

다중 객체 추적의 재인지 성능 개선을 위한 개체 쌍 추적 기법

*남다운, 임성용
한국전자통신연구원
*ndy9709@etri.re.kr

Paired Objects Tracking to Improve Re-Identification for Multiple Object Tracking

*Da Yun Nam, Seong Yong Lim
Electronics and Telecommunications Research Institute

요 약

다중 객체 추적 기술은 스포츠, 문화 예술 공연, VR 등 여러 방송 콘텐츠에서 자주 사용되고 있다. 방송 영상 안에 등장하는 여러 객체들은 객체간 상호작용에 의해 가려짐, 사라짐 (Occlusion) 등의 현상이 빈번하게 발생하고, 이 경우 기존에 추적되어온 객체들의 ID 가 소실되거나 교환되는 문제가 발생한다. 본 논문에서 더 강인한 다중 객체 추적을 위해, 주 개체 뿐만 아니라 주 개체에 종속되는 하위 개체 또한 함께 추적하는 개체-쌍-추적 기법을 제안한다. 한 쌍으로 묶인 주 개체와 종속 개체의 추적 정보와 매칭 정보는 상호보완적으로 사용되어, 소실 및 교환된 ID 도 복원할 수 있는 가능성을 높일 수 있다. 본 논문에서는 재인지 성능 향상을 위한 개체 쌍 추적 기법을 기술하였고, 성능 평가를 통해 제안 방법이 재인지 성능 향상에 기여할 수 있음을 확인하였다.

1. 서론

딥러닝 기반 객체 탐지 (Object Detection) [1][2] 기술이 발전해오면서, 객체 탐지 기술을 기반으로 객체 추적 기술 역시 활발히 연구되고 있다. 딥러닝 기반 다중 객체 추적 기술은 (Multi Object Tracking, MOT) [3][4] 기존 고전적 컴퓨터 비전 기술 기반 추적 기술 대비 뛰어난 성능을 달성하며, 다양한 산업계에서 활용되고 있다. 특히 스포츠 경기, 문화 예술 공연, VR/AR 등의 방송 콘텐츠에 적용되어 시청자들의 편의와 즐거움을 높이는 다양한 서비스를 제공하고 있다.

다중 객체 추적 기술은 주어진 시퀀스에서 하나 이상의 객체들을 동시에 추적하고, 각 객체의 궤적을 추출하고, 객체의 ID 정보를 기록하는 기능을 한다. 많은 응용 프로그램에서 트래킹되는 대표적인 추적 대상으로 ‘거리 위의 보행자’, ‘도로 위의 차량’ 등이 있다. 최근 딥러닝 기반 다중 추적 객체 기술은 두 가지 방향으로 발전되어 오고 있다. 첫 번째는 탐지 기반 추적 (Tracking by

Detection) [5]-[9] 모델이고, 두 번째는 공동 탐지 및 추적 (Joint Detection and Tracking, JDT) [10][11] 모델이다.

탐지 기반 추적 (Tracking by Detection) [5]-[9] 모델은 시퀀스의 매 프레임에 대해 객체를 탐지하고, 탐지된 객체 정보 (바운딩 박스 위치 및 크기, 객체의 클래스)를 이용하여 객체의 연결 정보를 찾아 추적한다. 일반적으로 객체 탐지 (Object Detection) 후에 시간상 인접한 프레임의 객체와 가장 상관성이 높은 현재 프레임의 객체를 찾아 연결 (Association) 하는 과정으로 동작된다. 대표적인 탐지 기반 추적 기술인 SORT [5] 은 프레임간 객체의 유사성을 계산하고, 칼만 필터를 사용하여 객체들의 움직임 예측하는 과정을 포함한다. 그리고 이러한 SORT 를 기반으로 한 기술들이 최근까지도 꾸준히 등장하고 있다 [6]-[9]. 공동 탐지 및 추적 (JDT) [10][11] 모델은 객체 탐지와 추적 기능을 동시에 수행하는 End-to-End 모델이다. 최근에는 Transformer 기반의 JDT 기법이 많이 등장하고 있다. Transformer 기술 발전에 따라 다

중 객체 추적의 정확도 및 처리속도 향상에 좋은 결과를 보이고 있으며, 높은 발전 가능성을 지닌 모델이다.

