

Comparison of Tracking-by-detection Algorithm on Aerial Battle Scenario based On-Device

Du-Hwan Hur*, Dae-Hyeon Park*, Deok-Woong Kim**, Junhyeong Bak***, Ingu Park****,
Ki-Nam Lee****, Seung-Hwan Bae*****

*Ph.D. candidate, Vision & Learning Lab, Inha University, Incheon, Korea

**M. S. candidate, Vision & Learning Lab, Inha University, Incheon, Korea

***Research Engineer, Missile System Seeker R&D, LIG Nex1 Co., Ltd., Yongin, Korea

****Chief Research Engineer, Missile System Seeker R&D, LIG Nex1 Co., Ltd., Yongin, Korea

*****Associate Professor, Vision & Learning Lab, Dept. of Electrical and Computer Engineering, Inha University,
Incheon, Korea

[Abstract]

In this paper, we present a guideline for multi-object tracking in aerial battle scenarios, where tracking is required from low-resolution infrared imagery and low-quality detection results. Albeit many MOT algorithms are compared on high-performance resource (e.g., high-performed devices, high-resolution inputs and powerful detectors), the analysis on the resource-constrained scenario is limited. To resolve this gap, we analyze how such resource-constrained conditions affect tracking performance and computational cost. We demonstrate that powerful detections leads to improved Average IoU and reduced false positives (FP) in the tracker. We also compare recent tracking-by-detection under both high and low resolution conditions. Based on these results, we provide tracking algorithm that clarify whether to reduce false positives (FP) or false negatives (FN). As a result, we provide the guideline in aerial battle scenarios, demonstrating that tracking accuracy depends on detection confidence and that the association algorithms focus on regulating both FP and FN.

▶ **Key words:** Tracking-by-detection, Multi-object tracking, Detection confidence, On-device tracking, Aerial battle scenario

-
- First Author: Du-Hwan Hur, Corresponding Author: Seung-Hwan Bae
 - *Du-Hwan Hur (gjenghks@inha.edu), Vision & Learning Lab, Inha University
 - *Dae-Hyeon Park (saintPalite2221@inha.edu), Vision & Learning Lab, Inha University
 - **Deok-Woong Kim (k5000plus@inha.edu), Vision & Learning Lab, Inha University
 - ***Junhyeong Bak (junhyeong.bak@lig.nex1.com), Missile System Seeker R&D, LIG Nex1 Co., Ltd.
 - ****Ingu Park (ingu.park2@lig.nex1.com), Missile System Seeker R&D, LIG Nex1 Co., Ltd.
 - ****Ki-Nam Lee (leekinam@lig.nex1.com), Missile System Seeker R&D, LIG Nex1 Co., Ltd.
 - *****Seung-Hwan Bae (shbae@inha.ac.kr), Vision & Learning Lab, Dept. of Electrical and Computer Engineering, Inha University
 - Received: 2025. 09. 04, Revised: 2025. 12. 11, Accepted: 2025. 12. 11.

[요 약]

본 논문에서는 저해상도 이미지 및 저수준 검출 결과로부터 추적이 요구되는 항공 전투 시나리오 환경에 대한 객체 추적 가이드라인을 제시한다. 기존 검출 기반 추적 알고리즘은 고성능 디바이스에서 평가되지만, 실제 전장 환경에서는 자원이 제한되기 때문에 고해상도 및 고수준 검출 결과를 활용하기 어렵다. 하지만 저수준 검출 결과만 사용할 수 있는 상황에 대해서는 분석이 미비하다. 본 논문에서는 검출 신뢰도에 따라 검출 신뢰도가 높을수록, 추적기가 Average IoU가 향상되고, FP (False Positive)가 감소되는 것을 입증한다. 그리고 고해상도/저해상도 환경에서 최신 검출 기반 추적 알고리즘 간의 추적 메트릭을 비교한다. 이를 통해 FP를 줄일지, FN (False Negative)을 줄일지에 대한 추적 전략을 제공한다. 결과적으로, 우리는 항공 전투 시나리오에서 추적 정확도는 검출 신뢰도에 의존한다는 것과, 연관 알고리즘은 FP와 FN을 조절하는 역할을 나타내는 추적 알고리즘 가이드라인을 제공한다.

▶ **주제어:** 검출 기반 추적, 다중 객체 추적, 검출 신뢰도, 온 디바이스 추적, 항공 전투 시나리오

I. Introduction

다중 객체 추적 알고리즘은 연속된 프레임 간의 검출 결과를 연관하여 ID를 할당하는 검출 기반 추적 (Tracking-by-detection) 패러다임을 준용한다 [1-7]. 따라서 다중 객체 추적 알고리즘의 추적 정확도는 검출 정확도에 매우 의존적이다. 많은 검출 기반 추적 연구는 고성능 온 서버(On-Server) 기반 환경에서, 고해상도 이미지와, 높은 정확도를 가진 고수준 객체 검출기(Object Detector)[8, 9]로 평가된다. 하지만 실제 전장 시나리오에서는 위와 같은 환경을 구축하기 어렵다. 첫째, 고전력을 사용하기 어렵기 때문에 온 서버에 준하는 환경 대신 온 디바이스(On-Device) 환경에서 모델이 실행된다. 둘째, 실제 환경에서는 센서나 비용 문제로 인해 고해상도의 이미지를 제공하기가 어려울 뿐만 아니라, 객체에 대해 충분히 분별할 수 있는 정제된 이미지를 제공하기 어렵다. 셋째, 온 디바이스 환경에서는 고성능 검출기를 사용하기 어렵다. 이런 전력 및 비용 문제로 인한 현실적 제약으로 인해 낮은 정확도의 검출기와 그 결과를 사용한다.

