

# 유사도 비교 모듈을 이용한 Tracking By Detection 모델 설계

양현성<sup>1</sup>, 정세훈<sup>2</sup>, 심춘보<sup>3</sup>

<sup>1</sup>순천대학교 IT-Bio융합시스템전공 석사과정

<sup>2</sup>순천대학교 컴퓨터공학과 교수

<sup>3</sup>순천대학교 인공지능공학부 교수

naiu8165@naver.com, cbsim@scnu.ac.kr

## Design of Tracking By Detection Model Using Similarity Comparison Module

Hyun-Sung Yang<sup>1</sup>, Se-Hoon Jung<sup>2</sup>, Chun-Bo Sim<sup>3</sup>

<sup>1</sup>Interdisciplinary Program in IT-Bio Convergence System, Suncheon National University

<sup>2</sup>Dept. of Computer Engineering, Suncheon National University

<sup>3</sup>Dept. of Artificial Intelligence Engineering, Suncheon National University

### 요 약

현대 컴퓨터 비전 분야에서는 객체 추적이 중요한 연구 주제 중 하나이다. 기존 Tracking By Detection 방식은 실시간 추적 속도와 Tracklet을 유지할 수 있는 정보 전달의 한계를 가지고 있다. 본 연구에서는 유사도 비교 모듈을 기반으로 Tracking By Detection 모델을 설계하고자 한다. 탐지 모델은 Anchor를 사용하지 않는 CenterNet을 사용하고 탐지된 값에 유사도 비교 알고리즘을 적용하여 객체 탐지와 객체 추적을 동시에 수행하는 모델을 제안한다. 제안하는 방법은 Occlusion으로 인한 객체 정보 손실을 완화하고, 새로운 객체 및 장애물에 대해 강건할 것으로 사료된다.

### 1. 서론

최근 MOT(Multiple Object Tracking)에서 주목 받는 패러다임은 TBD(Tracking By Detection)다 [1,2]. TBD 방법은 객체를 탐지하고, 탐지된 객체를 시간에 따라 추적하는 방식이다. 객체 탐지와 객체 추적은 각각 다른 목적을 갖지만, 객체를 찾고 활용하는 것에 있어 상호 연관되어 있다. 객체 탐지는 단일 이미지나 비디오 시퀀스의 개별 프레임에 작동하여 객체를 위치시키고, 각 프레임을 독립적인 객체로 분류하는 것이 목적이다. 반면, 객체 추적은 시간에 걸쳐 객체의 움직임을 모니터링하고, 객체의 상태를 여러 프레임에 걸쳐 유지한다.

객체 추적 과정에서 다양한 문제가 발생하지만, 주요 문제로 언급되는 것은 Occlusion이다. Occlusion이란 추적 대상이 다른 객체나 장애물에 의해 일시적이나 영구적으로 가려지는 현상이다. Occlusion이 발생하면 객체 일부 혹은 전체가 카메라 시야에서 사라져 객체의 신원을 판단하기 어렵다. 이는 추적 모델이 잘못된 방향으로 학습될 가능성이 존재하고, 객체가 다시 나타났을 때 이전 상태와의 연결에 어려움을 준다.

Occlusion 문제 해결에 관한 연구는 다양한 접근 방식

을 사용하여 진행되고 있다[3,4]. 그중 Re-ID(Re-Identification)를 이용한 연구는 객체가 시야에서 사라졌다가 재등장했을 때 이전 추적 대상과 같은지 판별한다. 이는 특히 Tracklet에 객체가 일시적으로 사라졌다가 다시 나타날 때 발생하는 Occlusion 처리에 유용하다.

Re-ID 알고리즘만으로는 완벽하게 Occlusion을 해결할 수 없다. Re-ID는 빠르게 변하는 외관 변화에 민감하게 반응할 수 있고, 여러 객체가 같은 프레임 내에서 밀접하게 위치하거나 중첩될 경우 복잡한 상황을 처리하지 못할 수 있다. 또한, 같은 클래스에 속하는 다양한 객체들 사이에서 Re-ID의 성능은 완벽할 수 없다.

본 연구에서는 이러한 문제를 해결하고자 TBD 패러다임을 따르고, Occlusion에 강건한 추적을 위해 각 모델 사이에 유사도 비교 모듈을 추가한 모델을 제안한다. 탐지 모델은 기존 CenterNet을 사용하고, 유사도 비교 모듈의 입력은 CenterNet 출력에 객체 중앙 Peak를 추가한다. 또한, 유사도 비교 모듈 출력 값과 Tracklet의 Value 값을 내적하여 추적의 강건성을 보장하고자 한다.