다중 객체 추적 분야에서 범용적으로 가장 많이 사용되고 있는 벤치마크 데이터 셋인 MOT17 [12], MOT20 [13], DanceTrack [14]에 대해서는 다양한 다중 객체 추적 알고리즘들의 추적 성능 결과를 확인할 수 있다. 그리고 현재까지 가장 우수한 성능을 달성한 기술들은 탐지 기반 추적 (Tracking by Detection) 기술들이라는 것을 확인할 수 있다. 하지만 탐지 기반 추적 (Tracking by Detection) 기술들은 객체의 사라짐, 가려짐 (Occlusion) 이 발생하는 경우, 쉽게 ID 가 소실되거나, 교환될 수 있다는 한계점을 가지고 있다. 일반적인 방송 영상은 Occlusion 이 매우 빈번하게 발생하고, 해당 영상에 다중 객체 추적 기술을 적용하면 동일 객체에 대해 동일 ID 를 오랫동안 유지하지 못한다. 따라서 다중 객체 추적에서는 사라지거나 가려진 객체에 대해 동일 ID 를 부여할 수 있는 재인지 성능의 개선이 필요하다.

본 논문은 하나의 객체를 추적하기 위해, 주 객체에 대한 추적을 하는 동시에 주 객체에 종속된 하위 객체도 함께 추적하는 기술을 제안한다. 주 객체와 종속 객체는 한 쌍으로 매칭하여 추적 정보를 연결할 수 있다. 예를 들어, 사람추적 분야에서 주 객체인 ‘사람’에 대한 추적이 실패하더라도, 종속 객체인 ‘머리’에 추적이 계속 진행되는 경우라면, ‘머리’의 매칭 정보와 추적 정보를 이용하여 ‘사람’의 ID 를 복원할 수 있고, 마찬가지로 ‘머리’의 추적이 실패한 경우에도 ‘사람’의 추적 정보를 활용하여 ‘머리’의 ID 를 다시 복원할 수 있는 가능성이 높아진다. 즉, 하나의 객체를 추적하기 위해 한 쌍으로 묶인 두 객체의 추적 정보를 연결해가면서, 재인지 (Re-Identification) 성능을 강화시킬 수 있다.

본 논문의 구성은 다음과 같다. 2 장에서는 제안 방법의 핵심 아이디어 및 전체적인 과정을 설명하고, 3 장에서는 제안 알고리즘의 성능 검증을 위한 실험 결과를 보이고, 마지막으로 4 장에서 본 논문의 결론을 맺는다.

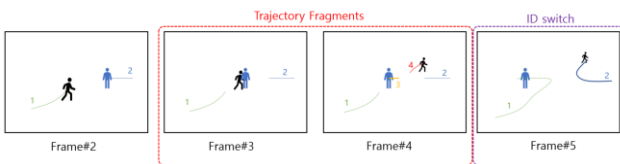


그림 1. 다중 객체 추적 과정에서 발생하는 ID 스위치 문제

2. 제안 방법

제안 방법은 탐지 기반 추적 (Tracking by Detection) 기술에 속한다. 하지만 기존 탐지 기반 추적 기술들이 ‘주 객체’에 대해서만 탐지 및 추적되었던 반면, 제안 기술은 ‘주 객체’와 이에 종속된 ‘종속 객체’에 대해 모두 탐지 및 추적을 수행하고, ‘주 객체’와 ‘종속 객체’ 를 한 쌍으로 매칭하는 과정을 포함한다. 한 쌍으로 매칭된 두 객체 [주 객체, 종속 객체] 중

하나가 소실되더라도 다른 하나의 추적이 진행되고 있다면, 매칭 정보를 통해 소실된 객체의 ID 를 복원시킬 수 있다. 즉 한 쌍으로 묶인 두 객체의 추적정보는 상호 보완적으로 적용되어, 기존에 추적되었던 객체가 ID 가 소실/교환/재생성되어 추적에 실패하더라도, 매칭되는 다른 쪽 객체에 추적 및 매칭 정보를 사용해, 원래 ID 를 복원할 수 있는 가능성을 높일 수 있다.

