

마스크 기반 재식별 최적화와 ID-Switch 개선 방안 연구

유수빈, 진광, 이도훈*
부산대학교

seal498@pusan.ac.kr, geemguang@pusan.ac.kr *dohoon@pusan.ac.kr

A Study on mask-based re-identification optimization and ID-Switch improvement methods

Yu Su Bin, Jin Guang, Lee Do Hoon*
Pusan National Univ.

요 약

본 논문은 Multi-Object Tracking(MOT)의 정확도를 높이고 Occlusion(폐색)으로 인한 ID(Identity) Switch의 발생을 줄이기 위한 Re-identification(Re-ID) 방법을 제안한다. 마스크 처리한 입력 이미지를 인코딩하여 객체의 특징을 추출하고 유사도를 계산하는 방법으로, 실험적으로 성능이 가장 좋은 매개변수를 결정하여 Tracker에 적용시킨다. 새로운 ID를 부여하기 전에 추적 경로를 잃은 객체들과 특징을 비교하여 유사도를 계산하고 동일한 객체라고 판단될 경우 경로를 재활성화 한다. 본 연구의 결과는 총 ID Switch를 약 20% 감소시켰고 이는 객체를 더 정확하게 추적한다.

I. 서론

Multi Object Tracking(MOT)은 비디오 내에서 여러 객체를 동시에 감지하고 추적하는 정교한 컴퓨터 비전 기술이다. MOT의 핵심 목적은 한 비디오의 여러 프레임에서 동일한 객체를 일관되고 정확하게 식별하고 그 궤적을 모니터링하는 것이다.[1] MOT에서는 성능 개선을 위해 폐색(Occlusion)과 외관 변화와 같은 문제를 해결하는 것이 필수적이다. 외관의 변화는 얼굴 특징, 귀 모양, 걸음걸이와 같은 생체 특징을 활용하여 추적해야 한다. 폐색이 발생하면 동일한 객체의 identity(ID)가 변경되는 ID Switch와 궤적이 끊어지는 fragmentation이 발생한다. ID Switch의 발생 횟수는 MOT의 평가 지표로도 사용된다. 폐색이 발생한 후 ID Switch를 줄이는 것은 MOT 성능 향상에 필수적이다.

MOT에서 현재 State-Of-The-Art (SOTA) Tracker는 여러 가지 방법으로 폐색을 해결한다. 추적 객체에 이중 임계값을 설정하여 추적 범위를 확장하거나[2], 폐색으로 인한 노이즈를 최소화하기 위해 Observation-centric Re-Update(ORU)를 적용하고, 객체 위치와 모양보다 덜 중요한 객체의 깊이와 속도 같은 보조 증거를 고려하여 추적한다. 그러나 현재의 SOTA Tracker는 폐색을 해결하지만 ID Switch를 주요 평가지표로 사용하지 않는다.

추적 시스템은 다양한 분야의 자동 관제 시스템에서 사용된다. 그러나 폐색과 같은 문제가 발생하는 복잡한 환경에서는 인간의 개입이 필요하다. 완전한 자동화를 위해 이러한 시스템은 인간의 개입 없이 정확하고 독립적으로 객체를 감지하고 추적해야 한다. ID Switch를 줄이는 것은 시스템 자율성에 매우 중요하다.

본 논문에서는 MOT에서 ID Switch를 줄이는 것을 목표로 새로운 Re-identification(Re-ID) 방법을 소개하고, 해당 방법을 적용한 Tracker의 유효성을 제시한다. Re-ID 방법은 입력 이미지를 패치로 분할하여 선택적으로 마스크 처리한 후 인코더를 사용하여 객체 특징을 추출한다. 실험을 통해 마스크 전략, 비율, 패치 크기 및 특징 벡터 크기를 포함한 다양한 매개변수를 결정하고, Re-ID를 위한 유사도 계산 방식을 결정한다. 실험적으로 112x112 입력 이미지를 8x8 사이즈 패치로 나누고 60% 마스크 비율로 센터 중심 마스크하여 512 차원의 특징 벡터 추출하고, 맨해튼 거리를 사용하여 특징 벡터의 유사성을 측정했을 때 76.75%의 Top-5 Accuracy로 가장 높은 성능을 보인다. 해당 매개변수의 Re-ID 방법을 Tracker에 적용하여 기존의 Tracker 성능과 비교할 때 SOTA Tracker의 성능과 비슷하며 ID Switch는 감소한다. 총 ID Switch가 169회에서 136회로 20% 감소하였다.

II. 본론

II. I. 재식별과 Tracker

본 논문에서는 Re-ID를 위해 마스크 처리된 이미지를 인코딩하여 특징 벡터를 추출하고, 유사도를 계산하는 방법을 제안한다. 사용된 인코더는 Vision Transformer(ViT)로, 입력 이미지를 특정 패치 크기로 나누어 인코더에 입력한다.[3] 연산 속도 향상과 객체의 정보를 최대한으로 입력 받기 위해 입력 이미지를 일정 비율 마스크한다. 마스크되지 않은 패치에 대해 특징 벡터를 추출한다. 제안하는 Re-ID 방법은 입력 이미지 사이즈, 패치 사이즈, 마스크 패턴, 마스크 비율, 특징 벡터 차원, 유사도 계산 방법에 따라 성능이 달라진다. 그림 1과 같이 마스크 패턴의 경우 랜덤, 격자패턴, 테두리, 중심부

마스킹으로 총 4 가지이며, 0 부터 90%까지 10%간격으로 마스킹 비율을 설정한다. 특징 벡터 간의 유사도 계산 방법은 코사인 유사도와 유클리드 거리, 맨해튼 거리로 계산한다.

