진다.

II. 관련 연구

1. 3차원 표현 기반 다수 휴먼 추적 모델

T3DP[2]에서는 2차원 표현에 비해 3차원 표현은 구분되기 쉽고(distinguishable) 시점에 불변(invariant)한 특성을 가지고 있으므로 휴먼 추적에 더 적합하다는 주장이 제기되었다. T3DP는 검출에 의한 추적 방식을 따르며, 3차원 자세, 위치, 그리고 외관 속성을 특징으로 추출 및 예측하여 휴먼 객체의 추적에 활용한다. T3DP의 확장 모델인 PHALP[3]에서도 동일한 속성의 3차원 표현이 휴먼 추적에 활용된다. PHALP는 PoseTrack18 데이터셋[4]에서 프레임 간 휴먼 객체들의 3차원 속성별 거리가 주어질 때 검출된 휴먼 객체가 추적되고 있는 휴먼 객체에 포함될 확률에 기반한 비용 계산 함수를 추적을 위한 데이터 연관(data association)에 활용한다. 4DHumans[5]는 트랜스포머(trans-former) 기반의 3차원 휴먼 복원 모델인 HMR2.0을 제안하여 보다 정확한 자세 특징을 휴먼 추적에 활용할 수 있도록 하였다. 또한 이전 모델들[2,3]과 다르게 SMPL 자세 파라미터[5]를 자세 특징으로서 활용하여 일반적인 3차원 자세 복원 모델이 추적에 적용될 수 있는 구조를 제안하였다.

2. 외관 변화를 고려한 다수 휴먼 추적 모델

검출에 의한 추적 구조를 가지는 기존의 모델들[6,7,8]은 일반적으로 검출에 대한 신뢰도 임계치 이하의 결과들을 추적에서 제외한다. 하지만 이러한 검출 결과들은 외관 변화가 발생한 객체를 포함할 수 있는데, 단순히 이러한 객체들을 필터링하는 방식은 추적 성능에 무시할 수 없는 악영향을 끼칠 수 있다. ByteTrack[9]에서는 모든 검출 결과들을 추적에 활용하되 검출 신뢰도를 기준으로 2단계로 수행되는 추적 알고리즘이 제안되었다. ByteTrack은 검출에 의한 추적 구조를 가지는 추적 모델에 일반적으로 적용될 수 있고, MOT 데이터셋[10]에 대해 향상된 추적 성능을 보였다. 그리고, PHALP+B는 기존의 3차원 표현 기반 추적 모델에 휴먼 객체의 외관 변화를 고려한 알고리즘을 적용하여 PoseTrack 데이터셋[4,11]에 대해 추적 성능이 향상됨을 보였다. 하지만 PHALP+B는 높은 검출 신뢰도를 가지는 휴먼 객체가 역동적인 자세를 가지거나 모션 블러가 발생할 때, 여전히 저하된 성능의 추적 결과를 산출할 수 있다. 본 연구에서는 PHALP+B를 baseline 모델로 하여 높은 신뢰도 범위의 검출 결과에 대한 자세 및 외관 특징 추출의 정확도를 높여 추적 성능을 향상시키는 방법을 제안한다.

III. 제안하는 방법

본 연구에서 제안하는 외관 변화를 고려한 적응적 다수 휴먼 추적 방법의 추적 과정은 그림 2와 같다. 첫 번째, 단안 비디오의 프레임이 검출기에 입력되면 휴먼 객체에 대한 검출 결과가 출력된다. 이 때 외관 변화 정도에 따라 검출 결과가 나누어진다. 두 번째, 검출 영역에서 각 추적 단계에 따른 최적의 특징이 추출된다. 세 번째, 추출된 특징을 바탕으로 검출된 휴먼 객체와 예측된 휴먼 객체 사이의 유사도가 계산된다. 네 번째, 유사도를 기반으로 데이터 연관이 수행되어 인접한 프레임에서 휴먼 객체 간 정합이 결정된다.

1. 검출 및 특징 추출

먼저 입력 비디오 V의 t번째 프레임 f_t가 입력되면 검출 결과 $D_t = \{d_t^i\}_1^{i=m}$가 출력된다. 여기서 d_t^i는 i번째 휴먼 객