### 2. 관련 연구

[6] 연구에서는 Anchor를 사용하지 않고

Heatmap을 사용하여 객체의 중심점을 예측하고, 이를 기반으로 객체를 탐지했다. 이 연구는 Keypoint Detection, Bounding Box Regression, Object Classification을 단일 네트워크로 통합하여 계산 효율적인 탐지를 진행했다.

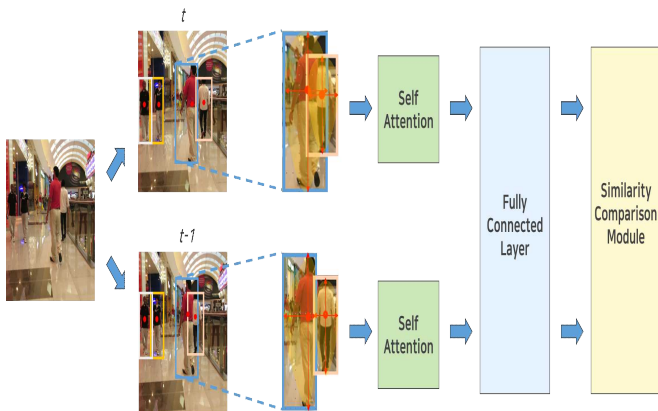
[7] 연구에서는 객체 검출과 Re-Identification을 유기적으로 결합한 방식을 제시했다. 이 방식은 Similarity 측정을 통해 서로 다른 시점에서 출현한 동일 객체의 신원을 판별하는 Re-ID 헤드와 객체 검출 헤드를 함께 학습했다. 이러한 통합 방식으로 복잡한 처리 없이 높은 정확도의 객체 추적을 가능하게 만들어, Similarity 계산을 더 효율적으로 수행했다.

[8] 연구에서는 Transformer 기반의 아키텍처를 이용하여 Attention 기반의 객체 추적 방법을 제안했다. 이전 프레임의 객체 특징을 현재 프레임의 Query로 사용하고, 사전 학습된 객체 Query 세트로 새로운 객체를 검출했다. 이 방식은 Single-shot으로 객체 탐지와 추적을 함께 수행하여, 복잡성을 완화했다.

**3. 제안하는 방법**

본 연구는 CenterNet과 Transformer 기반 추적 모델을 융합하여 실시간 객체 추적 모델을 설계한다. CenterNet은 Anchor-free 탐지 방식을 통해 객체의 중심점과 특성을 파악하고, Transformer는 추적에 필요한 Temporal Coherence를 고려한 모델링을 수행한다. 두 모델을 결합함으로써 실시간 처리 능력과 추적 정확도를 동시에 높이는 것이 본 연구의 목표다. 특히, 유사도 비교 모듈을 통해 배경과 장애물로 인한 추적 오류를 최소화하고,  $t$ 와  $t-1$  프레임 간의 객체 추적을 더욱 견고하게 만들어준다.

**3.1 다중 객체 유사도 비교 모듈**



(그림 1) Center based Similarity Comparison.

객체 추적에서 유사도 비교는  $t$  프레임과  $t-1$  프레임 객체를 연결하거나, 객체가 일시적으로 사라진 후 다시 나타날 때 기존 ID를 유지하기 위해 활용된다. 그림 1과 같이 다중 객체일 경우 한 객체를 나타내는 Anchor는 배경을 포함하고 있다. 배경 및 장애물 영역을 포함한 상태로 유사도 비교를 진행할 경우 ID 스위칭 오류를 범하기 쉽다. 따라서, CenterNet 출력에서 중앙 Peak와 수직수평 방향의 Local Feature Map을 추출한다. 추출된 Feature Map은 Self-Attention 모듈을 통과하여 중요한 특성을 강조한다. 이후, 강조된 Feature Map은 Fully Connected Layer를 통과하여 추가적인 특성 통합과 차원 축소를 진행한다. 이렇게 변환된 Feature는 배경 및 장애물과의 차별성을 높이는 데 도움을 준다. 마지막으로,  $t$ 와  $t-1$  프레임의 각 객체를 비교하기 위해 변환된 Feature는 코사인 Similarity를 이용하여 계산한다.

**3.2 유사도 비교를 활용한 객체 추적**

기존 객체 탐지와 추적을 독립적으로 수행하는 방법과 달리, 제안하는 모델은 두 작업을 통합하는 방법을 선택했다. 이 구조는 파라미터와 계산량이 증가하지만, 객체 정보의 손실을 줄이고 강력한 성능을 보인다. 특히, 영상 내에서 객체 정보는 새롭게 등장하는 객체, 장애물 그리고 겹침 때문에 고유 정보를 쉽게 잃을 수 있다.

