다중객체추적(Multiple Object Tracking) 기술 동향 분석

서 찬 호 , Nguyen Cong Quy, 허 동 욱 , 최 성 록
(서울과학기술대학교 컴퓨터공학과)

1. 서론

다중객체추적(multiple object tracking; 이하 MOT) 기술은 센서 데이터를 이용하여 관찰 공간 내에 존재하는 여러 (움직이는) 물체를 구분하여 동일 물체를 시간이 지나도 연속하여 구분하는 기술이다. 물체를 단순히 감지(detection)하는 것을 넘어 물체가 움직이더라도 동일한 물체를 지속적으로 추적(tracking)하는 MOT 기술은 물체의 과거 궤적과 현재 위치/속도 뿐만 아니라 미래의 예상 경로를 예측할 수 있기 때문에 다양한 자동화 시스템에 활용되는 중요한 기술이다. CCTV 카메라의 영상에서 관찰된 다수의 사람이나 차량을 추적하여 공간의 흐름(traffic)이나 혼잡도를 추정할 수 있고 이상상황(abnormality)을 감지할 수도 있다. 스포츠 경기에서도 선수들과 공의 움직임을 추적하여 개별 선수의 운동량이나 특징, 기여도 뿐만 아니라 각 팀의 전술적 상황도 파악할 수도 있다. 또 자율주행 자동차나 로봇, 드론은 MOT 기술을 이용해 주변 환경에 있는 물체를 고려한 경로 계획이나 장애물 회피를 할 수 있다.

MOT 기술은 물체의 특성과 센서의 종류에 따라 컴퓨터비전, 로봇공학, 제어공학 등과 같은 분야에서 다양한 방법론으로 연구되고 있다. MOT 문제는 물체의 개수나 움직임의 복잡도, 형태의 유사도에 따라 난이도가 달라진다. 즉 MOT 문제는 물체의 개수가 많을수록, 움직임이 불규칙적이며 클수록, 또 물체의 형태가 유사할수록 매우 복잡한 문제가 된다. 특히 물체는 일반적으로 움직이고, 물체가 센서 데이터의 시야를 벗어나 사라지기도 하고, 시야 내에 새로운 물체가 관찰되기도 한다. 또 주변 구조물이나 움직이는 물체로 인해 가려짐이나 겸침이 발생할 수 있기 때문에 더욱 복잡한 문제이다. 또 MOT에 사용되는 센서 데이터의 형태나 상황도 다양하다. LiDAR 데이터와 같이 물체의 3차원 위치를 바로 얻을 수도 있고, 영상(image) 데이터와 같이 물체가 2차원 평면에 투영되어 거리 정보가 사라지는 경우도 있다. 또 CCTV 카메라와 같이 센서가 고정되어 있는 경우도 있고, 자율주행 자동차에 설치된 카메라와 같이 센서가 이동하는 경우도 있다. 특히 MOT 기술은 그 응용에 따라 실시간(real-time) 처리가 반드시 필요한 경우가 있다. 따라서 MOT 기술의 정확도와 계산시간 사이의 조절(trade-off)은 주어진 상황의 복잡성과 함께 MOT 방법론 선택에 있어 중요한 이슈이다.

본 기고문에서는 MOT 기술의 방법론에 따른 최신 연구들을 리뷰하고 MOT 기술이 응용된 최근 예들을 소개하고자 한다. 우선 2장에서는 MOT 문제의 일반적인 수학적 정의를 살펴보고, 이를 통해 MOT 기술의 세 가지 방향의 접근법을 소개한다. 또 각 접근법의 대표적인 연구 아이디어나 최근 연구 결과를 리뷰한다. MOT 기술은 다양한 제품과 서비스에 활용되는데, 3장에서는 지능형 영상감시(intelligent visual surveillance)와 자율주행 분야에서의 최근 응용 예들을 살펴본다. MOT 기술은 요소기술

로써 다양한 연구와 응용이 이뤄지고 있다. 본 기고문에서는 최근의 학술적 연구 결과와 산업적 활용 예를 공유하여 MOT 기술의 선택과 활용에 대한 통찰을 제공하고자 한다.

2. MOT 기술 연구 동향

2.1. MOT 문제 정의

MOT 문제는 센서 데이터를 이용해 각 물체를 구분하고 상태를 추정하는 다변량 추정 문제[1]로 볼 수 있다. 이를 수학적으로 표현하면 다음과 같다. t 번째 시간에 획득한 센서 데이터 D_t를 통해 얻은 i번째 관측치를 O_t^i로 표현한다. 관측치는 물체의 위치(예: bounding box)나 종류(class), 특징벡터 또는 형태 데이터(예: image patch 또는 point cloud) 등을 포함한다. t번째 시간에 M_t개의 모든 관측치를 $O_t = \{o_t^1, o_t^2, \cdots o_t^{Mt}\}$로 표현한다. 또 1 번째 시간에서 t 번째 시간까지 모든 관측치를 $O_{1:t} = \{O_1, O_2, \cdots, O_t\}$로 표현한다. MOT를 통해 얻고자 하는 것은 각 관측치의 물체의 상태이다. t번째 시간에서 i번째 물체의 상태를 s_t^i로 표현한다. 물체의 상태는 물체를 구별할 수 있는 식별자(identifider; 이하 ID)를 비롯하여 물체의 종류(class), 위치, 속도 등을 포함할 수 있다. t 번째 시간에 M_t 개의 모든 물체의 상태를 $S_t = \{s_t^1, s_t^2, \cdots, s_t^{M_t}\}$로 표현한다. 또 1번째 시간에서 t번째 시간까지 모든 물체의 상태를 $S_{1:t} = \{S_1, S_2, \cdots, S_t\}$로 표현한다. 따라서 MOT는 t번째 시간의 모든 관측치 O_t에 대한 물체의 상태 S_t를 찾아야 하고, 이를 전체 시간으로 확대하면 아래와 같은 MAP (maximum a posteriori) 최적화 문제로 표현할 수 있다.

$$\hat{S}_{1:t} = \arg\max_{S_{1:t}} P(S_{1:t} \mid O_{1:t})$$

위의 MOT 문제의 정의는 특정 시점(t)에서 과거($t-k$)와 미래($t+k$)에 얻은 관측치를 모두 함께 이용할 수 있는 오프라인(offline) 기법을 포함하는 수식이다. 만약 현재의 관측치 O_t와 직전($t-1$)까지의 상태 추정 결과 $S_{\{1:t-1\}}$만을 이용하는 온라인(online) 기법은 아래와 같이 표현할 수 있다.

$$\hat{S}_t = \arg\max_{S_t} P(S_t \mid S_{1:t-1}, O_t)$$

MOT 문제에서 가장 중요한 부분은 동일 물체에 동일한 물체 ID를 계속 할당하는 것이다. 예를 들어, 온라인 추적기법에서는 o_t^i에 해당하는 s_t^i를 추정할 때, 해당 물체가 이전에 관찰되었다면 해당 물체를 $S_{1:t-1}$에서 찾고, s_t^i가 같은 해당 ID가 되도록 할당하여야 한다. 이러한 정합(association) 문제는 MOT의 핵심이다. 이러한 정합 문제를 풀기 위해 일반적으로 아래와 같은 물체의 움직임과 형태에 대한 조건들을 사용하고, 각각 운동모델(motion model)이나 형태모델(appreance model)의 형태로 MOT에 구현된다.