하지만 고성능 연산 자원, 고해상도 이미지, 그리고 고수준의 검출기를 사용하는 기존 연구와 달리, 제한된 연산 자원, 저해상도 이미지와 그리고 저신뢰 검출 결과에 대해 종합적으로 고려하고 개선하는 다중 객체 추적 연구는 물론, 다중 객체 추적 알고리즘의 정확도 변화를 검증하거나 자세한 분석을 진행한 연구는 여전히 미비하다. 일반적으로 객체 추적 알고리즘은 고수준 검출기를 전제로 설계되었기 때문에, 저신뢰 검출이 지배적인 실제 온 디바이스 기반 항공 전투 시나리오에서 어떤 연관 알고리즘이 적합

한지, 그리고 추적 메트릭과 속도가 어떻게 나타나는지에 대한 체계적 분석이 부족하다. 항공 전투 시나리오에서는 저성능 검출기의 낮은 객체 분별력으로 인한 폐색 (Occlusion), 조도 변화(Illumination change), 모션 블러(Motion blur), 노이즈(Noise), 객체 스케일 변화 (Object scale variance) 문제에 직면하기 때문에 검출 기반 추적이 실패할 수 있다. 예를 들어, 본 논문에서 다루는 항공 전투 시나리오 데이터셋은 256×256 저해상도에 열화상 기반 이미지이므로 객체의 외형 정보는 거의 소멸되며, 작은 객체가 배경과 구별되지 않는 경우도 빈번하다. 이것은 다음과 같은 MOT 연구와 독립적인 문제를 야기한다.

- (i) 외형 연관 모델(Re-ID embedding)은 저해상도/열화상 이미지이므로 사용하기 어렵다.
- (ii) 모양 연관 모델(IoU)은 박스 간 분별력이 낮다
- (iii) 모션 연관 모델(Kalman filter)은 폐색/비선형 움직임에 의해 쉽게 오염된다.

결과적으로, 온-디바이스 환경에서 저해상도 이미지로 저성능 검출기를 사용하는 것은 기존 MOT 연구와 다른 결과를 초래할 수 있음을 시사한다.

본 논문에서는 실제 항공 전투 시나리오 환경에서 저해상도 이미지와 저수준 검출 결과, 그리고 제한된 연산 자원에 기반한 다중 객체 추적 알고리즘의 결과를 분석한다. 본 연구의 주요 기여점은 크게 다음과 같이 정리된다.

- (1). 온 디바이스 환경에서 열화상 기반 저해상도 이미지로 구성된 항공 전투 시나리오 데이터셋에서, 검출 신뢰

도 threshold를 변화시켜 검출 퀄리티에 따라 추적 성능과 연산 복잡도에 미치는 정량적인 근거를 도출한다.

(2). 최신 검출 기반 연관 알고리즘(ByteTrack, OC-SORT, Hybrid-SORT)을 고해상도 환경(MOT17 [10], DanceTrack [11])과 저해상도 환경(항공 전투 시나리오 데이터셋)에서 추적 성능을 비교 분석한다.

(3). (1)과 (2)를 종합하여, 온 디바이스 환경, 저해상도 이미지, 저성능 검출기에서 선택해야 할 최적 연관 전략을 위한 가이드라인을 제시한다.

II. Preliminaries

1. Tracking-by-detection Paradigm

다중 객체 추적은 검출과 연관을 동시에 처리하는 종단간(End-to-End) 추적과, 검출 결과에 기반해 연관을 수행하는 검출 기반 추적(Tracking-by-Detection)으로 구분된다. 종단간 추적은 검출과 연관을 모두 GPU에서 처리할 수 있지만, 높은 연산량과 복잡한 구조로 인해 저전력 온 디바이스 환경에 적용하기 어렵다. 반면 검출 기반 추적 방식은 연관 알고리즘을 CPU에서 수행하므로, 고속 검출기 적용이 용이해 저전력 환경에서 더 적합하다.

일반적으로 검출 결과를 연속된 프레임에서 연관하는 검출 기반 추적 알고리즘을 채택한다 [12, 13]. 검출 기반 추적은 매 프레임마다 검출 결과를 생성하고, 생성된 검출 결과를 각 프레임마다 동일한지 유사도 모델을 통해 계산한다. 이를 통해 동일한 객체 여부를 판별하고, 추적 ID를 할당한다. 만약 새로운 추적 대상인 경우, 새로운 추적 ID를 할당한다. 동일 객체 여부를 판별하기 위한 유사도 모델은 일반적으로 모양 모델, 모션 모델, 그리고 외형 모델을 채택한다. 모양 모델은 추적 객체와 검출 결과 간의 IoU가 높을수록 높은 유사도를 나타낸다. 모양 모델은 칼만 필터(Kalman filter)를 사용해 객체의 다음 프레임의 예상 위치를 선형적으로 예측하고 갱신한다 [5]. 외형 모델 [6]은 객체의 외형 특징맵 간의 유사도를 비교한다. 하지만 이 접근 방법은 검출 결과의 퀄리티에 크게 의존한다. 검출 결과의 퀄리티가 낮을 경우, 추적 객체와 검출 결과 간의 모양 모델과 외형 모델의 유사도가 낮아지기 때문에 동일한 객체인지 판별하기 어렵다. 본 논문에서는 이 점에 주목하고, 검출 신뢰도에 따라 검출 기반 추적 알고리즘의 정확도/속도의 변화를 관찰하고 분석한다.

III. Preliminaries

본 논문에서는 실시간 검출기를 사용하기 위해 종단간 추적 알고리즘 대신 검출 기반 추적 패러다임을 채용한다. 특히, 낮은 해상도로 인해, 추적 객체의 외형 정보를 사용하기 어렵다. 본 장에서는 Re-ID를 사용하지 않은 검출 기반 추적 알고리즘인 ByteTrack, Hybrid-SORT, OC-SORT를 소개한다.

1. ByteTrack

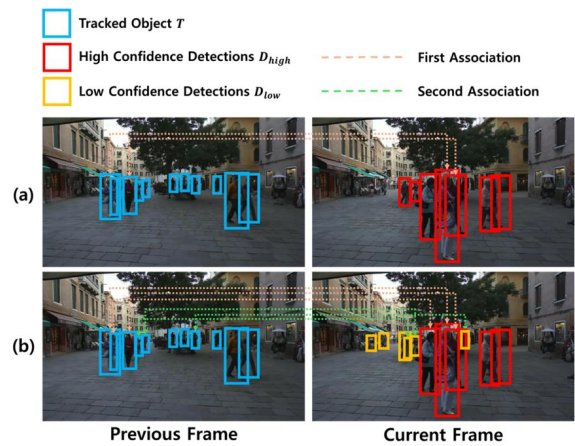


Fig. 1. The examples without ByteTrack (a) and with ByteTrack (b)

기존 검출 기반 추적 모델은 검출 결과 퀄리티에 따라 추적 퀄리티가 의존하기 때문에, 고수준의 검출 결과를 사용하여 연관한다[2, 3, 14]. 하지만 고수준의 검출기를 통해 객체를 검출하더라도 객체가 카메라로부터 멀어져 객체의 크기가 감소하거나 폐색으로 인해 객체 외형이 손상될 경우 객체의 신뢰도를 감소하여 객체 검출이 실패할 수 있다. 이로 인해 기존 추적 객체가 소실되어, 추적 객체의 프레임 간 위치의 연속된 기록(Trajectory)이 끊기거나, 기존 ID를 유지하지 못하고 새 ID가 부여되는 문제(ID 재할당)가 발생할 수 있음을 나타낸다. 이 문제를 해결하기 위해서는 낮은 신뢰도의 검출 결과를 모두 고려해야 함을 나타낸다.