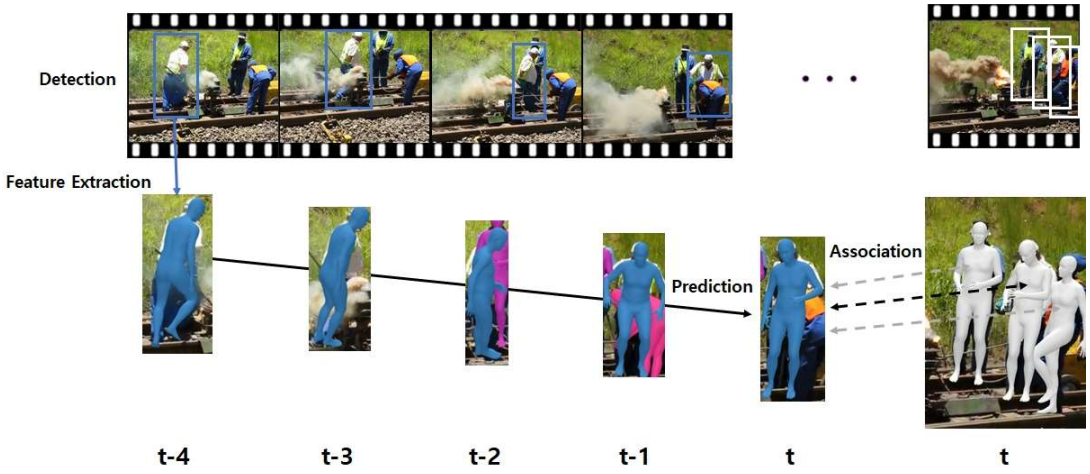

그림 2. 제안하는 모델의 개요
Fig. 2. The overview of the proposed method

체에 대한 바운딩 박스 b_t^i, 검출 신뢰도 점수 c_t^i, 이진 마스크 m_t^i의 집합이다. 검출 결과 d_t^i는 검출 신뢰도 c_t^i가 신뢰도 임계치 c_{high} 이상일 경우 높은 신뢰도 검출 결과의 집합 DH로 분류된다. 그렇지 않고 c_t^i가 c_{high} 미만이고 낮은 검출 신뢰도 임계치 c_{low}이상이면 d_t^i는 낮은 신뢰도 검출 결과의 집합 DL로 분류된다. 이러한 검출 및 특징 추출 과정은 그림 3에 나타나 있다.

검출 결과 d_t^i가 DH로 분류되는 경우 특징 추출기를 통해 3차원 표현 R_t^i가 추출된다. 3차원 표현 $R_t^i = \{p_t^i, l_t^i, a_t^i\}$는 자세, 위치, 외관에 대한 정보를 포함하는 특징 벡터의 집합으로 $p_t^i \in \mathbb{R}^{24\times3x3}$, $l_t^i \in \mathbb{R}^3$, $a_t^i \in \mathbb{R}^{512}$를 만족한다. 먼저, i번째 사람의 검출 영역 영상 f_t^i에 대해 디블러 모델[12]이 다음과 같이 적용된다: $deblur(f_t^i)$. 다음으로 디블러된 영상이 트랜스포머 기반의 휴먼 메쉬 복원 모듈인 HMR2.0[5]에 입력되어 SMPL 파라미터 $\Theta = [\theta, \beta, \pi]$가 출력된다. 여기서 $\theta \in \mathbb{R}^{24\times3x3}$, $\beta \in \mathbb{R}^{10}$, $\pi = (R, t)$는 각각 자세, 형태, 카메라 파라미터이며, 카메라 파라미터는 전역 회전(global orientation) $R \in \mathbb{R}^{3\times3}$과 전역 이동(global translation) $t \in \mathbb{R}^3$으로 구성된다. 복원된 SMPL 파라미터 중 θ는 추적을 위한 3차원 표현 중 자세 특징 p_t^i로 활용되고, 복원된 SMPL 메쉬의 골반(pelvis)의 2차원 픽셀 좌표인 (x_t^i, y_t^i)와 깊이(depth) 정보를 나타내는 nearness[13] n_t^i는 함께 위치 특징 $l_t^i = (x_t^i, y_t^i, n_t^i)$로 사용된다. 그리고 a_t^i는 HMAR[2]의 외관 특징 추출 모듈을 통해 추출된다.

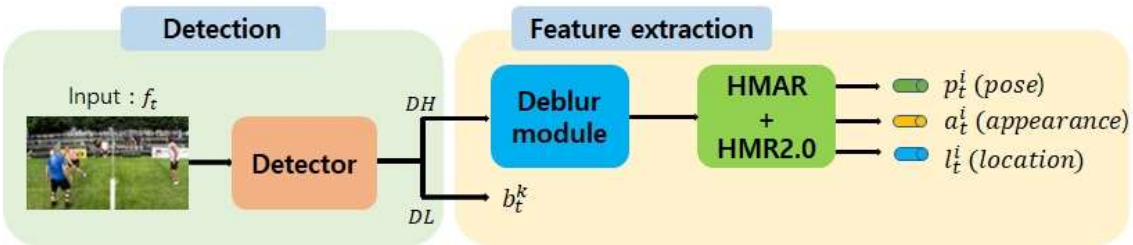

그림 3. 검출기 및 특징 추출기의 구조
Fig. 3. The pipeline of the detector and feature extractor

다음으로 검출 결과가 DL으로 분류되는 경우에는 별도의 추가적인 특징 추출 없이 바운딩 박스 b_t^i가 데이터 연관을 위한 특징으로 추출된다.