- 물체의 움직임은 연속적이므로 현재 위치와 직전의 위치는 가깝거나 운동 법칙에 의해 표현할 수 있다. → 운동모델
- 물체의 형태는 변화가 많지 않거나 연속적으로 변하므로 현재 형태와 직전의 형태는 유사하고 특정 지표를 통해 그 유사도나 차이를 정량하여 표현할 수 있다. → 형태모델

운동모델은 아래와 같이 물체의 상태와 관측치를 통해 다음 순간의 물체의 상태/관측치(예: 위치나 방향)를 추정할 수 있고, 이를 현재의 관측치와 비교하여 정합에 활용할 수 있다.

$$s_t, o_t = f_{\text{motion}}(s_{t-1}^i, o_{t-1}^i)$$

형태모델은 주어진 센서 데이터 D_t 물체의 관측치 O_t를 생성한다. 또 형태모델을 바탕으로 두 관측치(예: 특징벡터) 사이의 차이는 유사도 함수나 거리 함수 $\text{dist}(o^i, o^j)$를 통해 정량할 수 있고, 이를 이용하여 정합에 활용할 수 있다.

$$O_t = h_{\text{appearance}}(D_t)$$

운동모델과 형태모델은 운동학(kinematics)기반의 모델이나 특징추출 알고리즘이나 물체감지 알고리즘 등을 통해 구현될 수 있고, 최근에는 심층신경망(DNN)을 통해 표현되고 데이터셋을 통해 학습하기도 한다. 본 기고에서는 물체의 움직임 정보를 효과적으로 활용하는 연구와 물체의 형태 정보를 효과적으로 활용하는 연구, 그리고 심층신경망의 학습을 통해 MOT를 수행하는 중요 연구들을 리뷰하고자 한다.

2.2. 움직임(Motion) 기반 MOT 접근법

SORT (ICIP 2016)

Simple Online and Realtime Tracking (이하 SORT) [2]는 간단한 구조로 빠르고 정확하게 동작하여 이후 많은 MOT 기술에 영향을 미쳤다. SORT는 검출 기반 추적(tracking-by-detection) 기법으로 다음과 같이 동작한다. 1) 물체검출(object detection) 알고리즘을 통해 물체를 바운딩 박스(bounding box) 형태로 검출한다. SORT 논문에서는 Faster R-CNN 물체검출기를 사용하였는데, 다른 물체검출기를 가능하다. Github의 많은 구현들은 Faster R-CNN보다 빠르게 동작하는 YOLO 기반 물체검출기를 사용한다. 2) 칼만필터(Kalman filter)를 이용해 다음 프레임의 바운딩 박스를 예측한다. 다음 프레임의 바운딩 박스 예측에는 등속운동모델(constant velocity model)을 사용한다. 3) 두 프레임에 관찰된 객체들의 동일성은 cost matrix 형태로 표현한다. 두 객체의 동일성 판단하는 지표로 두 물체의 바운딩 박스의 IoU를 이용한다. 4) 헝가리안 알고리즘(Hungarian algorithm)을 이용하여 cost matrix를 고려한 최적의 정합 결과를 도출한다. 5) 이후 정합이 된 것과 정합이 되지 않은 것들을 다양한 조건에 따라 분류하여 새로운 물체나 사라진 물체 등으로 후처리한다. SORT는 물체의 형태 정보를 사용하지 않는 물체의 움직임 정보만을 사용하는 MOT 기술이다. 물체가 느리게 움직이거나 또는 카메라의 초당 프레임 수(framerate)가 빠른 경우, 물체의 이동 변위는 크기지 않고 SORT는 빠르고 효과적으로 물체를 매칭할 수 있다. 그러나 형태 정보를 전혀 사용하지 않기 때문에, 객체의 겸칩이나 가려짐에 매우 취약한 단점을 가지고 있기 때문에 혼잡한 환경에 사용하기에 부적합하다.

UCMCTrack (AAAI 2024)

Uniform Camera Motion Compensation Track (이하 UCMCTrack) [3]은 기존 MOT가 이미지 평면(image plane) 상에서 정합하는 방식을 한계로 지적하였다. 특히 이 문제는 물체 가림 문제에 대한 직접적 원인이기도 하며, 카

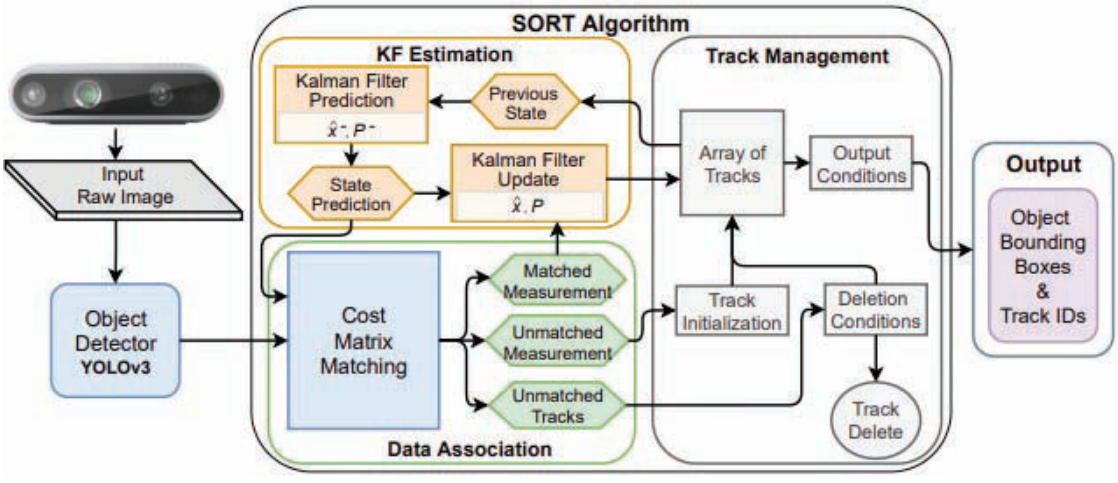

그림 1. SORT의 동작[4].

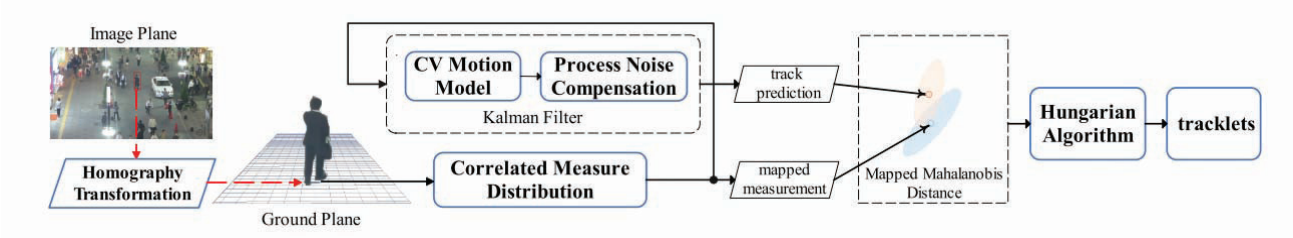

그림 2. UCMCTrack의 동작[3].