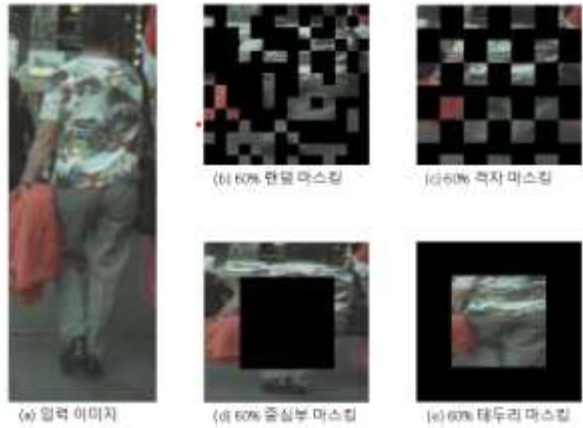


그림 1(a) 입력 이미지 원본, (b)~(e) 4 가지 마스킹 패턴 예시

Tracker 는 SOTA 를 기록한 ByteTrack 으로 채택한다. ByteTrack 은 칼만 필터를 이용하여 추적하는 SORT 를 응용한 Tracker 로, 추적 대상을 정하는 탐지 임계값을 이중임계값으로 설정하여 추적 대상 범위를 넓히고 칼만 필터를 이용하여 다음 위치를 예측하는 방식이다. 해당 Tracker 에 제안한 Re-ID 방법을 적용하여 새로운 Tracker 구조를 고안한다. 폐색 후 다시 나타난 객체가 이전 추적경로를 잃어 새 ID 를 부여하기 전에 특정 범위 내의 프레임에서 추적 경로를 잃은 객체들의 특징과 비교하여 동일한 객체를 찾아낸다. 기존의 Tracker 는 탐지 이중임계값 중 강한 임계값보다 높은 점수를 가지는 객체에 대해 추적이 되지 않은 경우 새 ID 를 부여한다. 이때 추적 방법은 추적 활성화된 객체들의 위치를 칼만 필터로 예측하고 탐지 객체와 매치한다. 본 논문에서 제안된 Tracker 는 해당 단계에서 새 ID 를 부여하기 전 일정 시간 이상 추적되지 않아 추적 경로를 잃은 객체들과 Re-ID 방법을 이용하여 유사도를 검사하고 유사도가 높고 기준치 이상일 때 동일 객체로 판단하여 경로를 재활성화한다. 실험은 MOT Benchmark dataset 인 MOT17 중 GroundTruth 가 공개된 것을 사용한다.[4]

	IDF1	ID Switch(IDs)
ByteTrack	86.4	169
Track (ours)	80.7	136

II. II. 결과

표 1 ByteTrack 과 제안된 Tracker 의 성능 비교

실험적으로 얻은 매개변수에 따른 Re-ID 결과는 112x112 입력 이미지, 8x8 패치 사이즈, 60% 마스킹 비율, 센터 중심 마스킹, 512 차원의 특징 벡터, 맨해튼 거리를 사용하여 특징 벡터의 유사성을 측정했을 때 76.75%의 Top-5 Accuracy 로 가장 높은 성능을 보인다. 제안된 Tracker 의 성능은 표 1 과 같다. 기존의 ByteTrack 과 제안된 Tracker 의 추적 성능을 비교할 때 IDF1 의 값이 낮아졌다. 그러나 ID Switch 횟수는 감소하였으며 표 2 의 영상별 ID Switch 비율을 보면 ID

Switch 비율이 1.1%로, 가장 높은 MOT17-05 에서 ID Switch 가 0.4% 감소하였다.

	ByteTrack	Track (ours)
	IDs (%)	
MOT17-02-FRCNN	0.6	0.6
MOT17-04-FRCNN	0.0	0.0
MOT17-05-FRCNN	1.1	0.7
MOT17-09-FRCNN	0.7	0.7
MOT17-10-FRCNN	0.5	0.3
MOT17-11-FRCNN	0.4	0.2
MOT17-13-FRCNN	0.2	0.2

표 2 MOT17 Dataset 영상 별 IDs 비율. 두 값 중 높은 것은 빨간색, 낮은 것은 파란색으로 표시함.

III. 결론

본 논문에서는 MOT 를 위한 Re-ID 방법을 제안하고, Tracker 에 적용시켜 유효성을 입증했다. Re-ID 방법의 경우 마스킹 여부에 따라 정확도가 크게 차이났다. 마스킹을 하지 않은 경우 Top-5 정확도가 68%, 마스킹 비율이 60%일 때 Top-5 정확도가 76% 였다. Re-ID 방법을 적용한 Tracker 는 기존의 Tracker 에 비해 성능은 낮아졌지만, 추적 객체의 ID Switch 를 줄여 추적 객체들의 구분이 명확해졌다.

ACKNOWLEDGMENT

본 연구는 「부산 연구산업진흥단지 육성사업(과학기술정보통신부, 부산광역시)」의 재원을 지원받아 수행하였음(2023-02-연구산업진흥단지육성-부산-1 세부-산학연 1-2).

참 고 문 헌

- [1] Sun, S., Akhtar, N., Song, H., Mian, A., Shah, M., "Deep affinity network for multiple object tracking", 2019, (https://arxiv.org/abs/1810.11780).
- [2] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X., "Bytetrack: Multi-object tracking by associating every detection box", 2022, (https://arxiv.org/abs/2110.06864).
- [3] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Housby, N., "An image is worth 16x16 words: Transformers for image recognition at scale", 2021, (https://arxiv.org/abs/2010.11929).
- [4] Dendorfer, P., Oš ep, A., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S., Leal-Taixé, L., "Motchallenge: A benchmark for single-camera multiple target tracking ", 2020, (https://arxiv.org/abs/2010.07548).