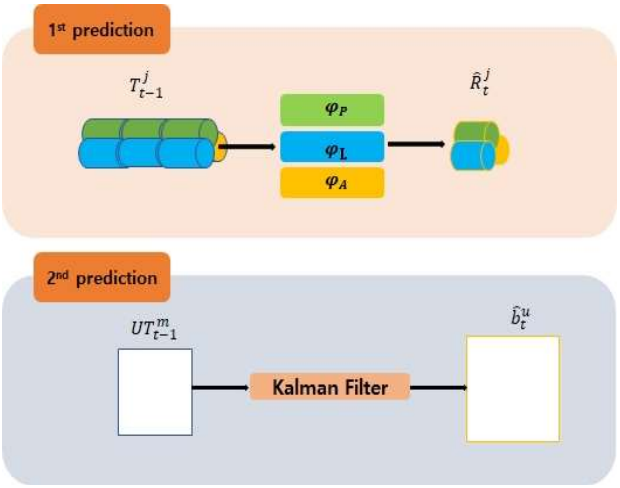

그림 4. 예측 모듈의 구조
Fig. 4. The pipeline of the predictor

2. 예측

검출 및 특징 추출과 동일하게 예측 과정도 그림 4와 같이 두 단계로 수행된다. 첫 번째 예측 단계에서는 DH에 속하는 검출 결과들과 3차원 표현에 기반한 유사도 계산을 하기 위해 현재까지의 추적 결과인 트랙릿(tracklet)의 3차원 표현에 대한 예측을 수행한다. 추적 ID가 j인 휴먼 객체에 대한 트랙릿 T_{t-1}^j의 3차원 표현에 대한 예측을 수행한다고 가정하자. 이때 트랙릿 T_{t-1}^j가 포함하고 있는 3차원 표현 시퀀스를 $R^j=\{P^j, L^j, A^j\}$라고 하자. 여기서 P^j, L^j, A^j은 이전 q프레임 동안의 자세, 위치, 외관 특징 벡터의 시퀀스로서 $P^j=\{p_{t-1}^j, p_{t-2,\dots,}^j p_{t-q}^j\}$, $L^j=\{l_{t-1}^j, l_{t-2,\dots,}^j l_{t-q}^j\}$, $A^j=\{a_{t-1}^j, a_{t-2,\dots,}^j a_{t-q}^j\}$로 표현될 수 있다.

먼저 P^j가 트랜스포머 기반의 자세 예측 모델[3]에 입력되어 자세 특징 벡터 $\hat{p}_t^j=\varphi_p(P^j)$가 예측된다. 위치 특징 벡터의 경우 간단한 선형 회귀 함수 $\varphi_L(L^j)$에 의해 각 좌표가 독립적으로 예측된다. 외관 특징 벡터의 예측을 위해서는 A^j가 사용되지 않고 해당 휴먼 객체의 UV 영상 $T_t^j\in\mathbb{R}^{3\times256\times256}$와 가시성 지도 $v_t^j\in\mathbb{R}^{1\times256\times256}$가 결합된 $A_t^j=[T_t^j, v_t^j]\in\mathbb{R}^{4\times256\times256}$가 사용된다. 외관 특징 벡터 $\hat{a}_t^i$의 예측을 위해 먼저 $\hat{A}_t^j$가 다음과 같이 A_t^j와 $\hat{A}_{t-1}^j$의 가중합(weighted sum) 연산 φ_A를 통해 계산된다:

$$\hat{A}_t^j=\varphi_A(\hat{A}_{t-1}^j, A_t^j)=(1-\beta)\hat{A}_{t-1}^j+\beta A_t^j. \quad (1)$$

그 후 계산된 $\hat{A}_t^j$는 HMAR의 인코더에 입력되어 외관 특징 벡터 $\hat{a}_t^j$가 출력된다.

두 번째 예측 단계에서는 DL에 속하는 검출 결과들과 intersection over union(IoU)를 계산하기 위해 트랙릿의 바운딩 박스를 예측한다. 이를 위해 첫 번째 데이터 연관 과정을 통해 정합되지 않은 트랙릿 T_{t-1}^j의 바운딩 박스 b_{t-1}^j가 칼만 필터(Kalman filter)에 입력되어 $\hat{b}_t^j$가 예측된다.

3. 비용 계산

첫 번째 단계에서의 비용 함수는 확률적으로 정의된다. 검출 결과로부터 계산된 특징 벡터들 p_t^i, $l_t^i=[x_t^i, y_t^i, n_t^i]^T$, a_t^i과 트랙릿 T_{t-1}^j의 예측된 특징 벡터들 $\hat{p}_t^j$, $\hat{l}_t^j$, $\hat{a}_t^j$사이의 유클리드 거리 $\Delta_p, \Delta_{xy}, \Delta_n, \Delta_a$가 주어졌을 때, 검출 결과 d_t^i가 트랙릿 T_{t-1}^j에 속할 조건부 확률은 다음과 같다:

$$P(d_t^i\in T_{t-1}^j|\Delta_p,\Delta_{xy},\Delta_n,\Delta_a)\propto P_P P_{XY} P_N P_A, \quad (2)$$

