JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 31 No. 7, pp. 95-106, July 2026
https://doi.org/10.9708/jksci.2026.31.07.095

Advanced Strategy for On-Device Single Object Tracking Algorithm Robust to Multi-Domain Shift and Occlusion

Ji-Su Yoon*, Jae-Yeob Kim**, Hyun-Sung Jang***, Seung-Wook Park***

*Research Engineer, EO/IR Systems R&D, LIG Defense&Aerospace Co., Ltd., Pangyo, Korea
**Senior Research Engineer, EO/IR Systems R&D, LIG Defense&Aerospace Co., Ltd., Pangyo, Korea
***Chief Research Engineer, EO/IR Systems R&D, LIG Defense&Aerospace Co., Ltd., Pangyo, Korea

[Abstract]

In this paper, we propose a robust object tracking tuning strategy suitable for gimbal control in mobile platforms, ensuring stable operation across diverse scenario. The core method applies domain generalization for robustness against unseen environments and introduces an occlusion-aware head, trained with spatial Cox process-based random masking and pseudo-labels, to autonomously recognize occlusion states. Additionally, to prevent bounding box jittering during tracking failures, we employ a rule-based post-processing algorithm that detects outliers in the bounding box center shift and real-time response map statistics using a Z-score approach. This immediately corrects trajectory instability, maintaining stable outputs for gimbal systems. Experimental results show that the proposed model achieves superior robustness over baselines. Furthermore, it operates at 77.5 FPS in a high-performance GPU environment and 30.3 FPS in an edge device environment, proving its high efficiency in satisfying both performance and real-time requirements in resource-constrained on-device environments.

▸ **Key words**: Object Tracking, Domain Generalization, Occlusion Awareness, On-device, Tracking Stability

[요 약]

본 논문에서는 이동형 플랫폼의 짐벌 시스템과 연동을 고려하여, 다양한 센서 및 환경 시나리오에서도 안정적으로 동작하는 강건한 객체 추적 모델 고도화 방안을 제안한다. 제안하는 방법의 핵심은 도메인 일반화 기법을 적용하여 미지의 환경 변화에 대한 강건성을 확보하고, 공간적 콕스 과정 기반 랜덤 마스킹과 의사 라벨을 학습한 가림 판단 헤드를 통해 모델 스스로 타겟의 가려짐 상태를 인지하게 만드는 것이다. 또한 추적 실패 시 바운딩 박스의 지터링 현상을 방지하기 위해, Z-score를 기반으로 박스 중심점의 위치 변화와 실시간 응답 맵 통계의 이상치를 복합적으로 탐지하는 규칙 기반의 후처리 알고리즘을 적용하였다. 이를 통해 결과 바운딩 박스의 불안정성을 즉각 보정하여 짐벌 시스템을 위한 안정적인 출력을 유지한다. 실험을 통해 제안된 객체 추적 모델은 기존 모델 대비 높은 강건성을 달성할 뿐만 아니라, 고성능 GPU 환경에서 77.5 FPS, 엣지 디바이스 환경에서 30.3 FPS의 추론 속도를 기록하며 자원이 제한된 온디바이스 환경에서도 고성능과 실시간성을 동시에 충족하는 효율적인 시스템임을 입증하였다.

▸ **주제어**: 단일 객체 추적, 도메인 일반화, 가림 판단, 온디바이스, 추적 안정성

• First Author: Ji-Su Yoon, Corresponding Author: Seung-Wook Park
*Ji-Su Yoon (jisu.yoon25@ligdna.com), EO/IR Systems R&D, LIG Defense&Aerospace Co., Ltd.
**Jae-Yeob Kim (jaeyeob.kim@ligdna.com), EO/IR Systems R&D, LIG Defense&Aerospace Co., Ltd.
***Hyun-Sung Jang (hyunsung.jang@ligdna.com), EO/IR Systems R&D, LIG Defense&Aerospace Co., Ltd.
***Seung-Wook Park (seungwook.park@ligdna.com), EO/IR Systems R&D, LIG Defense&Aerospace Co., Ltd.
• Received: 2026. 04. 30, Revised: 2026. 05. 27, Accepted: 2026. 06. 17.

Copyright © 2026 The Korea Society of Computer and Information
http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945
www.kci.go.kr

I. Introduction

컴퓨터 비전 분야에서 비디오 시퀀스 내 객체의 움직임을 이해하고 분석하는 기술은 딥러닝의 발전과 함께 빠른 성장을 이루었다. 그중에서도 단일 객체 추적(Single Object Tracking, SOT)은 비디오의 첫 프레임에서 특정 타겟의 초기 상태 정보를 바탕으로, 이어지는 연속적인 프레임들에서 타겟의 위치 및 상태를 실시간으로 추정하는 핵심 기술이다. 이러한 SOT 기술은 최근 무인 항공기(UAV), 지능형 로봇 등 이동형 플랫폼에서의 수요가 급증하고 있으며, 특히 대상 객체를 화면 중앙에 안정적으로 유지하기 위해 짐벌(Gimbal) 시스템과 결합된 객체 추적 기술은 무인기, 정찰 및 감시 등 다양한 분야에서 필수성이 대두되고 있다. 이러한 이동형 시스템은 급변하는 외부 환경에 즉각적으로 대응해야 하므로, 지연 없는 실시간(Real-time)동작을 보장하는 것이 매우 중요하다. 이를 위해서는 중앙 서버를 거치지 않고 카메라가 장착된 엣지 레벨에서 직접 영상을 처리하는 온디바이스(On-device)기반의 추적 시스템 구축이 요구된다. 또한 최근의 시스템들은 가용 정보량을 극대화하고 환경적 제약을 극복하기 단일 센서가 아닌 다중 센서를 복합적으로 활용하는 추세이다. 따라서 이질적인 다양한 센서 입력 환경에서도 강인하게 동작할 수 있는 시스템 개발의 중요성이 더욱 커지고 있다. 이에 본 논문은 다중 파장 대역 센서를 활용한 짐벌 기반의 온디바이스 환경을 가정하고, 다중 센서 및 다양한 환경 시나리오의 변화에도 유연하게 대응할 수 있는 강건한 추적기를 설계 및 구현하고자 한다.