메라의 심한 흔들림(camera jittering)에 대한 취약성을 내포하기 때문에 그 한계가 강조된다. 이를 해결하고자 이미지 평면상의 움직임 정보를 3차원 평면(ground plane)으로 옮기는 과정을 도입하여 문제를 해결하고자 한다. 1) 먼저 이미지 내 위치 정보 값을 호모그래피(homography)를 이용해 전역 좌표계(ground plane)로 옮긴다. 2) 그 후 이미지 평면상의 측정 오류(measurement error)를 3차원상으로 투영한다(correlated measure distribution; CMD). 3) 3차원상에 투영된 값을 가지고 마할라노비스 거리(Mahalanobis distance)를 적용해 이를 비용 요소로 사용한다(mapped Mahalanobis distance; MMD). 칼만필터에서는 등속운동모델을 사용한다. 이때 카메라 떨림에 대한 가속도계를 모델링하여 이를 칼만필터 내 공분산 행렬에 적용한다. 상기된 방법론을 통하여 UCMCTrack 은 3차원상에서 물체의 움직임을 가지고 정합하는 방법론을 제안했다. 하지만 호모그래피를 수기로 작업해야 한다는 문제, 또 그 호모그래피를 통한 대응도 카메라 왜곡 문제를 해결하지 못해 또 다른 불확실성을 내포한다는 점에서 추가적인 논의가 필요하다고 보인다.

OC-SORT (CVPR 2023)

Observation-Centric SORT (이하 OC-SORT) [5]는 SORT에서 발생하는 문제들, 특히 객체의 가림(occlusion), 비선형 움직임(non-linear motion)으로 인한 문제들을 해결하기 위한 방법들을 제시한다.

OC-SORT는 SORT의 세 가지 주요 한계점을 지적한다. 첫 번째는 객체의 움직임을 선형으로 근사화하는 데 필요한, 높은 프레임 속도에서 야기되는 상태 추정의 잡음(noise) 민감성 증폭 문제이다. 이 속도 추정 잡음은 전이(transition) 프로세스에서 위치 추정에 누적되어 부정확한 추적으로 이끈다. 두 번째 한계는 칼만필터(Kalman filter) 업데이트 과정에서 관측치(observation)가 없는 경우, 시간이 지남에 따라 잡음이 누적되어 추적 정확도에 부정적인 영향을 끼친다는 점이다. 추적 대상을 추적하지 않다가 다시 추적하는 재정합(re-association)과정 이후에도 추적에 실패하는 등 추적 정확도에 영향을 끼친다. 세 번째 한계는 선형 동작의 가정과 오차 누적으로 인해 불확실성을 내재한 추정치에 의존하는 추정 중심 접근법(estimation-centric approach)을 채택하고 있다는 점이다. 객체 검출기(object detector)가 제공하는 비교적 낮은 분산(variance)을 가진 관측치를 활용하는 대신, 고정된 전이 함수에 의해 전파된 상태 추정치를 사용하여 성능을 제한하고 정확도에 부정적 영향을 끼칠 수 있다.

OC-SORT는 이러한 한계를 극복하기 위해 세 가지 핵심 구성요소를 도입한다. 첫 번째, Observation-centric Re-Update (ORU)는 가림(occlusion) 등으로 일시적으로 추적이 끊긴 객체가 다시 정합되었을 때, 그동안 누적된 오류를 줄이기 위해 제안되었다. 이 방법은 추적이 재활성화(reactivation)된 후, 해당 객체를 잃어버렸던 기간의 시작과 끝에 해당하는 관측치를 참조하여 생성된 가상 궤적에 기반하여 다시 업데이트(re-update)하는 것이다. 두 번째, Observation-Centric Momentum (OCM)은 선형 동작 가정의 한계를 극복하고, 추적 과정에서 발생하는 잡음을 완화하기 위한 방법이다. 추적 대상의 움직임 방향에 대한 일관성(consistency)을 기존 추적 링크와 지난 관측과 새로운 관측 사이의 링크가 이루는 각도의 쌍(pair)

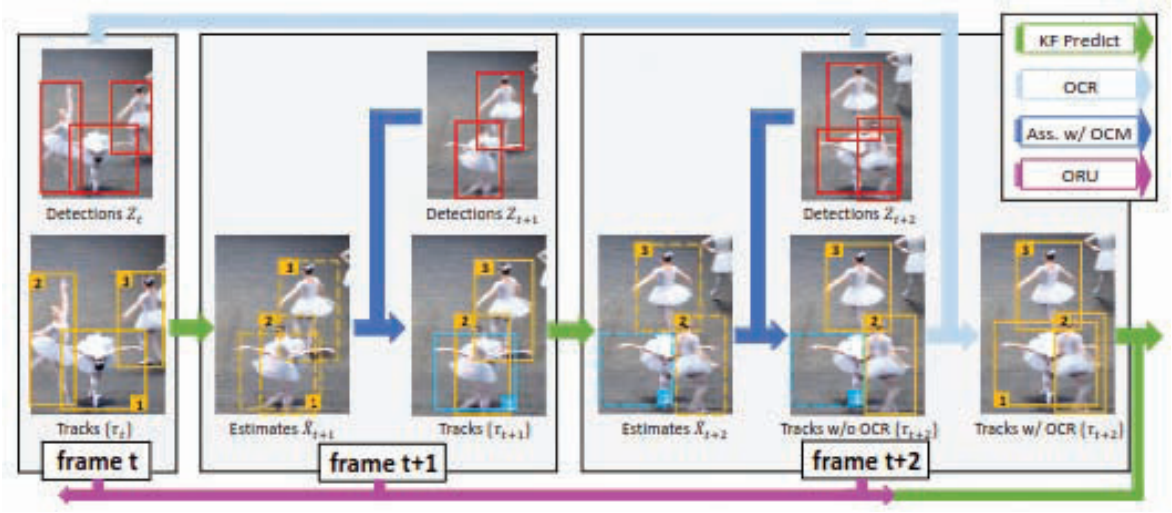

그림 3. OC-SORT의 동작[5].

들로 정의한다. 이를 기반으로 비용 행렬을 계산함으로써, 기존 추적 경로와 새로운 관측 사이의 방향 일관성을 정합 비용 계산에 사용한다. 마지막으로, Observation-Centric Recovery (OCR)은 객체가 일시적으로 멈추거나, 짧은 시간 동안 가려지는 상황에서 추적에 도움을 주는 휴리스틱 방법이다. 정합에 실패한 이후, 추적의 마지막 관측치와 매칭되지 않은 관측치 사이에서 두 번째 정합을 시도하는 방식이다.