ByteTrack [3]은 저수준의 검출 결과를 포함해 가능한 모든 검출 결과를 활용하기 위해 검출 결과를 2단계에 걸쳐 구분한 후에 연관하는 방식을 제안한다. 먼저, 검출 기로부터 검출 결과 집합 D 를 생성한다. 집합 D 는 n 개의 검출 결과 d 로 구성되며, 검출 결과 d 의 검출 신뢰도 (Detection confidence)를 $\theta(d)$ 라 한다. 이 때 검출 신뢰도는 분류 신뢰도(Classification score)와 같다. 고수준

임계값 θ_{high} 와 저수준 임계값 θ_{low} 는 검출 신뢰도 threshold이며, 각각 0.5, 0.1로 설정한다. $\theta(d) \geq \theta_{high}$ 을 만족하는 d 의 집합을 고수준 검출 결과 D_{high} 로 정의한다. $\theta_{high} > \theta(d) \geq \theta_{low}$ 을 만족하는 d 의 집합을 저수준 검출 결과 D_{low} 로 정의한다. 그 외의 검출 결과는 노이즈로 판단하여 제거한다. 1차 연관에서 고신뢰도 검출 결과 D_{high} 와 추적 객체 T 를 서로 연관한다. 이 때 연관되지 않은 검출 결과 D_{remain} 와 연관되지 않은 추적 결과 T_{remain} 를 2차 연관 때 활용한다. 2차 연관에서는 저수준의 검출 결과 D_{low} 와 1차 연관 되지 않은 검출 결과 D_{remain} 를 1차 연관되지 않은 추적 객체 T_{remain} 간의 연관을 수행한다. 즉, $D_{remain} + D_{low}$ 와 T_{remain} 간에 2차 연관을 수행한다. 이 때 각 연관 알고리즘은 IoU 기반 모양 유사도 모델을 통해 연관하며, IoU 연관 알고리즘을 통해 연관하기 전에 칼만 필터 기반 모션 유사도 모델을 통해 칼만 필터를 업데이트한다. 이 때 1차, 2차 연관 모두 모양 유사도 모델을 통해 연관한다. 외형 유사도 모델의 경우, 작은 객체 크기 혹은 폐색으로 인한 객체 외형 손상으로 인해 추적 정확도가 하락할 수 있기 때문에 연관 시 적용하지 않는다. 결과적으로, Fig. 1에 나타나듯이 폐색 등으로 인한 기존 포함되지 못했던 검출 결과를 모두 연관 대상에 포함할 수 있게 된다. 또한 이를 통해 폐색이 발생하거나 추적 객체의 외형이 손상되더라도 장기간 추적이 가능해진다. 특히 군중 밀집으로 인한 폐색(Occlusion)이 빈번한 시퀀스에서, 기존 저수준 검출 결과에 의한 추적 객체들이 복구된다.

2. OC-SORT

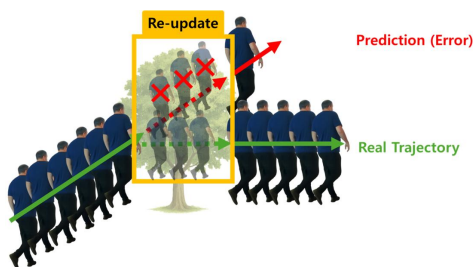


Fig. 2. Re-updating the trajectories with using OC-SORT

검출기의 검출 정확도가 개선된 이후로도 여전히 폐색에 의한 추적 객체 망실에 대한 적합한 해결 방법을 찾는 일은 어려운 문제다. 특히 추적 객체가 망실되는 상황은 객체가 폐색되었을 때 추적 객체가 선형으로 움직이지 않

았을 때 발생한다. 하지만 기존 모션 유사도 모델은 추적 객체의 경로가 선형이라는 가정 하에서 객체 위치를 업데이트한다. 하지만 이것은 폐색된 상태에서 실제 객체가 선형적으로 항상 움직인다고 가정하기 때문에 실제 추적 객체가 폐색된 상태에서 선형으로 움직이지 않는다면 해당 추적 객체의 칼만 필터는 선형으로 움직인다고 가정하고 폐색 기간 동안 연속적으로 칼만 필터를 업데이트 했기 때문에, 기존 칼만 상태 벡터는 노이즈에 오염된다. 따라서 잦은 폐색 혹은 비선형적인 추적 객체의 이동은 추적을 실패하게 만들 수 있음을 암시한다. 특히, 시퀀스의 프레임 레이트가 높아질수록 추적 객체의 이동 거리(픽셀 거리)는 감소하지만, 비선형적인 모션의 경우 객체 이동 방향이 쉽게 변경될 수 있기 때문에 위치 오차가 누적될 수 있다. 특히, 폐색 상태에서는 검출 결과를 이용할 수 없어 모션 유사도 모델을 통한 객체 위치 및 칼만 필터 업데이트만을 사용할 수 있기 때문에, 오차가 누적되기 쉽다. 이로 인해 장시간 추적이 끊긴 후 추적 객체가 복구될 때, 추적 객체가 비선형 모션을 나타내고, 검출 결과가 없다면 추적 결과가 복구되지 않거나, ID 스위치가 발생할 수 있다. OC-SORT [4]는 이 문제를 해결하기 위해, 추적 객체가 복구될 때 칼만 필터의 누적 에러를 최소화하도록 모델링된다. Fig. 2에 나타나듯이 추적 객체가 망실되기 전의 가장 최근의 추적 객체 위치와, 가장 최근에 연관에 성공한 추적 객체 위치가 주어질 경우 망실 구간 전후에 실제 검출 결과를 잇는 가상 검출 결과로 구성된 추적 궤적을 생성한다. 그리고 가상 검출 결과를 이용해 칼만 필터를 재갱신한다. 이를 통해 기존 칼만 상태 벡터가 노이즈에 오염되더라도 다시 재갱신함으로써 잦은 폐색 혹은 객체의 흔들림(jitter) 현상에 의한 객체 위치의 오차에 의한 칼만 상태 벡터 손상에 대응할 수 있다.