여기서 P_P, P_{XY}, P_N, P_A는 각 특징 벡터에 대한 조건부 확률로써, 이는 PoseTrack18 학습 데이터셋에 대한 검출 결과와 트랙릿 예측 사이의 정합에 대한 true positive 분포를 통해 모델링 될 수 있다[3]. 이제 검출 결과와 트랙릿 사이의 비용 함수 φ_{C1}는 다음과 같이 정의된다:

$$\begin{aligned}\varphi_{C1}(d_t^i, T_{t-1}^j) &= -\log\left(P(d_t^i\in T_{t-1}^j|\Delta_p,\Delta_{xy},\Delta_n,\Delta_a)\right) \\ &= -\log(P_P)-\log(P_{XY})-\log(P_N)-\log(P_A),\end{aligned} \quad (3)$$

여기서 각 특징 벡터 별 조건부 확률 P_P, P_{XY}, P_N, P_A은 스케일 파라미터(scale parameter)를 포함한다. 이러한 스케일 파라미터들은 PoseTrack18 검증 데이터셋에 대한 프레임 단위의 데이터 연관 오차를 넬더-미드(Nelder-Mead) 알고리즘[14]으로 최소화하여 결정된다.

두 번째 비용 계산에서는 DL에 속한 검출 결과들의 바운딩 박스들과 첫 번째 데이터 연관 단계에서 정합되지 않은 트랙릿에 대해 예측된 바운딩 박스들 사이의 IoU가 계산된다. 이 때의 비용은 각 IoU 값에 -1을 곱한 값으로 정의된다.

4. 데이터 연관

데이터 연관 단계에서는 헝가리안 알고리즘(Hungarian algorithm)[15]이 사용된다. 비용 계산 단계에서 획득된 n개의 검출 결과들과 m개의 트랙릿들 사이의 정합 비용으로부터 비용 행렬 $C \in \mathbb{R}^{n \times m}$이 구성된다. 헝가리안 알고리즘은 비용 행렬로부터 최소 가중치(minimum weight) 정합을 효율적으로 계산하여 트랙릿이 최적의 검출 결과를 포함할 수 있도록 해 준다. 첫 번째 데이터 연관 과정을 통해, 정합된 트랙릿 $mT1$, 정합되지 않은 트랙릿 $uT1$, 정합되지 않은 검출 $uD1$이 출력된다. 두 번째 데이터 연관에서도 $mT2$와 $uT2$가 출력되고, 정합되지 않은 검출인 $uD2$는 삭제된다. 최종적으로 $mT1$와 $mT2$에 의해 이전 트랙릿이 갱신되고, $uD1$이 새로운 트랙릿으로 할당된다. $uT2$에는 최대 수명(max age)인 t_{max}가 부여되고, t_{max} 개의 프레임이 지난 후에도 정합되지 않는다면 삭제된다.

Ⅳ. 실험 결과 및 분석

1. 데이터셋, 평가 방법, 구현 세부사항

본 연구에서는 제안하는 방법을 평가하기 위해 다수 휴먼 객체에 대해 밀집한 바운딩 박스와 추적 ID가 어노테이션(annotation) 되어 있는 PoseTrack 데이터셋 중 PoseTrack21 데이터셋[11]을 사용한다.

PoseTrack 데이터셋은 MPII 휴먼 자세 데이터셋[16]을 확장하여 구성된다. 593개, 170개, 375개의 학습, 검증, 테스트 시퀀스로 나뉘어지며 총 46,933개의 프레임으로 이루어진다. PoseTrack21은 다수 휴먼에 대한 자세 추정 및 추적 성능 평가를 위해 제안된 PoseTrack18에 기반한 데이터셋이다. 두 데이터셋은 동일한 시퀀스로 구성되어 있지만 PoseTrack18 데이터셋과는 달리 PoseTrack21 데이터셋에서는 작거나, 밀집되어 있거나, 잘려진(truncated) 사람들에 대한 어노테이션이 제공된다. 또한 평가 시 제외되는 무시 영역(ignore region)이 PoseTrack18 데이터셋 보다 적게 설정되었다. 이러한 특성을 바탕으로 외관 변화가 큰 휴먼 객체를 포함하는 동영상에서의 휴먼 추적을 수행하는 제안 모델에 대한 평가에 PoseTrack21 데이터셋을 활용한다.

우리는 제안하는 방법을 평가하기 위해 HOTA[17]를 측정하여 보고한다. HOTA는 시퀀스 별로 추적 결과에 대한 바운딩 박스와 추적 ID를 기준으로 검출 정확도인 DetA (detection accuracy)와 데이터 연관 정확도인 AssA(association accuracy) 두 평가 지표를 균형 있게 반영한 평가 척도이다.

제안하는 방법에서 검출기로는 Mask-RCNN[18]이 사용되고, c_{high}는 PHALP+B와 같이 0.8로 설정된다. c_{low}는 PoseTrack21 학습 데이터셋에 대한 그리드 서치(grid search)를 통해 0.5로 설정하였다.