2.3. 형태(Appearance) 기반 MOT 접근법

DeepSORT (ICIP 2017)

DeepSORT [6]는 기존 SORT의 칼만필터 예측이 상태 예측 불확실성(state estimation uncertainty)이 낮을 때에만 정확하다는 점을 지적한다. 이는 등속 운동 모델이 가정하는 상황에 반하는 경우, 특히 물체 가림(occlusion)에 의한 취약성을 시사한다. 또한 움직임 기반 메트릭만 사용하는 경우, 정합에 대한 장기적 관점이 부족하다는 점, 카메라의 흔들림과 같은 이미지 평면(image plane)상의 갑작스러운 물체 변위에 대해 대처 능력이 없다는 점을 지적한다. DeepSORT는 이를 해결하고자 움직임 기반과 외형 기반을 동시 사용하여 문제를 해결한다. 외형 지표를 설계 하기 위해 먼저 CNN 기반 외형 표현자(appearance descriptor)를 도입한다. 객체 바운딩 박스를 입력으로 받아 외형 표현자를 출력으로 내는 CNN 네트워크를 학습한다. 이때 외형 표현자를 축적하는 갤러리라는 데이터를 도입한다. 현재 프레임으로부터 100프레임 이전 동안 한 트랙에 대한 외형 표현자를 갤러리에 저장한다. 이 갤러리 내에서 코사인 유사도가 가장 큰 값을 가지고 비교하여 트랙과 현재 인식한 객체의 정합에 필요한 metric을 측정한다. DeepSORT는 외형 기반 정보를 이용하여 장기간 가려진 객체에 대해 향상된 추적 성능을 보여준다. 외형 정보를 사용한다는 측면에서 물체 재식별(re-identification: 인식된 물체에 대해서 이전에 인식한 같은 물체를 제공(retrieval)하는 동작 및 연구 분야를 일컫는다; Re-ID)과 밀접한 관계가 있다. 더욱이 다중 카메라에 대한 물체 추적으로 확장될 경우, 다른 두 카메라 간 정합을 시도할 때 움직임(motion)은 사용할 수 없다는 제약이 생기기 때문에 그러한 점에서 이점을 가진다. 하지만 외형 표현자 방식은 그대로 Re-ID에서의 한계점을 가지게 되는데, 이는 외형 형식의 변형(deformation), 물체의 가림 등에 의한 성능 저하이다.

Hybrid-SORT (AAAI 2024)

Hybrid-SORT [8]는 기존의 움직임 및 외형 기반 기술

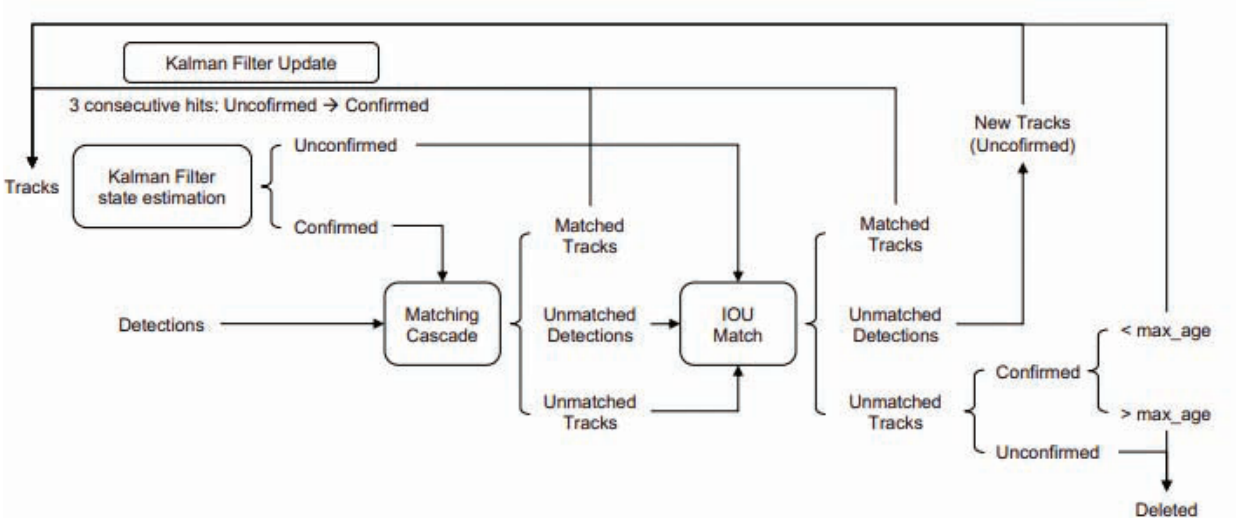

그림 4. DeepSORT의 동작[7].

들의 가려짐 및 밀집(clustering)상황에서 성능 저하 문제를 해결하기 위한 방법을 제시한다. 기존의 공간 및 외형(spatial and appearance)의 강한 단서(strong cues)와 함께 tracklet의 신뢰 상태(confidence state), 높이 상태(height state), 속도 방향(velocity direction)과 같은 약한 단서(weak cues)를 통합하여, 앞서있는 객체(foreground object)에 의한 문제를 완화한다.

Tracklet Confidence Modeling (TCM)은 tracklet의 신뢰도와 속도 요소를 칼만필터 상태에 통합하고, 선형 예측(linear prediction)을 결합한 방식이다. 이를 통해 필터의 지연 문제(delay)를 해소하고, 앞서 있는 객체와 뒤에 있는 객체(background object) 사이의 관계를 명시적으로 제공한다. Height Modulated IoU (HMIoU)는 바운딩 박스의 높이(height) 정보를 활용해 깊이 정보를 고려함으로써, 다양한 자세(pose)와 밀집된 상황에서 추적 성능을 강화한다. 이때, 너비(width) 정보는 사람의 자세 변화나 팔다리의 움직임에 따라 불규칙적으로 변하여 칼만필터에 활용하기 어렵기 때문에 높이 정보를 활용한다. Robust Object

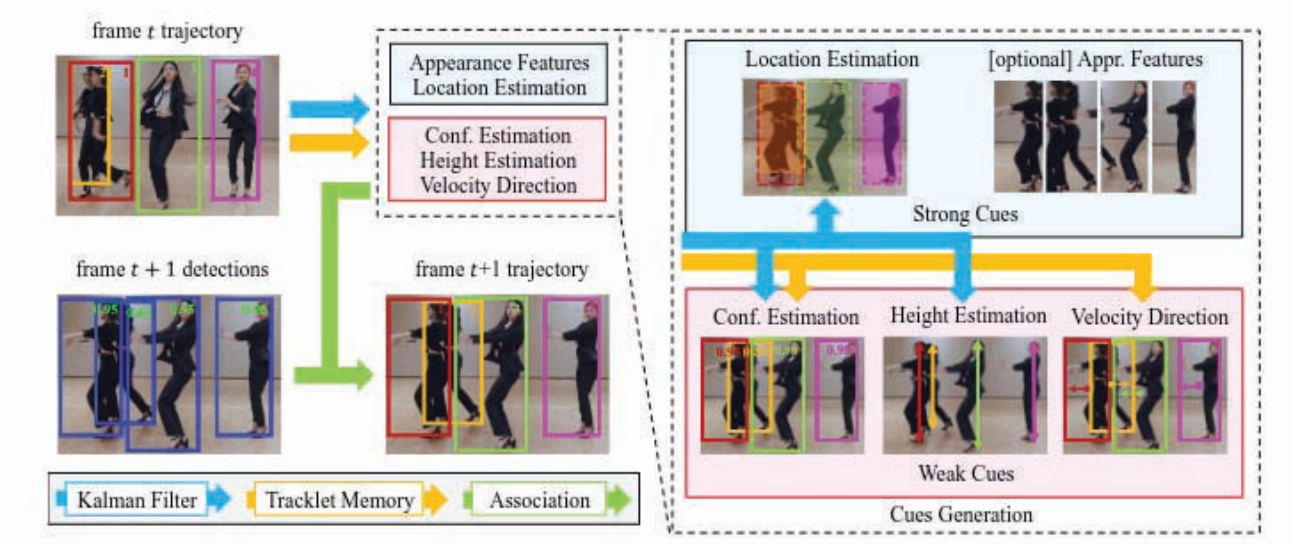

그림 5. Hybrid-SORT의 동작[8].