3. Hybrid-SORT

기존 추적 객체에 추가적인 단서를 좀 더 활용하는 것에 중점을 둔다. 기존 추적 객체의 모션 유사도 모델은 객체의 위치와 스케일을 주로 고려한다. 예를 들어, SORT [7]의 모션 유사도 모델의 칼만 상태 벡터는 추적 객체의 중앙 좌표와 그 변화량, 스케일과 그 변화량, 종횡비에 대한 요소로 업데이트 및 구성된다. 하지만 많은 객체가 좁은 공간에 겹치는 경우 이 요소만으로는 객체를 분별하기가 어렵다. 따라서, Hybrid-SORT [5]는 위 칼만 상태 벡터에 추가적으로 신뢰도와 그 변화량을 추가하여 칼만 상태 벡터를 다시 재모델링한다. 이를 통해 기존 추적 객체의 상태(예, 중앙좌표, 종횡비 등)를 객체 폐색 여부 등에 의한

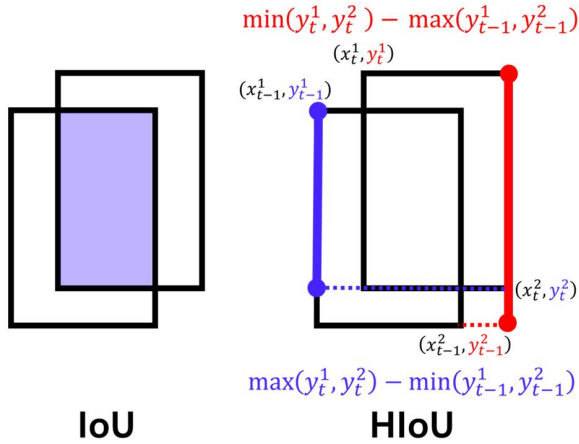


Fig. 3. Comparison IoU with HIoU

신뢰도 간의 상관관계를 직접 모델링할 수 있게 된다. 또한 기존 모양 유사도 모델에서 사용하는 IoU 대신, 객체의 높이를 추가적으로 활용할 수 있는 HMIoU (Height Modulated IoU)를 추가적으로 고려한다. 객체가 멀어질수록 높이(Height)가 감소하기 때문에 추적 객체의 거리(Depth)를 간접적으로 모델링할 수 있다. $t-1$ 프레임에서의 검출 결과 $d_{t-1} = (x_{t-1}^1, y_{t-1}^1, x_{t-1}^2, y_{t-1}^2)$ 와 t 프레임에서의 검출 결과 $d_t = (x_t^1, y_t^1, x_t^2, y_t^2)$ 라 하고, x^1, y^1, x^2, y^2 를 각각 왼쪽 상단 꼭짓점(x^1, y^1)과 오른쪽 하단 꼭짓점(x^2, y^2) 좌표라 할 때, HMIoU는 다음과 같다.

$$\text{HIoU} = \frac{\min(y_t^1, y_t^2) - \max(y_{t-1}^1, y_{t-1}^2)}{\max(y_t^1, y_t^2) - \min(y_{t-1}^1, y_{t-1}^2)} \dots (1)$$

$$\text{HMIoU} = \text{IoU} \cdot \text{HIoU} \dots (2)$$

이를 통해 추적 객체의 너비가 크게 변화하더라도, 높이를 안정적으로 추론할 수 있어 군집된 객체에서 추적 객체를 연관할 수 있다.

IV. Experiments

본 논문에서는 먼저, 검출 결과의 신뢰도에 따라 온 디바이스 환경에서 어떻게 추적 성능이 변경되는지 분석한다. 이 결과를 기반으로 두 번째 실험에서는 다양한 검출 기반 추적 알고리즘을 온 디바이스/온 서버 환경과 고해상도 환경(MOT17/DanceTrack)과 저해상도 환경(항공 전투 시나리오 데이터셋)에서 ByteTrack, OC-SORT,

Hybrid-SORT를 서로 비교해 연관 알고리즘 가이드라인을 제공한다.

Table 1. Detection Performance each detector and datasets.

YOLOX-X		
	MOT17 val.	DanceTrack val.
mAP	61.1%	79.5%
AP _s	19.9%	-
AP _m	53.0%	66.7%
AP _l	72.2%	80.1%
YOLOX-Tiny		
Aerial Battle Scenario Sets		
mAP	35.5%	
AP _s	35.7%	

1. Implementation Details

다양한 CPU 기반 추적 알고리즘을 비교하기 위해, ByteTrack, OC-SORT, Hybrid-SORT를 서로 비교한다.

Table 2. Comparison between MOT17/DanceTrack and aerial battle scenario sets

	Classes	Resolution
MOT17	1 Classes (Pedestrian)	480p~1080p
DanceTrack	1 Classes (Person)	720p~1080p
Aerial Battle Scenario	1 Classes (Aircraft)	256x256

θ_{high} 와 θ_{low} 은 ByteTrack [3]에서 경험적으로 설정된 0.5, 0.1로 각각 셋팅한다. ByteTrack과 동일한 고수준 검출 결과를 사용하여 비교하기 위해, OC-SORT와 Hybrid-SORT은 검출 신뢰도 0.5 이상의 검출 결과만 사용한다. OC-SORT와 Hybrid-SORT에 ByteTrack을 적용할 경우 ByteTrack과 동등하게 θ_{high} 와 θ_{low} 을 각각 0.5, 0.1로 셋팅한다. 추적 객체 유지 프레임은 MOT17 validation set과 DanceTrack validation set, 그리고 항공 전투 시나리오 데이터셋 모두 30 FPS로 설정하고, 30 프레임이 초과되는 경우 제거한다.