2. 정량적 평가 결과

우리는 제안하는 외관 변화를 고려한 적응적인 추적 방법이 성능 개선에 정량적으로 도움을 준다는 것을 보인다. 이를 위해 우리는 baseline 모델인 PHALP+B를 단계적으로 확장하여 baseline 모델과의 정량적 비교를 수행한다. 평가는 PoseTrack21 데이터셋의 검증 데이터셋에 대하여 수행되었다.

표 1은 PoseTrack21 데이터셋에 대해 baseline 방법인 PHALP+B와 제안하는 방법의 정량적 성능을 보여준다. 여기서 PHALP'는 기존 PHALP에서 자세 복원 모듈만 HMR2.0으로 바뀐 모델이다. 먼저, PHALP'+B는 PHALP' 모델과 비교하여 0.22 높은 HOTA를 나타내는데, 이는 PHALP' 모델에 대해서도 외관 변화를 고려한 2단계 추적

표 1. Baseline 모델과 제안하는 방법(ours)의 PoseTrack21 검증 데이터셋에 대한 성능 비교
Table 1. Performance comparison of baseline models and proposed method (ours) on PoseTrack21 validation dataset

Model	HOTA	DetA	AssA
PHALP+B	58.46	48.16	72.55
PHALP'	58.94	48.56	73.13
PHALP'+B	59.16	49.05	73.08
PHALP'+B_deblur (ours)	59.28	49.04	73.22

알고리즘의 적용이 추적 성능에 도움을 준다는 것을 보여준다. 그리고 PHALP'+B는 PHALP+B보다 0.70 높은 HOTA를 달성한다. 이를 통해 자세 특징에 대한 추출 정확도를 높이는 것이 추적 성능에 도움을 준다는 것을 알 수 있다[5]. 또한 PHALP'+B의 첫 번째 추적 단계에서 DH에 디블러를 적용했을 때, HOTA가 추가적으로 0.12 향상되었다. 이를 통해 휴먼 객체에 대해서 모션 블러를 감소시키는 것이 외관 특징 추출의 정확도를 높이고, 결과적으로 추적 성능에 도움이 됨을 알 수 있다. 이에 대한 구체적인 분석은 4.3장의 정성적 평가 결과에서 확인할 수 있다.

3. 정성적 평가 결과

그림 5, 6, 7은 PoseTrack21 데이터셋에 대해 제안하는 모델인 PHALP'+B_deblur의 추적 성능을 나타내는 정성적인 결과이다. 그림 5에서는 검출 신뢰도에 따른 검출 결과가 제시되며, PHALP'과 PHALP'+B의 추적 결과를 비교하였다. PHALP'+B 은 외관 변화가 큰 무대 뒤편의 흐릿한 사람 및 무대 위의 춤추며 가려진 배우에 대해서 PHALP'와 달리 추적을 성공적으로 수행한다. 그림 6에서는 baseline모델인 PHALP+B와 PHALP'+B의 추적 결과를 비교하였다. PHALP'+B는 낙하하며 역동적인 자세를 보이는 선수에 대하여 정확한 자세 복원 성능을 보이고, 이에 기반하여 PHALP+B와 달리 ID 스위칭 없이 추적을 성공적으로 수행한다. 그림 7에서는 PHALP'+B와 PHALP'+B_deblur의 추적 결과를 비교하였다. PHALP'+B_deblur은 서로에 대한 가리워짐과 모션 블러가 발생한 두 선수에 대하여 ID 스위칭 없이 추적을 성공적으로 수행한다.

우리는 디블러가 적용된 경우의 텍스쳐 영상(textue image)[3]을 확인하여 디블러가 어떤 식으로 추적 성능에 도움이 되었는지 확인해 보았다. 그림 8은 그림 7의 PHALP'+B