이러한 문제를 해결하기 위해, 본 연구에서는 유사도 비교 모듈을 통해 얻은 Similarity 값을 Transformer의 Self-attention 메커니즘을 거친 Track Query의 Value 값과 내적한다. 이러한 계산은 중요한 이유와 이점을 가진다. 첫째, 객체 중심점은 객체의 고유한 특성을 반영한다. 이는 객체의 위치, 크기, 그리고 형태에 대한 핵심 정보를 담고 있다. Transformer의 Value 값은 객체를 대표하는 다른 중요한 특성들을 포함하고 있다. 이 두 정보를 내적으로 결합함으로써, 객체 정보를 정확하고 풍부하게 표현할 수 있다. 둘째, 내적 연산은 정보의 일치도를 고려한다. 만약 객체의 중심점과 Value 값이 높은 내적 값을 가진다면, 이는 두 정보 사이에 높은 유사성이 있다고 해석할 수 있다. 즉, 객체가 이전 프레임에서의 정보와 현재 프레임에서의 정보가 일관되게 유지되고 있다는 것을 나타낼 수 있다. 이는 객체가 시간에 따라 얼마나 일관되게 나타나는지를 판단하는 데 도움이 된다.

객체 추적의 전체 흐름은 MOTRv2를 따른다. Transformer를 기반으로 하는 선행 연구들은 Detector 출력( $\Delta x, \Delta y, \Delta w, \Delta h$ )을 Embedding 하여 사용한다. 본 연구에서는 기존 Detector 출력에 유사도 비교 모듈의 출력을 통합하여 ( $\Delta x, \Delta y, \Delta w, \Delta h, \Delta c$ )를 출력하고자 한다. 이 값은 객체 연결을 위한 Decoder의 입력으로 활용된다.

객체 연결은 TBD 패러다임 내에서 큰 중요성을 지니고, 일치하는 부분에 대한 알고리즘에 따라 결과가 달라질 수 있다. ByteTrack은 이 문제를 해결하기 위해 간결하면서도 효과적인 방법을 제공하기 때문에 위 방법의 Matching Cascade를 사용한다[9].

#### 4. 결론

Transformer 기반의 MOT 연구들은 성능 향상과 실시간 추적을 위해 연구되고 있다. 본 연구는 객체의 Occlusion에도 강건한 성능을 달성하기 위해 기존 TBD 방식 기반의 유사도 비교 모듈을 추가한 모델을 설계했다. 제안하는 방법은 CenterNet의 출력 Center Peak를 유사도 비교 모듈의 입력으로 사용한다. 또한, 객체의 정보를 유지하기 위해 Transformer의 Value 값과 Similarity 값을 내적하여 Feature 정보를 강건하게 유지한다. 향후 이를 구현하여 탐지 과정에서 발생하는 속도 문제를 완화하고, 정보 전달손실을 감소시켜 실시간 Tracking을 하고자 한다.

#### 사사문구

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2022-2020-0-01489) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

#### 참고문헌

[1] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos and Ben Upcroft, "Simple online and realtime tracking", *IEEE international conference on image processing*, pp. 3464-3468, 2016.  
 [2] Nicolai Wojke, Alex Bewley and Dietrich Paulus, "Simple online and realtime tracking with

a deep association metric", *IEEE international conference on image processing*, pp. 3645-3649, 2017.

[3] Zhengcai Cao, Junnian Li, Dong Zhang, Mengchu Zhou and Abdullah Abusorrah, "Occlusion-aware real-time object tracking", *IEEE Transactions on Multimedia*, Vol. 19. No. 4, pp. 763-771, 2016.

[4] Zhengcai Cao, Junnian Li, Dong Zhang, Mengchu Zhou and Abdullah Abusorrah, "A multi-object tracking algorithm with center-based feature extraction and occlusion handling", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 24. No. 4, pp. 4464-4473, 2022.

[5] Zhang, Yuang, Tiancai Wang, and Xiangyu Zhang, "Motrv2: Bootstrapping end-to-end multi-object tracking by pretrained object detectors", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22056-22065, 2023.

[6] Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang and Qi Tian, "Centernet: Keypoint triplets for object detection", *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6569-6578, 2019.

[7] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng and Wenyu Liu, "Fairmot: On the fairness of detection and re-identification in multiple object tracking", *International Journal of Computer Vision*, pp. 1-19, 2021.

[8] Sun Peize et al, "Transtrack: Multiple object tracking with transformer", *arXiv preprint arXiv:2012.15460*, 2020.

[9] Zhang Yifu and et al, "Bytetrack: Multi-object tracking by associating every detection box", *In Computer Vision - ECCV 17th European Conference*, pp. 1-21, 2022.