추적 시 검출 결과의 검출 정확도는 Table 1에 나타난다. 검출 결과의 메트릭은 AP (Average Precision)[16]를 나타낸다. 검출 결과와 Ground Truth 간의 IoU threshold를 0.50에서 0.95까지 0.05 간격으로 10개 평가하고 평균낸 값이다. AP_s, AP_m, AP_l은 객체 크기(area)에 따라 분리된 subset에서 동일한 방식(AP)으로 계산한

Table 3. Comparison the confidence threshold for detailed analysis on aerial battle scenario datasets.

Model	Average IoU	FP	FN	Initialization	Association	Track Management
$\theta > 0.1$	71.53%	121,145	1,057	0.26ms (5.5%)	1.89ms (40.6%)	2.37ms (50.9%)
$\theta > 0.2$	71.80%	97,567	3,120	0.24ms (6.0%)	1.89ms (47.0%)	1.74ms (43.2%)
$\theta > 0.3$	72.09%	73,602	6,289	0.19ms (6.0%)	1.67ms (53.1%)	1.13ms (36.0%)
$\theta > 0.4$	72.53%	50,016	9,715	0.17ms (6.6%)	1.42ms (54.8%)	0.85ms (32.8%)
$\theta > 0.5$	73.14%	33,778	12,951	0.15ms (7.0%)	1.22ms (57.1%)	0.61ms (28.5%)
$\theta > 0.6$	74.00%	17,701	16,843	0.12ms (7.0%)	0.99ms (57.5%)	0.44ms (25.7%)
$\theta > 0.7$	75.04%	10,918	20,058	0.11ms (9.1%)	0.84ms (59.5%)	0.29ms (20.5%)
$\theta > 0.8$	76.04%	4,708	24,581	0.10ms (8.9%)	0.69ms (59.7%)	0.20ms (16.8%)
$\theta > 0.9$	79.81%	1,919	34,664	0.08ms (9.7%)	0.41ms (50.0%)	0.16ms (19.5%)

AP이다 AP_s는 32x32 픽셀 미만, AP_m은 32x32 픽셀 이상, 96x96 픽셀 미만, AP_l은 96x96 픽셀 이상의 객체이다. 그리고 MOT17과 DanceTrack에서 객체 크기가 클수록 검출 정확도가 향상됨을 나타낸다. 항공 전투 시나리오 데이터 셋에서는 객체 크기가 작아 AP_s만 제공된다.

고해상도 이미지 입력이 주어지고, 고수준 검출기를 사용할 수 있는 환경(MOTChallenge MOT17, DanceTrack)에서는 추적 연관 알고리즘의 정량적 평가 및 분석을 하기 위해, 고성능 검출기인 YOLOX-X를 사용한다. 반면, 항공 전투 시나리오 데이터셋에서는 저해상도 입력 이미지가 주어지고, 저성능 검출기만 사용할 수 있는 환경을 고려하기 위해, YOLOX-Tiny [15]를 사용한다. 우리는 두 가지 환경에서 연관 시 발생하는 연관 알고리즘의 정량적 결과로부터 그 차이를 분석한다. 온 서버 환경은 인텔 제온 Gold 6330@3.1Ghz (TDP 205W) 2ea, RTX 3090 (TDP 350W) 1ea로 구성한다. 온 디바이스 환경은 저전력 환경을 구축하기 위해, 라즈베리 파이 4에도 채택된 바 있는 ARM Cortex A72@2.0Ghz 16 Core (TDP <30W)와 Hailo-8 NPU (26 TOPS, 2.5W)를 사용한다. 온 디바이스에서 YOLOX-Tiny를 평가하기 위해, [17]에 따라 모델 경량화를 진행한다. 특히, 본 논문에서는 동일한 검출기를 사용하는 대신 추적 알고리즘만을 바꾸어 사용한다. 따라서 동일한 검출 결과(검출된 객체 위치 및 박스, 그리고 신뢰도)가 각 연관 알고리즘 별로 동일하게 입력된다.

2. Datasets

본 논문에서는 고해상도 환경을 고려한 데이터셋인 MOTChallenge MOT17 validation set과 DanceTrack 데이터셋, 그리고 저해상도 및 낮은 분별력의 객체를 고려하기 위해, 실제 전장 환경을 고려하여 작성된 항공 전투 시나리오 데이터셋으로 구성된다. Table 2에 나타나듯이, MOT17-val/DanceTrack와 달리 전장 항공 전투 시나리오 데이터셋은 256x256 크기의 55,932장의 이미지로 구축된다. 이것은 항공 전투 시나리오에서 객체를 외형 정보로부터 판별하는 것이 매우 어려움을 나타낸다. 항공 전투 시나리오 데이터셋은 항공기(Aircraft) 클래스로 구성되며, 다양한 시점(top-view, side view)으로 생성되었다. 그리고 메트릭은 FP, FN, Average IoU를 사용한다. Average IoU는 추적 객체에 대한 Ground Truth와 예측 결과 간의 IoU를 계산하고, 전체 프레임에서 IoU가 0.5 이상인 대상에 대해 평균한 값이다. Association Speed의 경우 추적기가 하나의 프레임을 처리할 때 소모하는 시간이다. 항공 전투 시나리오에서 모델을 평가하기 위해, YOLOX에서 COCO17로 pretrained된 가중치로부터, 497,133장의 열화상 기반 항공 전투 시나리오 기반 학습 데이터로 300 Epoch와 배치 사이즈 128, 학습률 (Learning rate) 0.0005로 학습한다. MOT17과 DanceTrack은 다중 객체 추적 분야에서 공인된 벤치마크 데이터셋으로, MOT17 벤치마크에서는 다양한 도시 환경에서 보행자(Pedestrian) 클래스를 추적한다. DanceTrack은 동일한 복잡과 유사한 체형을 가진 댄서가 비선형 패턴으로 움직인다. 이것은 유사 외형을 가진

Table 4. Detailed comparison with various association algorithm on aerial battle scenario datasets.