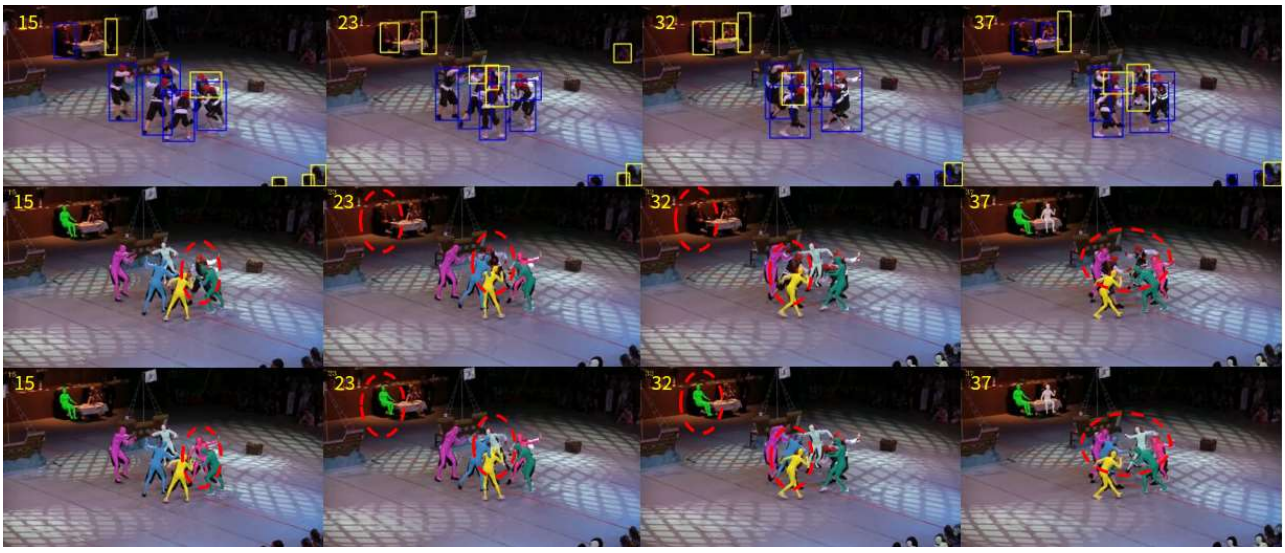

그림 5. PoseTrack21 데이터셋에 대한 검출 결과(1행)에서 높은 신뢰도의 검출 결과는 파란색, 낮은 신뢰도의 검출 결과는 노란색 바운딩 박스로 표시됨. PHALP'(2행)와 PHALP'+B(3행)의 정성적 비교
Fig. 5. Detection results for the PoseTrack21 dataset (row 1), with high confidence detections colored in blue and low confidence detections colored in yellow bounding boxes. Qualitative comparison of PHALP' (row 2) and PHALP'+B (row 3)

그림 6. PoseTrack21 데이터셋에 대한 PHALP+B(1행)와 PHALP'+B(2행)의 정성적 비교
Fig. 6. Qualitative comparison of PHALP+B(row1) and PHALP'+B(row2) on PoseTrack21 dataset

그림 7. PoseTrack21 데이터셋에 대한 PHALP'+B(1행)와 PHALP'+B_deblur(2행)의 정성적 비교
Fig. 7. Qualitative comparison of PHALP'+B(row1) and PHALP'+B_deblur(row2) on PoseTrack21 dataset

그림 8. 입력 영상(a)에 디블러를 적용한 후 얻어진 텍스쳐 영상(c, f)과 디블러를 적용하지 않고 얻어진 텍스쳐 영상(b, e)의 비교
Fig. 8. Comparison of texture images with deblur(c, f) and without deblur(b, e)

추적 결과에서 모션 블러로 인해 ID 스위칭이 발생하는 프레임에 대한 각 선수의 텍스쳐 영상이다. 그림 8의 b, e는 블러를 적용하지 않았을 경우의 텍스쳐 영상을, c, f는 디블러를 적용한 경우의 텍스쳐 영상을 나타낸다. 그림 d는 텍스쳐 영상의 각 픽셀에 대응되는 신체 부위의 예를 보여준다. 흰색 유니폼 선수에 대한 결과에서, b에 비해 c에서 머리나 유니폼 영역에서의 텍스쳐가 좀 더 잘 추정된 것을 볼 수 있다. 또한 자주색 유니폼 선수에 대한 결과에서도, e에 비해 f에서 팔 영역의 텍스쳐가 잘 추정되고, 유니폼의 무늬가 잘 표현된 것을 볼 수 있다. 디

마지막으로 제안하는 방법이 추적에 실패하는 경우에 대해서 분석하였다. 제안하는 방법은 검출에 대한 신뢰도 점수를 활용하여 휴먼 객체의 외관 변화 정도를 결정한다. 이러한 검출 결과의 예가 그림 9의 1행에 제시되어 있다. 16번