Centric Momentum (ROCM)은 OC-SORT에서 사용되는 중앙 좌표를 기반으로 한 속도 방향 계산 방식을 개선하여, 바운딩 박스의 4개의 코너의 속도 방향을 계산하고 평균을 냄으로써 정합 과정에서 잡음의 영향을 줄였다. 외형 모델링은 BoT-SORT의 접근 방식을 따라, 객체를 감지한 후 패치로 나누어 Re-ID 모델에 공급하는 방식을 채택한다. Exponential Moving Average (EMA)를 이용해 tracklet의 외형 정보를 모델링하고, 코사인 거리를 이용해 tracklet과 감지된 외형 특징(detection appearance features) 사이에 비용을 계산한다.

StrongSORT (T-MM 2023)

StrongSORT [9]는 DeepSORT의 한계를 보완하고자 세부 모듈을 바꾼 알고리즘이다. 물체 검출기로서 Faster R-CNN 대신 YOLOX-X를 사용하였고, 외형 표현자를 학습하기 위해 기존의 CNN 대신 BoT라는 물체 재식별(re-identification) 모델을 도입하였다. 또한 DeepSORT에서는 gallary라는 저장 방식을 이용하여 장기적 정보를 획득하고자 했는데, 이는 객체 검출 잡음에 취약하다는 단점을 가지고 있다. 이를 해결하기 위해 외형 상태(appearance state)에 대한 Exponential Moving Avarage (EMA)를 적용한다. 또한 카메라의 움직임을 모델링하여 시스템에 적용하기 위해 Enhanced Correlation Coefficient Maximization (ECC)을 도입하는데, 이는 래핑 파라미터(wrapping parameter)를 최소화하는 래핑 함수(warping function)를 찾

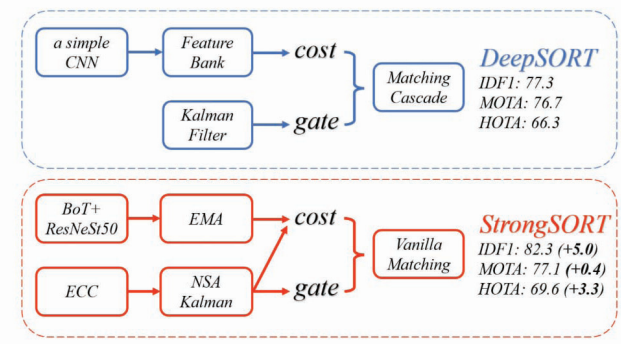

그림 6. StrongSORT의 동작[9].

은 후 사진 보정을 진행하는 방식이다. 이는 카메라 움직임에 대한 잡음을 보완하는 효과가 있다. StrongSORT는 EMA를 통해 기존의 갤러리 방식을 효율적으로 대체한다. 기존 방식은 현재 프레임으로부터 100프레임 동안의 외형 정보를 저장해야 했으나, StrongSORT에서는 moving average 방식을 적용해 저장공간이 필요하지 않으면서 업데이트하는 방식으로 효율성을 증대한다.

2.4. 학습(Learning) 기반 MOT 접근법

앞서 언급한 방식은 MOT의 주요 방법론에서 파생된 연구를 소개하였다. 하지만, 이 방식들에는 여전히 제한사항이 존재한다. 외형 기반 방법은 많은 객체가 겹치는 복잡한 시나리오에서 어려움을 겪는다. 또한, 움직임 기반 방법에서는 비선형 3D 동작을 2D 이미지 도메인으로 변환하는 것이 어려운 문제로 남아있다. 추가로, 본문에 소개되지 않은 회귀 기반 추적(tracking-by-regression), 세그멘테이션 기반 추적 (tracking-by-segmentation), 그리고 그래프 이론을 적용한 검출 기반 추적 방법들은 각각 학습 방식의 복잡성, MOT에 특화된 학습 데이터의 부족, 과도한 최적화로 인한 계산 비용 증가 및 온라인 추적 시 적용 한계로 인해 문제가 있다. 이러한 문제들을 해결하기 위해, 학습 기반(learning based) 이론을 활용하는 다양한 학습 방법이 MOT 문제에 적용되고 있다.

TrackFormer (CVPR 2022)

TrackFormer [10]는 추적 작업을 단일 집합 예측 문제로 처리하고, Transformer 알고리즘을 통해 문제를 해결하려고 한다. 이 방법은 연관 단계(association)를 포함하여 트랙 초기화, 동일성 유지, 시공간 궤적을 추적하는 데에서도 어텐션 메커니즘을 적용한다. 이 접근법은 특징 수준(feature-level)의 어텐션에만 초점을 두어, 기존 방식인 그래프 최적화와 외형/움직임 모델이 불필요하다. TrackFormer의 아키텍처는 이미지 특징을 추출하기 위한 CNN, 이 특징들을 인코딩하기 위한 트랜스포머 인코더, 그리고 셀프 어텐션 및 인코더-디코더 어텐션을 사용하

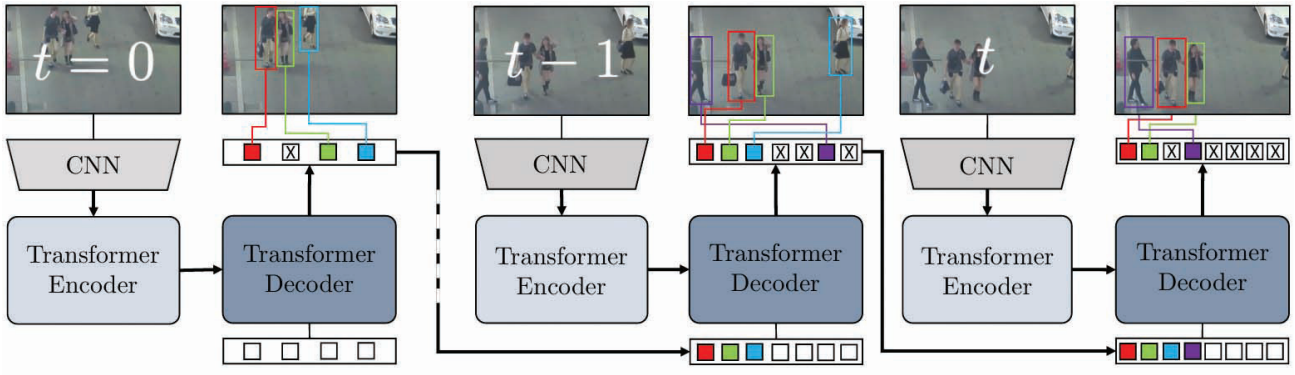

그림 7. TrackFormer의 동작[10].