	Model	Average IoU	FP	FN	On-Server Asso. Speed	On-Device Asso. Speed
Aerial Battle Scenario Sets	ByteTrack	71.07%	122,538	265	0.632ms	4.5ms
	OC-SORT	74.00%	14,965	15,858	0.718ms	2.2ms
	OC-SORT+Byte	72.00%	42,243	7,674	0.809ms	2.7ms
	Hybrid-SORT	73.68%	14,653	16,120	1.025ms	3.4ms
	Hybrid-SORT+Byte	72.40%	41,555	7,855	1.179ms	4.1ms

Table 5. Detailed comparison with various association algorithm on MOT17 and DanceTrack validation sets

	Model	MOTA	IDs	MT	IDF1	FP	FN	Asso. Speed
MOT17 Val. sets	ByteTrack	76.5	171	198	79.0	3373	9,119	4.50ms
	OC-SORT	74.0	213	163	77.4	1587	12,194	7.58ms
	OC-SORT+Byte	73.2	257	196	76.4	5487	8,716	8.62ms
	Hybrid-SORT	80.1	110	160	83.3	994	7,634	12.92ms
	Hybrid-SORT+Byte	81.9	121	177	83.6	2085	5,756	13.44ms
DanceTrack Val. sets	ByteTrack	88.3	1,864	247	54.1	8,959	15,582	3.76ms
	OC-SORT	87.2	1,898	205	51.7	2,205	24,625	4.37ms
	OC-SORT+Byte	88.6	2,256	240	52.0	7,683	15,618	4.40ms
	Hybrid-SORT	87.8	1,469	205	56.7	2,206	23,752	5.96ms
	Hybrid-SORT+Byte	89.5	1,720	237	60.7	6,044	15,821	6.32ms

객체가 매우 복잡한 움직임 패턴으로 움직이기 때문에, 추적을 매우 어렵게 한다. MOT17은 서로 다른 외형과 상대적으로 복잡하지 않은 움직임 패턴을 가진 반면, DanceTrack은 서로 유사한 외형 객체를 기반으로, 추적 객체가 복잡한 비선형 패턴을 가지기 때문에 두 데이터셋에서 모두 평가함으로써 추적 객체의 움직임에 대해 편향되지 않고 평가할 수 있다. 두 데이터셋은 MOTA, IDsw (ID Switching), MT (Mostly track), IDF1, FP (False Positive), FN (False Negative)을 메트릭으로 사용한다.

이 때 MOT17 validation은 MOT17-02/04/05/09/10/11/13으로 평가되며, 그 외 MOT17 학습 데이터셋의 시퀀스로 학습된 가중치로 모델을 초기화한다.

3. Sensitivity analysis for different confidence on aerial battle scenario datasets

본 섹션에서는 검출 신뢰도에 따라 추적 성능이 어떻게 변화하는 지 분석한다. 우리는 ByteTrack을 채택하지만, 동일한 검출 신뢰도로 평가하기 위해, 검출 신뢰도로 계층적 연관하는 대신, 연관 스코어로 계층적 연관한다. 연관 스코어는 모양 모델 유사도와 모션 모델 유사도 간의 곱으로 계산한다. 이 때 1단계 연관 스코어는 ByteTrack에서 제공하는 것과 동일하게, 1단계 연관 시엔 0.9, 2단계 연관 시엔 0.5로 설정한다. θ 는 0.5부터 0.9까지 0.1 간격으로 평가한다. Table 3의 괄호 안의 값은 전체 연산 비중의

백분율이다. Initialization은 검출 결과로부터 추적기를 생성하는 작업이며, Association은 ByteTrack의 1차 및 2차 연관 작업을 포함한다. Track Management는 연관 작업을 마치고, 추적 트랙의 종료, 추적 트랙 유지, 추적 트랙의 대기 작업을 나타낸다. 추적 정확도 관점에서, Table 3는 θ 가 증가할수록, Average IoU, FN이 증가하고 FP가 감소하는 것을 나타낸다. 이것은 고신뢰를 가진 검출 결과만을 사용하는 것이 추적 연관 정확도(Average IoU)를 개선하는 것을 나타낸다. 하지만, 고신뢰도를 가진 검출 결과만을 사용하기 때문에 FN이 증가하여 일관된 추적을 할 수 없음을 나타낸다. 특히, False Positive가 감소하는 것은 고신뢰도의 검출 결과만을 사용하기 때문에 추적 실패를 감소시킴을 나타낸다. 연산 복잡도에 관점에서, θ 가 증가할수록, 트랙 관리에 의한 연산 복잡도 비율이 감소하는 것을 확인한다. 이것은 FP가 감소함으로써 트랙 관리 연산에 의한 병목이 크게 감소한다는 것을 나타낸다.

4. Comparison tracking-by-detection on aerial battle scenarios datasets

본 섹션에서는 항공 전투 시나리오에서 최신 다중 객체 추적 알고리즘 비교함으로써 저해상도 환경에서 추적하기 위한 단서를 확인하고 분석한다. 이를 위해 우리는 ByteTrack, OC-SORT, Hybrid-SORT를 서로 비교한다. 온 서버 및 온 디바이스 모두 동일한 검출 결과를 기반으

로 평가한다. '+Byte'는 기존 추적 알고리즘의 연관 알고리즘을 1차 연관 알고리즘으로 하고, ByteTrack의 2차 연관을 추가적으로 적용한다. Table 4에 나타나듯이, ByteTrack은 가장 적은 FN과 가장 높은 FP, 그리고 Average IoU가 가장 낮다. 이것은 False Positive 또한 추적 연관 대상에 포함했기 때문이다. OC-SORT는 가장 높은 속도 2.2ms, 그리고 가장 높은 Average IoU를 나타낸다. 이것은 칼만 상태 벡터 보정 효과를 통해 효율적으로 추적 상태를 유지할 수 있음을 의미한다. Hybrid-SORT는 추적 객체의 높이에 대한 추가 단서를 이용해 칼만 필터를 갱신 및 IoU를 계산한다. 하지만 항공 전투 시나리오는 객체 크기가 매우 작을 뿐만 아니라, 고수준의 검출기를 사용할 수 없어 낮은 속도 3.4ms를 나타내며 불구하고, OC-SORT 대비하여 낮은 Average IoU를 나타낸다. +Byte 모델의 경우, 적용 전후 모두 FP가 증가하고, FN, Average IoU가 이 감소하는 것을 확인한다. 우리는 5개의 연관 알고리즘으로부터 Average IoU의 최대-최소 차이가 2.93%에 불과함을 확인한다. 이것은 검출 퀄리티가 개선되지 않는다면, 연관 알고리즘의 정확도 향상은 제한적임을 나타낸다.