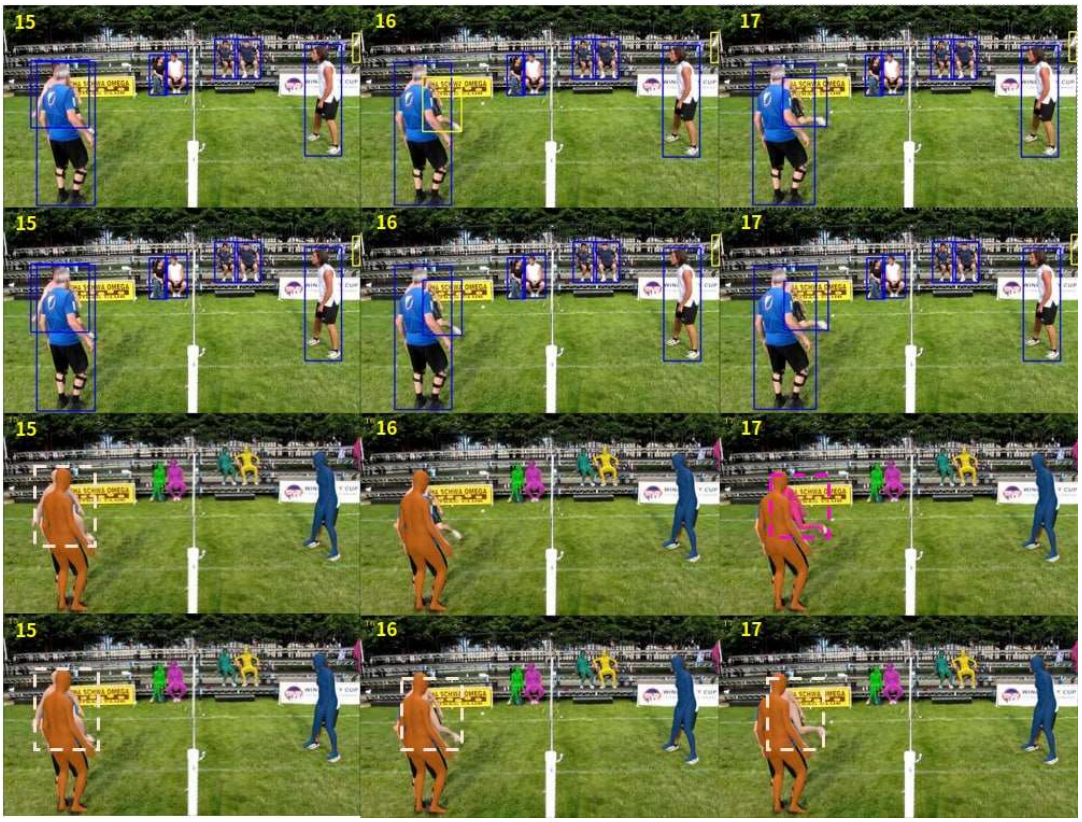

그림 9. 검출 신뢰도를 기반으로 *DL*(노란색), *DH*(파란색)로 분류된 검출 결과(1행), 1행의 검출 결과에 따른 PHALP'+B_deblur의 추적 결과(3행), 3행의 추적 결과 ID 스위칭이 발생한 16번 프레임의 사람을 임의로 *DH*로 할당한 검출 결과(2행), 2행의 검출 결과에 따른 PHALP'+B_deblur의 추적 결과(4행)

Fig. 9. Detection result with *DL*(yellow) and *DH*(blue) based on detection confidence score(row1), tracking result of PHALP'+B_deblur based on detection result in row1(row3), detection result of arbitrary assigning the human subject in frame 16 where ID switching occurred in row3 to *DH*(row2), tracking result of PHALP'+B_deblur based on detection result in row2(row4)

째 프레임에서 가리워진 사람에 대해 검출 결과가 DL로 분류되고, 결국 PHALP'+B_deblur의 추적 결과(3행)에서 ID 스위칭이 발생한다. 이는 16번째 프레임에서 가리워진 부분이 제외된 바운딩 박스 영역으로 인해 낮은 IoU 유사도를 갖기 때문이다. 한편 2행의 검출 결과와 같이 가리워진 사람에 대해 임의로 DH로 분류 후 추적을 진행할 경우, 4행을 통해 16번 프레임의 가리워진 사람에 대한 검출에 대해서도 SMPL 복원이 올바르게 수행되고 결과적으로 3차원 표현을 통해 ID 스위칭 없이 추적되는 것을 확인할 수 있다. 이러한 사례는 외관 변화 정도를 판단하기 위해 검출 신뢰도 점수만 활용하는 현재의 방법을 보다 개선할 필요가 있음을 보여준다.

Ⅴ. 결 론

본 연구에서 우리는 휴먼 객체에 대한 외관 변화가 발생할 때 다수 휴먼 추적의 성능이 저하되는 문제를 개선하기 위해 외관 변화에 적응적인 다수 휴먼 추적 방법을 제안하였다. 제안하는 방법은 검출 신뢰도를 기준으로 추적을 2단계로 나누고, 높은 신뢰도의 검출에 대해서 디블러 모델과 개선된 3차원 복원 모델을 적용하여 외관 변화에 강인한 특징을 추출한다. PoseTrack21 데이터셋을 사용한 실험을 통해 우리는 제안 모델이 기존 모델과 비교하여 역동적인 자세, 가리워짐, 모션 블러와 같은 외관 변화를 가지는 휴먼 객체에 대해 향상된 추적 성능을 달성함을 확인하였다.

참 고 문 헌(References)

[1] D. Jeon and J. Y. Chang, “Two-step multi-person tracking method considering appearance changes,” *The Korean Institute of Broadcast and Media Engineers Summer Conference*, Jun. 2023.