제안 방법의 전체적인 과정을 보면, 가장 먼저 주 객체와 종속 객체에 대해 독립적으로 객체 탐지 (Object Detection)과 객체 추적 (Object Tracking)을 한다. 그 다음, 주 객체와 종속 객체를 하나의 쌍으로 매칭시킨다. 이 후 한 쌍의 객체 쌍 [주 객체, 종속 객체] 중 한 객체에서만 추적이 실패가 발견된다면, 다른 쪽 객체의 추적 정보를 활용할 수 있는지 검사하고, 조건에 맞는 경우, 추적 실패한 객체의 ID 를 복원한다.

본 논문의 아이디어를 사람 추적 분야에 적용하면, 주 객체는 ‘Person’, 종속 객체는 ‘Head’ 로 지정하여 그림 3 과 같은 과정으로 동작된다. 제안 알고리즘은 기존 객체 탐지기술 [1][2]을 적용하여 객체를 탐지한다 (Person, Head Detection). 이 후, 각 Head 와 Person 클래스로 감지된 객체들에 대해 인접 프레임 간의 유사성 높은 객체와 연결시켜 추적 정보를 갱신한다. (Person, Head Association). 이 후, 추적된 Person 과 Head 객체에 대해 하나의 쌍으로 매칭시킨다 (Person-Head Pair Matching). 그리고 추적에 실패한 Person (또는 Head)가 존재할 경우, 이와 매칭되는 Head (또는 Person)의 추적 정보를 활용할 수 있는지 조사하고, 참일 경우 현재까지 기록된 Head-Person 쌍 매칭 정보를 이용해 소실된 Person (또는 Head)의 추적 정보를 복원한다 (Person and Head Track Recovery).

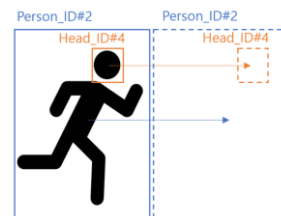


그림 2. 한 쌍의 Person - Head 객체들에 대한 동시 추적 기능

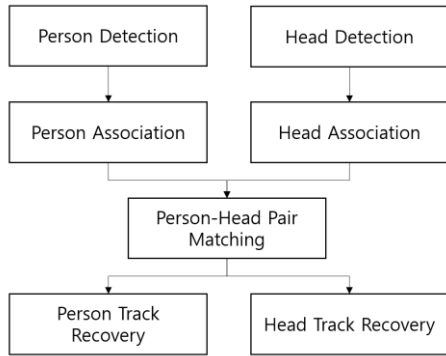


그림 3. 제안 알고리즘의 동작 과정 (프레임 단위로 순차적으로 동작)

3. 실험 결과

본 장에서는 제안 방법의 성능 검증을 위한 실험 결과를 보여준다. 제안 알고리즘의 구현을 위해, 객체 탐지 (Object Detection) 방법으로 YOLOv5 [1]을 사용하였고, 객체 추적 (Association) 알고리즘으로 ByteTrack [7] 방법을 사용하였다. 이 때 YOLOv5 는 CrowdHuman 데이터 셋을 이용하여 'Person'과 'Head'를 탐지할 수 있도록 학습되었다. 제안 알고리즘의 성능 검증을 위한 비교 알고리즘으로, 최신 Tracking by Detection 기술인 ByteTrack [7]과 OC-SORT [9] 를 사용하였다. 이 중 ByteTrack [7] 은 MOT17 [12] 과 MOT20 [13] 데이터 셋에서, OC-SORT[9]는 DanceTrack [14] 에서 가장 우수한 성능을 달성했던 기술이다.