5. Comparison between tracking-by-detection between high and low resolution inputs.

본 섹션에서 우리는 고해상도 입력 환경(MOT17/DanceTrack)과 저해상도 입력 환경(항공 전투 시나리오 데이터셋)에서 서로 다른 추적 알고리즘 5개를 서로 비교하여 입력 해상도에 따른 추적 알고리즘 별 경향성을 비교한다. Table 4는 항공 전투 시나리오 데이터셋에서 평가되었으며, Table 5는 MOT17/DanceTrack에서 비교된다. Table 4, Table 5에서 모두 ByteTrack을 OC-SORT/Hybrid-SORT에 적용할 때, FP가 증가하고, Average IoU와 FN이 감소하는 것을 관찰한다. 이것은 ByteTrack에 의한 추적 후보 증가에 의한 경향성이 데이터셋과 상관 없이 일관적으로 나타내는 것을 확인한다. 특히, Table 5에 나타나듯이, 온 서버 환경에서 ByteTrack은 항상 가장 빠른 연관 알고리즘으로, 온 서버 환경에서는 연관 대상을 늘리는 것보다는 연관 알고리즘에 의한 복잡도가 속도에 영향을 미친다는 것을 나타낸다. OC-SORT는 MOT17/DanceTrack에서 가장 높은 FN을 나타낸다. 이것은 칼만 필터 보정만으로는 누적 오차를 완전히 제거하기 어렵다는 점을 시사한다. Hybrid-SORT는 복잡한 모션 모델로 인해 가장 느린 속도를 나타낸다.

6. Complexity analysis between on-server and on-device.

Table 6. Comparison the speed for detailed analysis on aerial battle scenario datasets.

Model	Association	Track Management
On-Server	0.33ms (52.4%)	0.20ms (31.7%)
On-Device	1.49ms (33.1%)	2.53ms (56.2%)

Table 4에서 우리는 ByteTrack이 온 디바이스에서는 가장 느린 속도를 나타내지만, 온 서버에서는 가장 빠른 속도를 나타내는 것을 확인한다. 이 원인을 확인하기 위해, 온 서버와 온 디바이스 간의 연관 구성 요소 별로 속도를 비교한다. Table 6는 연관 속도가 4.5배 감소하는 것에 비해, 트랙 관리 속도는 12.7배 감소하는 것을 확인하여 트랙 관리에 의한 병목 현상이 발생하는 것을 확인한다. 이는 Table 1에서 확인한 바와 같이, 높은 FP로 인해 트랙 관리에 의한 복잡도가 증가하기 때문이며, 온 디바이스에서는 하드웨어 리소스 한계로 인해 많은 추적 대상에 대해 충분히 고려하기 어렵다는 점을 나타낸다.

7. Guideline for aerial battle scenario datasets

본 섹션에서는 앞서 평가한 결과를 통해 항공 전투 시나리오 등의 저해상도, 저성능 검출기 환경에서 검출 기반 추적 알고리즘을 설계하기 위한 전략을 제시한다. 우리는 위 실험에서부터 다음과 같은 사실을 발견하였다. 먼저, Table 3로부터 고신뢰도를 가진 검출 결과를 사용할 때 추적 정확도는 향상된다는 점을 확인한다. 그리고 동일한 검출 결과 하에서는 추적 정확도(Average IoU)가 크게 변동하지 않음을 확인한다. 우리는 이 결과로부터 다음과 같은 가이드라인을 제시한다.

(i) 고신뢰도를 가진 검출 결과를 사용하는 것이 검출 기반 추적 알고리즘의 정확도를 개선하는 것과 합치한다.

(ii) 시나리오에 따라 False Positive와 False Negative 중 줄여야 할 대상을 결정하기 위한 연관 알고리즘 채택을 하는 것이 합리적이다. 추적 대상을 놓치면 안되는 경우, ByteTrack과 같이 저신뢰도 검출 결과를 모두 연관 대상으로 채택해야 한다. 반면, 추적 노이즈를 감소시키고 싶은 경우 OC-SORT/Hybrid-SORT 같이 기존 추적 결과를 더 많은 단서를 통해 개선할 수 있는 연관 알고리즘을 적용해야 한다.

V. Conclusions

본 논문에서는 열화상 기반 저해상도 이미지에서 저수준 검출기 기반 검출 기반 추적 알고리즘에 대한 가이드라인을 제시한다. 우리는 검출 신뢰도 민감도 분석을 통해, 항공 전투 시나리오에서는 저수준 검출 결과가 FP/FN/연산 복잡도를 지배하는 핵심 요소임을 확인하였다. 저수준 검출 결과를 포함할 경우 추적 속도와 추적 정확도가 감소하지만 FN이 증가했고, 고신뢰 검출 결과만 사용할 경우 그 반대임을 확인했다. 종합하면, 추적 정확도를 개선하기 위해서는 고성능 검출기를 사용하는 것이 근본적인 해결책이다. 그리고 연관 알고리즘은 시나리오에 따라 연관 전략을 채택해야한다. 예를 들어, 저수준 검출이 불가피한 온 디바이스 항공 전투 시나리오에서 연관 알고리즘은 목표물 누락을 허용할 수 없는 상황에서는 저수준검출 결과까지 사용할 필요가 있음을 나타낸다. 본 연구는 이러한 분석을 통해, 군사 및 항공 분야와 같이 실제 온 디바이스 환경에서 검출 품질 제약을 전제로 어떤 연관 전략을 결정해야 하는지에 대한 정량적 근거 기반 가이드라인을 제공하였다. 본 연구진은 본 논문이 각 산업 분야 및 연구에서 적합한 연관 알고리즘을 결정하는 것에 도움이 되길 기원한다.