여 경계 상자 및 클래스 정보를 포함하는 출력 임베딩을 생성하는 트랜스포머 디코더로 구성된다.

TransTrack (arXiv 2020)

기존 MOT 기술에서 중요한 딜레마는 객체 검출과 Re-ID이 독립적으로 동작하여 서로에게 이익을 주지 못한다는 것이다. MOT 문제를 해결하려면, 검출와 객체 정합 간의 지식(knowledge)을 공유하기 위한 joint-detection-and-tracking 프레임워크가 필요하다. TransTrack [11]는 Transformer를 기반으로 한 MOT 프레임워크를 구축한다는 점에서 기본적으로 TrackFormer와 매우 유사한 개념을 가지고 있다. 그러나 TransTrack은 질의(query)를 정의하는 방식에서 차이점을 가진다. 공통 객체 감지 결과를 제공하기 위한 객체 질의(object queries), 다음 프레임에서 관련된 객체를 찾기 위한 추적 질의(track queries)이다. 이 두 가지 질의 세트는 두 개의 병렬 디코더를 통해 전달된다. 그 후 'detection boxes'와 'tracking boxes'라고 불리는 두 개의 바운딩 박스 세트를 병렬로 생성한다. 마지막으로, TransTrack은 IoU를 비용으로 헝가리안 알고리즘을 사용하는 구조로 되어있다.

MOTR (ECCV 2022)

Multiple-Object Tracking with Transformer (MOTR)[12]은 DETR을 기반으로 한 end-to-end 프레임워크 추적 알고리즘이다. 그렇기 때문에 Non-Maximum Suppression (NMS) 추적 또는 IoU matching과 같은 후처리 단계를 도

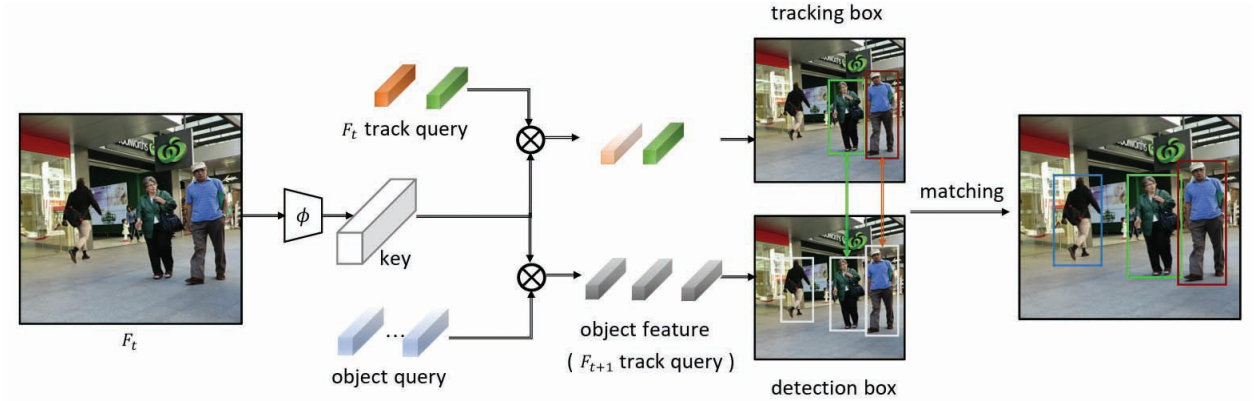

그림 8. TransTrack의 동작[11].

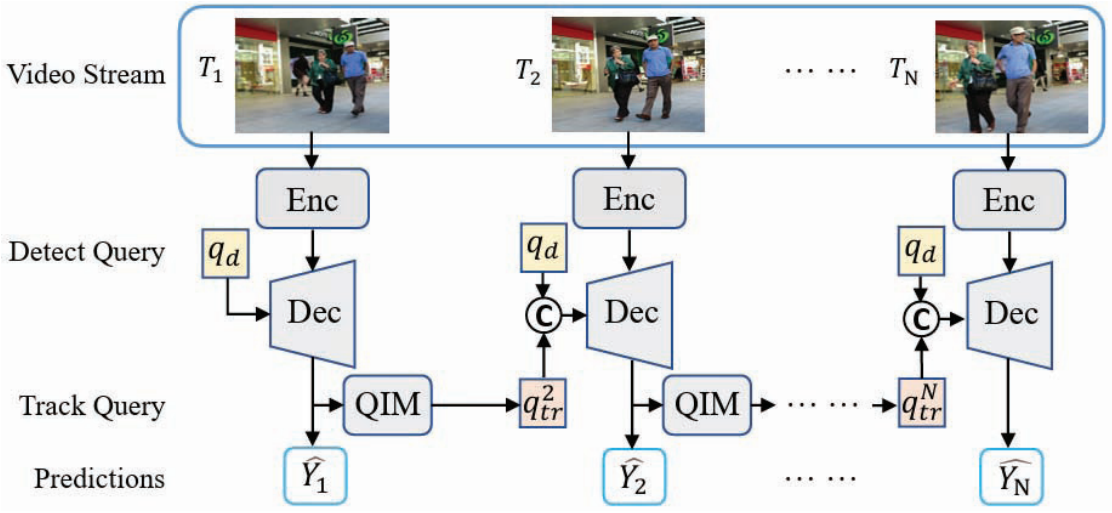

그림 9. MOTR의 동작[12].

입하지 않고 단일 네트워크로 모든 동작을 수행하여 기존 학습 기반 MOT보다 더 개선된 구조를 제시한다. TransTrack은 여러 개의 독립적인 짧은 tracklets의 조합으로 전체 트랙을 모델링한다. TrackFormer는 인접한 두 프레임 내에서 학습하므로 시간적 학습이 상대적으로 약하다. 그렇기 때문에 이전 시간 정보를 활용할 수 없어 TrackFormer는 NMS 및 Re-ID와 같은 경험적 기법을 사용하여 중복 트랙을 필터링한다. 반면, MOTR은 Collective Average Loss (CAL) 및 Temporal Aggregation Network (TAN) 같은 세부 네트워크를 통해 전체 시간에 대한 정보를 활용하여 추적을 수행함으로써, 추가적인 후처리 단계 없이도 효율적인 추적 결과를 제공한다.

3. MOT 기술 응용 동향

MOT 기술은 물체를 단순히 감지하는 것을 넘어 시간에 따라 물체가 이동함에도 동일 물체로 구분하여 정합시켜준다. 따라서 MOT 기술은 감지된 물체를 공간적 움직임을 시간 축에 따라 분석하고 이해하고자 할 수 있어 다양한 현장에 응용되고 있다. 특히 1) 다수의 사람 또는 물체의 모니터링, 2) 이벤트나 이상 행동 또는 흐름 감지, 3) 도시 규모 광역 공간의 모니터링 등 사람이 수작업으로 직접 수행 어려운 작업을 자동화하는데 사용될 수 있다. 대표적 응용 분야는 공공장소 감시 및 모니터링 시스템, 스포츠 영상 분석, 야생동물 생태 분석, 자율주행 자동차/로봇/드론의 물체 및 장애물 감지 등이 있다. 본 기고문에서는 지능형 영상 감시 시스템과 자율주행의 공간인지의 두 측면에서의 최근 응용 예들을 소개한다.