공정한 실험을 위해 CrowdHuman 데이터 셋으로 학습된 YOLOv5 모델을 공통된 객체 탐지 모델로 사용하였다. 실험 데이터 셋으로 사용한 Seq1, Seq2, Seq3, Seq4, Seq5 는 모두 DanceTrack 훈련용 데이터 셋에 포함된 시퀀스이며, 각각의 시퀀스는 dancetrack0001, dancetrack0006, dancetrack0015, dancetrack0016, dancetrack0027 영상에 해당한다. 표 1 은 시퀀스 내 등장한 사람 수, 해상도, 프레임 레이트 등의 시퀀스 정보를 확인할 수 있다.

표 1. 성능평가과정에서 사용된 시퀀스 정보

	Seq1	Seq2	Seq3	Seq4	Seq5
People Count	7	9	9	6	9
Resolution	1280 x 720	1920 x 1080	1280 x 720	1920 x 1080	1280 x 720
Frame Rate	20	20	20	20	20
Sequence Length	703	1202	1203	2163	403

표 1 은 하나의 시퀀스 내에서 추적된 Trajectory 의 총 개수를 비교한 결과를 보여준다. 이때 추적된 Trajectory 의 총 개수는 한 시퀀스 내 생성된 ID 의 총 개수를 의미한다. 다중 객체 추적 기술이 높은 추적 정확도를 가지고 있다면, 생성된 ID 개수 (혹은 Trajectory 개수)는 실제 영상 내 등장한 객체의 수가 가까울 것이다.

표 2. MOT 알고리즘 별 시퀀스당 생성된 Trajectory 개수 비교

	Seq1	Seq2	Seq3	Seq4	Seq5
ByteTrack	25	42	16	37	30
OC-SORT	22	38	14	27	39
Proposed method	21	37	14	27	28

표 2 에서 제안 알고리즘이 비교적 적은 추적된 객체 수를 가지는 것을 확인할 수 있고, 이는 객체 간 ID 스위치가 비교적 적게 일어났기 때문이라고 예측할 수 있다.

또한 결과 동영상을 비교하여, 제안하는 다중 객체 추적 기술의 재인지 성능을 확인하였다. 그림 4 와 그림 5 는 일부 프레임에 대한 객체 추적 결과를 보여준다. 그림 4 와 그림 5 는 각각 771 번째 프레임과 596 번째 프레임에서 Occlusion 이 발생한다. 기존 방법의 결과는 이후 프레임에서 ID 를 유지하지 못하고 새로운 ID 가 부여되는 반면 제안 방법을 적용한 결과에서는 동일 ID 로 복원됨을 확인할 수 있다.

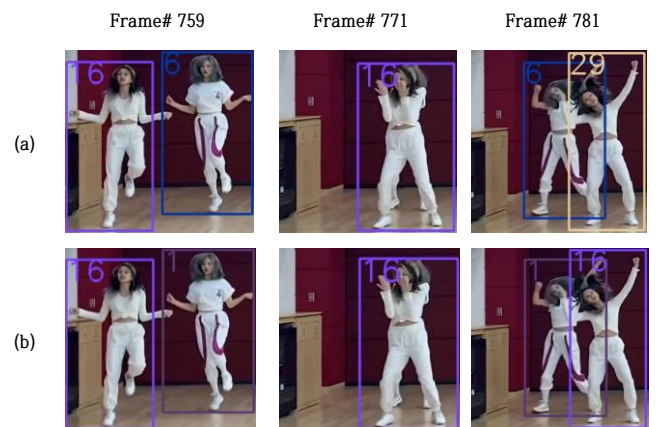


그림 4. Seq2 에 대한 다중 객체 추적 결과영상: (a) ByteTrack 적용 결과, (b) 제안 기술 적용 결과



그림 5. Seq4 에 대한 다중 객체 추적 결과영상: (a) ByteTrack 적용 결과, (b) 제안 기술 적용 결과

표 3. MOT 알고리즘 별 실행 시간 비교 (Frame/Sec)