ACKNOWLEDGEMENT

This work was supported by Korea Research Institute for defense Technology planning and advancement(KRIT) grant funded by the Korea government(DAPA(Defense Acquisition Program Administration)) (No. 21-107-F00-015(KRIT-CT-22-024-02), 2023)

REFERENCES

- [1] Z. Wang, L. Zheng, Y. Liu, Y. Li, and S. Wang, "Towards real-time multi-object tracking." European conference on computer vision. Cham: Springer International Publishing, pp. 107-122, Glasgow's Scottish Event Campus, UK, August 2020. DOI: 10.48550/arXiv.1909.12605
- [2] Y. Zhang, C. Wang, X. Wang, W. Zeng and W. Liu, "Fairmot: On the fairness of detection and re-identification in multiple object tracking," International journal of computer vision, Vol. 129, pp. 3069-3087, September, 2021. DOI: 10.1007/s11263-021-01513-4
- [3] 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. Cham: Springer Nature Switzerland, pp. 1-21, Expo Tel Aviv International Convention Center, Israel, October 2022. DOI: 10.1007/978-3-031-20047-2_1
- [4] J. Cao, J. Pang, X. Weng, R. Khirodkar, and K. Kitani, "Observation-centric sort: Rethinking sort for robust multi-object tracking." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9686-9696, Vancouver Convention Centre, Canada, June 2023. DOI: 10.1109/CVPR52729.2023.00934
- [5] 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." Proceedings of the AAAI conference on artificial intelligence., pp. 6504-6512, Vancouver Convention Centre, Canada, February 2024. DOI: 10.1609/aaai.v38i7.28471
- [6] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric." ICIP. IEEE, pp. 3645-3649, China National Convention Center, China, September 2017. DOI: 10.1109/ICIP.2017.8296962
- [7] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking." 2016 IEEE international conference on image processing (ICIP). pp. 3464-3468, Phoenix Convention Centre, USA, September 2016. DOI: 10.1109/ICIP.2016.7533003
- [8] H. Zhang, F. Li, S. Liu, L. Zhang, H. Su, J. Zhu, L. Ni, and H. Shum, "Dino: Detr with improved denoising anchor boxes for end-to-end object detection." International Conference on Learning Representations. Kigali Convention Centre, Rwanda, May 2022. DOI: 10.48550/arXiv.2203.03605
- [9] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable detr: Deformable transformers for end-to-end object detection." International Conference on Learning Representations. Virtual Conference, Austria, 2021. DOI: 10.48550/arXiv.2010.04159
- [10] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, "MOT16: A benchmark for multi-object tracking." arXiv preprint arXiv:1603.00831. June 2016. DOI: 10.48550/arXiv.1603.00831
- [11] P. Sun, J. Cao, Y. Jiang, Z. Yuan, S. Bai, K. Kitani, and P. Luo, "Dancetrack: Multi-object tracking in uniform appearance and diverse motion." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20993-21002, Morial Convention Center, USA, 2022. DOI: 10.1109/CVPR52688.2022.02032
- [12] G. Ciaparrone, F. Sanchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, "Deep learning in video multi-object tracking: A survey." Neurocomputing, Vol. 381, pp. 61-88, March 2020. DOI: 10.1016/j.neucom.2019.11.023
- [13] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and TK Kim, "Multiple object tracking: A literature review." Artificial

intelligence, Vol. 293, April 2021. DOI: 10.1016/j.artint.2020.103448

- [14] T. Meinhardt, A. Kirillov, L. Leal-Taixe, and C. Feichtenhofer, "Trackformer: Multi-object tracking with transformers." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8844-8854, Morial Convention Center, USA, 2022. DOI:10.1109/CVPR52688.2022.00864
- [15] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021." arXiv preprint arXiv:2107.08430, August 2021. DOI: 10.48550/arXiv.2107.08430
- [16] TY. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and CL. Zitnick, "Microsoft coco: Common objects in context." European conference on computer vision. Cham: Springer Nature Switzerland, pp. 740-755, Zurich, Switzerland, September 2014. https://doi.org/10.1007/978-3-319-10602-1_48
- [17] DH Hur, DH Park, DW. Kim, JW. Baek, and SH. Bae, "Object Detection Performance Analysis between On-GPU and On-Board Analysis for Military Domain Images." Journal of the Korea Society of Computer and Information, Vol. 29, No. 8, pp. 157-164, August, 2024. DOI: 10.9708/jksci.2024.29.08.157

Authors



Du-Hwan Hur received the B.S. degree in Computer Engineering from Hanbat National University in 2021, and is currently pursuing the M.S. - Ph.D. integrated course degree with the Department of Electrical Computer

Engineering at Inha University, Korea. His current research interests are object detection, model compression, efficient learning, and on-device AI.



Dae-Hyeon Park received the B.S degree in Computer Engineering from Inha University in 2020 and the M.S degree with the Department of Electronic Computer Engineering at Inha University in 2023.

He is currently pursuing the Ph.D. degree with the Department of Electronic Computer Engineering at Inha University, Korea and His current research interests are single/multi-object tracking, multi-modal learning, real-time system and self-attention mechanism.



Deok-Woong Kim received the B.S. degree with Department of Industrial engineering, Public administration, Inha University, South Korea. He is currently pursuing the M.S. - Ph.D. integrated course degree.

His research interests are knowledge distillation, quantization, data-free and on-board AI.



Junhyeong Bak received his B.S. degree in Electronics Engineering from Hannam University in 2020, and M.S. degree in Electrical and Computer Engineering from Inha University in 2022.

He is currently a research engineer at Missile System Seeker R&D, LIG Nex1 Co., Ltd. He is interested in fast visual object detection and tracking on low-power hardware.



Ingu Park received the B.S., in Computer Science and Engineering and M.S. degrees in Computer Engineering from Pusan National University, Korea, in 2004 and 2006, respectively.

Mr. Park is currently a chief research engineer in Missile System Seeker R&D, LIG Nex1. He is interested in embedded system, and object detection & tracking using deep learning.



Ki-Nam Lee received his B.S. degree in Mechanical Engineering from Myongji University in 1997, and M.S. degree in Mechanical Engineering from Korea University in 2013, respectively.

He is currently a chief research engineer at Missile System Seeker R&D, LIG Nex1 Co., Ltd. He is interested in control system, signal processing and sensor.



Seung-Hwan Bae received the BS degree in information and communication engineering from Chungbuk National University, in 2009 and the MS and PhD degrees in information and communications from the Gwangju

Institute of Science and Technology (GIST), in 2010 and 2015, respectively. He was a senior researcher at Electronics and Telecommunications Research Institute (ETRI) in Korea from 2015 to 2017. He was an assistant professor in the Department of Computer Science and Engineering at Incheon National University, Korea from 2017 to 2020. He is currently an Associate Professor with the Department of Computer Engineering at Inha University, His research interests include object tracking, object detection, generative model learning, continual learning, on-device ML, etc.