3.1. 지능형 영상 감시 시스템

인공지능 및 사물인터넷(internet-of-things; 이하 IoT)의 발전은 MOT 기술을 포함한 모니터링 시스템 분야의 혁신적인 진화를 촉진하였다. 이러한 진보는 고도화된 지능형 영상 감시 시스템(intelligent video surveillance; 이하 IVS)의 구현을 가능하게 함으로써, 효율성과 안정성을 향상하여 다양한 분야에서 사회적 문제 해결에 중요한 역할을 하고 있다.

특히, 국방 분야에서는 출산율의 저하로 병력 감소에 대응하기 위해 과학화 경계 시스템의 도입이 활발하다. CCTV를 활용한 외곽 경계 시스템은 MOT 기술을 활용하여 사람, 동물 등을 식별하고 검출하며, 추적을 통해 감시 및 경보를 발령한다. 특히 산악 지형인 우리나라의 지형에 적합하며, 기온 변화 및 동식물 등 다양한 외부 환경 조건에도 강한 신뢰성을 보여준다[13].

더불어, 제7회 국방과학기술대제전에서는 MOT를 기반으로 한 감시정찰 솔루션이 선보였다. 이 솔루션은 영상 및 카메라 기술을 활용하여 다중 객체의 정보, 공간 측위 및 GPS 추적을 가능하게 하며, 영상 내 특정 지점의 위치 정보를 기반으로 위도와 경도를 계산한다. 이와 같은 기술은 드론과 같은 무인이동체에 적용될 수 있으며, 국방뿐만 아니라 건설 및 제조 현장에서 중장비 충돌 방지 예방과 스마트 시티 관제 영역에서도 활용할 수 있음을 보여준다[14].

대구 달서구청은 2023년 11월부터 쓰레기 무단 투기 자동 추적 시스템을 도입하여, CCTV를 통해 쓰레기 무단 투기 행위를 실시간으로 감지하고, 통합관제센터와 연계하여 이를 관리한다. 이 시스템은 다중 카메라를 기반으로 동일 인물의 동선을 분석하고 추적하여, 효율적인 감시 및 관리를 실현한다[15].

선박 특화 지능형 CCTV 감시 시스템은 선박을 자동으로 인식하고, 입출항 여부 및 선박의 이동 상황을 실시간으로 추적한다. 이 시스템은 기상 상황의 변화에도 불구하고 체계적인 관리를 가능하게 하며, 항로상 위험 구역 내 선박이나 사람 감지 시 상황센터에 알람을 울려 신속한 조치를 취할 수 있는 재난 예방 효과를 제공한다[16].

한편, ETRI에서 개발한 다중 카메라 교통 객체 추적 기술은 관심 객체의 자동 검출 및 인식, 차량 추적 및 재식별, 대기열 추정 및 교통량 추산 등을 통해 주차 관리 및 차량 출입 통제 시스템에 활용될 수 있다. 이 기술은 또한 구간별 교통량을 추산하여 교통정보 처리 시스템에도 응용될 수 있다[17].

감시 시스템의 하드웨어 기술 발전도 주목할 만하다. PTZ (pan-tilt-zoom) 카메라와 적외선(infra-red; IR) 카메라의 결합을 통해, 낮과 밤에 구분 없이 360도 모니터링이 가능한 시스템이 개발되었다. 이러한 기술의 발전은 IVS의 효율성 및 정확성을 크게 향상시키고 있다[18].

3.2. 자율주행의 공간인지

Advanced Driver Assistance Systems (ADAS)는 운전 및 주차 기능을 위해 차량에 적용된 총체적 기술을 일컫는다. ADAS는 카메라, 라이다(LiDAR), 레이더(Radar) 등을 이용한 도로 상황 인식(perception)을 주 기능으로 포함하게 된다. 도로 상황 인식은 특히 차량 및 보행자에 대해서 어디에 있는지(object localization), 어떤 것인지(object classification), 각 객체 간 궤적을 인식 및 예상(object tracking)할 수 있어야 한다.

테슬라(Tesla)의 AutoPilot은 기술 오직 카메라 정보만을 이용해 독자적인 인식 알고리즘을 적용한다. 테슬라는 특히 인식, 결정 및 학습에 대한 전체 과정을 end-to-end 딥러닝 방식을 적용하는 특징을 가지고 있다. HydraNet은 여러 카메라의 이미지를 결합하는 다중 카메라 융합(multi-cam fusion)을 통해 '슈퍼 이미지'를 생성하고, 이전의 '슈퍼 이미지'와 결합한다. HydraNet이 수행하는 다양한 테스크(task)는 공유된 백본(backbone) 신경망 뒤에 여러 헤드로 분리되어 수행되고, AutoPilot은 객체 추적 기술을 포함하여 도로 내 자동차의 다양한 테스크를 통합적으로 수행할 수 있는 모델로 주목을 받고 있다[19].

차량의 자율주행 분야에는 특히 LiDAR를 이용해 물체를 추적하는 기술이 주목 받고 있다. 이는 앞서 소개한 카메라 기반 추적과 비교적 다른 방식을 적용하게 된다. 먼저 LiDAR를 통해 물체를 인식하기 위해서 크게 두 가지 경우의 방식이 사용된다. 첫 번째는 클러스터링 알고리즘을 적용하는 경우이다. 클러스터링 알고리즘에는 대표적으로 K-means과 DBSCAN이 있는데, 사전에 클러스터

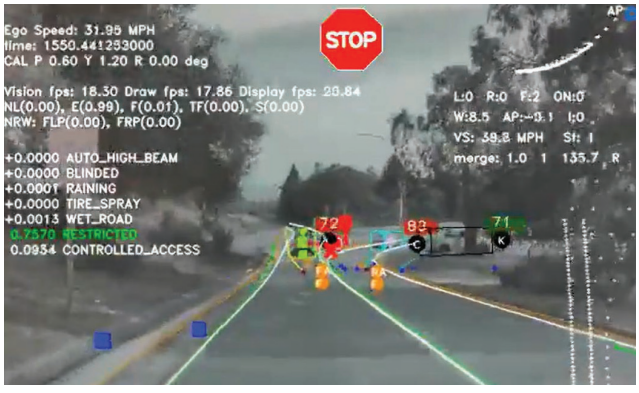

그림 10. Tesla AutoPilot의 공간 인지의 예[19]

개수를 정의하지 않아도 되는 DBSCAN이 주로 적용된다. 혹은 점군 기반 3차원 객체 인식(point cloud-based 3D object detection)을 통하여 한 프레임 내 정합하고자 하는 후보군을 생성한다. 이를 칼만필터를 이용하여 위치와 속도를 예측한다. 일반적으로 확장칼만필터(extended Kalman filter)를 적용한다.