	Seq1	Seq2	Seq3	Seq4	Seq5
ByteTrack	56.38	40.60	50.80	41.83	47.71
OC-SORT	44.11	28.28	38.99	26.56	36.81
Proposed method	48.18	29.76	46.98	36.40	39.80

표 2 와 표 3 의 실험 결과를 통해, 제안 방법이 실제 객체 수에 가장 가깝게 추적 경로 (Trajectory) 를 기록하는 반면, 최신 다중 객체 추적 기술들과 비교하여 큰 속도 저하 없이 동작할 수 있음을 확인할 수 있다.

4. 결론

본 논문은 다중 객체 추적 방법으로, 주개체와 종속개체를 동시에 추적하여 한 쌍으로 묶고, 두 개체를 상호보완적으로 이용하여 ID 를 복원율을 높이는 개체 쌍 추적 방법을 제안하였다. 실험 결과를 통해, 제안 방법이 소실된 ID 를 다시 복원시키는데 효과가 있음을 확인하였다. 더 나아가 본 논문의 방법이 긴 시퀀스에서 장기간 추적의 성능개선을 위해 활용되는 것을 기대해볼 수 있다. 현재까지 주 개체, 종속 개체를 모두 탐지 및 추적하는 과정에서 속도가 다소 느려진다는 한계점이 있으나, 향후 다중 객체 추적 기술 연구를 계속하여, 제안 알고리즘의 한계점을 보완하고 추적 성능을 개선시킬 계획이다.

감사의 글

본 연구는 2021 년도 정부(과학기술정보통신부)의 재원으로

정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-02084, 비대면 실감 경험 공유를 위한 XR, Volumetric 실감미디어 생성, 전송 기술 개발 및 한-유럽 국제공동연구)

참고문헌

- [1] [online] <https://github.com/ultralytics/yolov5>
- [2] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, Jian Sun, "YOLOX: Exceeding YOLO Series in 2021", CVPR2021, 2021
- [3] Park, Y.; Dang, L.M.; Lee, S.; Han, D.; Moon, H. Multiple Object Tracking in Deep Learning Approaches: A Survey. *Electronics*.<https://doi.org/10.3390/electronics10192406>
- [4] Patrick Dendorfer, Aljoša Ošep, Anton Milan, Konrad Schindler, Daniel Cremers, Ian Reid, Stefan Roth, Laura Leal-Taixé, "MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking", IJCV, 2020.
- [5] A. Bewley, Z. Ge, L. Ott, F. Ramos and B. Upcroft, "Simple online and realtime tracking," 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 3464-3468, doi: 10.1109/ICIP.2016.7533003.
- [6] Nicolai Wojke, Alex Bewley, Dietrich Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric", 2017. arXiv:1703.07402
- [7] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, Xinggang Wang, "ByteTrack: Multi-Object Tracking by Associating Every Detection Box", 2022, arXiv:2110.06864
- [8] Yunhao Du, Yang Song, Bo Yang, Yanyun Zhao, "StrongSORT: Make DeepSORT Great Again", 2022, arXiv:2202.13514
- [9] Jinkun Cao, Xinshuo Weng, Rawal Khirodkar, Jiangmiao Pang, Kris Kitani, "Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking", 2022, arXiv:2203.14360
- [10] Fangao Zeng, Bin Dong, Yuang Zhang, Tiancai Wang, Xiangyu Zhang, Yichen Wei, "MOTR: End-to-End Multiple-Object Tracking with Transformer", 2022.arXiv:2105.03247
- [11] Liting Lin, Heng Fan, Yong Xu, Haibin Ling, SwinTrack: A Simple and Strong Baseline for Transformer Tracking, 2021. arXiv:2112.00995
- [12] [online] <https://motchallenge.net/data/MOT17/>
- [13] [online] <https://motchallenge.net/data/MOT20/>
- [14] [online] <https://github.com/DanceTrack/DanceTrack>