[2] J. Rajasegaran, G. Pavlakos, A. Kanazawa, and J. Malik, “Tracking people with 3D representations,” *Advances in Neural Information Processing Systems,* vol. 34, pp. 23703-23713, 2021.
url: https://proceedings.neurips.cc/paper_files/paper/2021/file/c74c4bf0dad9cbae3d80faa054b7d8ca-Paper.pdf

[3] J. Rajasegaran, G. Pavlakos, A. Kanazawa, and J. Malik, “Tracking people with 3D appearance, location & pose,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, pp. 2740-2749, 2022.
doi: https://doi.org/10.1109/CVPR52688.2022.00276

[4] M. Andriluka, U. Iqbal, E. Insafutdinov, L. Pishchulin, A. Milan, J. Gall, and B. Schiele, “PoseTrack: A benchmark for human pose estimation and tracking,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 5167-5176, 2018.
doi: https://doi.org/10.1109/CVPR.2018.00542

[5] S. Goel, G. Pavlakos, J. Rajasegaran, A. Kanazawa, and J. Malik, “Humans in 4D: Reconstructing and tracking humans with transformers,” *IEEE/CVF International Conference on Computer Vision*, Paris, France, pp. 14783-14794, 2023.
doi: https://doi.org/10.48550/arXiv.2305.20091

[6] Z. Lu, V. Rathod, R. Votel, and J. Huang, “RetinaTrack: Online single stage joint detection and tracking,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14668-14678, 2020.
doi: https://doi.org/10.1109/CVPR42600.2020.01468

[7] P. Tokmakov, J. Li, W. Burgard, and A. Gaidon, “Learning to track with object permanence,” *IEEE/CVF International Conference on Computer Vision*, pp. 10860-10869, 2021.
doi: https://doi.org/10.1109/ICCV48922.2021.01068

[8] Z. Wang, L. Zheng, Y. Liu, Y. Li, and S. Wang, “Toward real-time multi-object tracking,” *European Conference on Computer Vision*, pp. 107-122, 2020.
doi: https://doi.org/10.1007/978-3-030-58621-8_7

[9] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, “ByteTrack: Multi-object tracking by associating every detection box,” *European Conference on Computer Vision*, pp. 1-21, 2022.
doi: https://doi.org/10.1007/978-3-031-20047-2_1

[10] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler, “MOT16: A benchmark for multi-object tracking,” *arXiv:1603.00831*, 2016.
doi: https://doi.org/10.48550/arXiv.1603.00831

[11] A. Döring, D. Chen, S. Zhang, B. Schiele, and J. Gall, “PoseTrack21: A dataset for person search, multi-object tracking and multi-person pose tracking,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, pp. 20963-20972, 2022.
doi: https://doi.org/10.1109/CVPR52688.2022.02029

[12] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, “DeblurGAN-v2: Deblurring(orders-of-magnitude) faster and better,” *IEEE/CVF International Conference on Computer Vision*, Seoul, South Korea, pp. 8878-8887, 2019.
doi: https://doi.org/10.1109/ICCV.2019.00897

[13] J. J. Koenderink, “Optic flow,” *Vision research*, vol. 26, no. 1, pp. 161-179, 1986.
doi: https://doi.org/10.1016/0042-6989(86)90078-7

[14] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *The computer journal*, vol. 7, no. 4, pp. 308-313, 1965.
doi: https://doi.org/10.1093comjnl/7.4.308

[15] H. W. Kuhn, “The hungarian method for the assignment problem,” *Naval Research Logistics Quarterly*, vol. 2, pp. 83-97, 1955.
doi: https://doi.org/10.1002/nav.3800020109

[16] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, “2D humans pose estimation: New benchmark and state of the art analysis,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Columbus, USA, pp. 3686-3693, 2014.

[17] J. Luiten, A. Osep, P. Dendorfer, P. Torr, A. Geiger, L. Leal-Taixé, and B. Leibe, “HOTA: A higher order metric for evaluation multi-object tracking,” *International journal of computer vision*, vol. 129, pp. 548-578, 2021.
doi: http://doi.org/10.1007/s11263-020-01375-2

[18] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” *IEEE/CVF International Conference on Computer Vision*, Venice, Italy, pp. 2961-2969, 2017.
doi: https://doi.org/10.1109/ICCV.2017.322

저 자 소 개

전 도 현

- 2022년 2월 : 광운대학교 전자통신공학과 학사
- 2022년 3월 ~ 현재 : 광운대학교 전자통신공학과 석사과정
- ORCID : https://orcid.org/0009-0004-7285-2576
- 주관심분야 : 컴퓨터 비전 및 머신러닝

장 주 용

- 2001년 2월 : 서울대학교 전기공학부 학사
- 2008년 2월 : 서울대학교 전기컴퓨터공학부 박사
- 2008년 2월 ~ 2009년 1월 : Mitsubishi Electric Research Laboratories (MERL) Postdoctoral Researcher
- 2009년 4월 ~ 2011년 1월 : 삼성전자 DMC 연구소 책임연구원
- 2011년 4월 ~ 2012년 2월 : 서울대학교 BK 조교수
- 2012년 3월 ~ 2017년 2월 : 한국전자통신연구원 선임연구원
- 2017년 3월 ~ 현재 : 광운대학교 전자통신공학과 교수
- ORCID : https://orcid.org/0000-0003-3710-7314
- 주관심분야 : 컴퓨터비전 및 머신러닝