도시바(TOSHIBA)는 세계 최초로 99.9%의 추적 정확도를 가진 LiDAR 기반 물체추적 기술을 개발했다. 2D/3D 퓨전 AI, 비/안개 제거 알고리즘, 가변 측정 범위 기술 등을 포함하여 다양한 시나리오에 대처할 수 있고, 물리적 공간과 객체의 실시간, 정확한 모델링을 가능하게 한다. 도시바는 2025년 상용화를 목표로 하고 있다[20].

4. 맺음말

본 기고문에서는 영상 감시, 자율 주행, 로보틱스, 스포츠 분석 등 다양한 분야에서 활용되는 다중객체추적(MOT) 기술을 리뷰하였다. 우선 MOT 문제에 대해 수학적으로 정의하고, 물체의 정합에 중요한 정보가 되는 운동모델과 형상모델에 대해 소개하였다. 이러한 개념을 바탕으로 다양한 MOT 연구들을 물체의 움직임(motion) 정보를 보다 적극적으로 활용하는 기술과 외형(appearance) 정보를 보다 심도있게 활용하는 기술, 그리고 학습(learning) 기반으로 이를 대체하는 기술로 나누어 살펴보았다. MOT 기술은 다양한 분야에 널리 적용되고 있는데, 본 기고문에서는 지능형 영상 감시와 자율주행의 공간인지 부분에서의 최근 응용 예들을 소개하였다.

MOT 기술은 다양한 응용의 핵심 기술로 여전히 많은 연구가 이뤄지고 있다. 주요 기술적 난제로는 다중 카메라 물체추적으로 확장, 물체의 복잡한 가려짐이나 비선형적인 움직임이 있고, 그 외에도 개인 정보 보호, 데이터의 오용과 남용 가능성과 같은 윤리적 및 사회적 문제도 중요한 이슈이다. 본 기고문을 통해 MOT 분야의 놀라운 연구를 모두 다룰 수는 없었지만, 독자들로 하여금 다중 객체 추적의 중요성과 가능성에 대한 관심을 불러일으킬 수 있기를 바란다.

감사의 글

본 기고는 과학기술정보통신부 및 한국연구재단의 'BRIDGE융합연구개발'사업의 지원으로 작성되었습니다. (과제명: AI기반 3차원 곡면에서의 위치 인식 및 이동 경로 생성 기술, 과제번호: 2021M3C1C3096810)

REFERENCES

[1] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, "Multiple object tracking: A literature review", *Artificial Intelligence,* vol. 293, 2021.

[2] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking", in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2016.

[3] K. Yi, K. Luo, X. Luo, J. Huang, H. Wu, R. Hu, and W. Hao, "UCMCTrack: Multi-object tracking with uniform camera motion compensation", in *Proceedings of AAAI Conference on Artificial Intelligence,* vol. 38, no. 7, 2024.

[4] R. Pereira, G. Carvalho, L. Garrote, and U. Nunes, "Sort and Deep-SORT Based Multi-Object Tracking for Mobile Robotics: Evaluation with New Data Association Metrics", *Applied Sciences,* vol. 12, no. 1, 2022.

[5] J. Cao, J. Pang, X. Weng, R. Khirodkar, and K. Kitani, "Observation-centric SORT: Rethinking SORT for robust multi-object tracking", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 2023.

[6] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric." in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649. IEEE, 2017.

[7] B. Veeramani, J. W. Raymond, and P. Chanda, "DeepSort: Deep convolutional networks for sorting haploid maize seeds," *BMC Bioinformatics*, vol. 19, 2018.

[8] M. Yang, G. Han, B. Yan, W. Zhang, J. Qi, H. Lu, and D. Wang, "Hybrid-Sort: Weak Cues Matter for Online Multi-Object Tracking", In *Proceedings of AAAI Conference on Artificial Intelligence,* vol. 38, no. 7, 2024.

[9] Y. Du, Z. Zhao, Y. Song, Y. Zhao, F. Su, T. Gong, and H. Meng, "StrongSort: Make deepsort great again", *IEEE Transactions on Multimedia,* vol. 25, 2023.

[10] T. Meinhardt, A. Kirillov, L. Leal-Taixe, and C. Feichtenhofer, "Trackformer: Multi-object tracking with transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 2022.

[11] P. Sun, J. Cao, Y. Jiang, R. Zhang, E. Xie, Z. Yuan, C. Wang, and P. Luo, "TransTrack: Multiple object tracking with transformer", *arXiv preprint (arXiv:2012.15460)*, 2020.

[12] F. Zeng, B. Dong, Y. Zhang, T. Wang, X. Zhang, and Y. Wei, "MOTR: End-to-end multiple-object tracking with transformer," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2022.

[13] 엄호식,"과학화 경계 시스템의 필수 요소 '외곽감시 시스템' 시장 분석", 보안뉴스, 2024.02.02, https://m.boannews.com/html/detail.html?idx=126193

[14] 김성수, "CCTV 등 영상통해 객체정보 식별하고 추적하는 AI 솔루션", 산업일보, 2023.08.05, https://kidd.co.kr/news/233583

[15] 김병철, "대구시 / 쓰레기 무단투기 "이동경로 자동 추적 시스템 도입", 환경신문, 2023.11.21, https://www.fksm.co.kr/m/view.php?idx=64744

[16] 박미영, "KOMSA, 지능형 CCTV로 여객선 안전운항 모니터링 강화, 보안뉴스, 2023.05.27, https://m.boannews.com/html/detail.html?idx=118559

[17] 이은실, "[대한민국 ICT융합엑스포] ETRI(한국전자통신연구원) 대경권 연구센터, '다중 카메라 교통 객체 추적 기술' 소개", 에이빙 뉴스, 2019.11.01, https://kr.aving.net/news/articleViewAmp.html?idxno=1539360

[18] 유희석, "한화비전, AI PTZ Plus 라인업 출시... 객체 '자동추적'", 뉴시스, 2023.06.16, https://mobile.newsis.com/view_amp.html?ar_id=NISX20230616_0002341814

[19] "This is what Tesla Autopilot sees", WhichCar, https://www.whichcar.com.au/car-advice/this-is-what-tesla-autopilot-sees

[20] "Toshiba's Image Recognition Technologies for Human-Robot Collaboration", Toshiba Global, https://www.global.toshiba/ww/technology/corporate/rdc/rd/topics/23/2309-02.html

서 찬 호

2017년–2023년 서울과학기술대학교 전기정보공학과 (학사)
2023년–현 재 서울과학기술대학교 컴퓨터공학과 (석사)

Nguyen Cong Quy

2016년–2021년 다낭과학기술대학교 전자통신부 (학사)
2023년–현 재 서울과학기술대학교 컴퓨터공학과 (석사)

허 동 욱

2014년–2018년 명지전문대학 기계과 (전문학사)
2019년–2022년 서울과학기술대학교 기계시스템 디자인공학과 (학사)
2024년–현 재 서울과학기술대학교 컴퓨터공학과 (석사)

최 성 록

2001년–2006년 서울대학교 기계항공공학과 (공학사)
2006년–2008년 KAIST 로봇공학학제전공 (공학석사)
2014년–2019년 KAIST 로봇공학학제전공 (공학박사)
2008년–2020년 ETRI 지능로보틱스연구본부 (선임연구원)
2021년–현 재 서울과학기술대학교 컴퓨터공학과 (조교수)