이러한 다중 센서 및 동적 복합 시나리오에서 강건한 추적 시스템의 구현을 위해서는 여러 도메인 데이터에 대한 모델의 적응력이 필수적이나, 실제 환경에 적용하기에는 다음과 같은 현실적인 문제점들이 존재한다. 대부분의 딥러닝 기반 객체 추적 연구는 특정 도메인에 한정된 데이터를 이용해 학습을 수행한다. 그 결과 학습데이터의 분포를 벗어난 이종의 센서 데이터가 입력되거나, 환경 조건 변화가 발생할 경우, 성능 저하가 발생하는 한계가 존재한다. 이를 방지하기 위해 모든 잠재적 도메인 데이터로 학습시킴으로써 성능을 보완할 수 있으나, 이는 도메인별로 방대한 규모의 데이터를 확보하고, 정밀한 라벨링 작업을 수행해야 하는 등 과도한 자원 소모를 수반하므로 현실적인 확장성에 한계가 있다.

더불어, 대다수의 객체 추적 연구는 정밀도(Precision Score), 성공률(Success Rate) 등 정량적 지표에 의존한다. 이러한 평가 체계는 실제 시나리오에서 빈번히 발생하는 가림, 유사 객체 간섭, 시야 이탈과 같은 추적 실패 상황에 대한 대응 메커니즘을 반영하지 못한다. 그 결과, 추적 실패가 발생하면 모델이 출력하는 바운딩 박스의 크기와 위치가 급격하게 요동치는 불안정성(Jittering)이 나타난다. 이러한 불안정성은 정밀한 기계적 구동이 요구되는 짐벌 시스템의 실제 적용 가능성과 신뢰성을 근본적으로 저하한다.

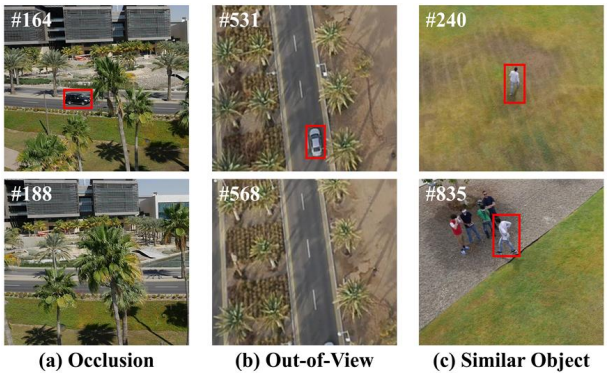

Fig. 1. Challenging scenarios in object tracking: occlusion, out-of-view, and similar objects

본 연구는 다중 센서 및 동적 복합 시나리오에서 안정적으로 동작하는 짐벌 기반 객체 추적 모델의 효율적인 파인튜닝 기법을 검증한다. 앞서 언급한 바와 같이, 다중 센서 및 복잡한 실세계 환경을 모두 아우르는 방대한 데이터를 구축하는 것은 현실적인 한계가 따른다. 따라서 제한된 데이터만으로도 다양한 파장 대역의 센서 및 가변적인 환경 요인에 강인하게 동작하는 모델을 학습하기 위해, 기 연구된 도메인 일반화 기법 Random Convolution[1]을 활용한다. 구체적으로, 사전 학습된 백본 네트워크의 기초 특징 추출 능력을 보존하면서도 텍스처 변화에 적응할 수 있도록, 네트워크의 심층 레이어 일부를 튜닝하여 모델이 객체의 온전한 형태 정보에 집중하도록 유도한다.

나아가, 본 연구는 가려짐이나 시야 이탈과 같이 추적 대상이 사라지는 예외 상황에서도 모델이 강건하게 대응할 수 있는 학습 전략을 제안한다. 대다수의 딥러닝 기반 추적 모델들은 타겟 객체가 현재 프레임에 존재한다는 전제를 두며, 이는 결과 불안정성의 원인이 된다. 이를 해결하기 위해 모델 학습 단계에서 인위적인 마스킹을 적용하고, 각 마스킹 비율에 따른 가시성(Visibility) 클래스를 자동으로 부여한다. 즉, 모델이 현재 프레임에서 객체의 가려짐 정도를 스스로 인지하는 학습 전략을 사용한다.실험 결과, RandConv 기반의 파인튜닝 기법을 적용했을 때, 타겟 도메인에서 베이스라인 모델 대비 최대 1.85%의 정밀도(Precision Score) 향상을 달성하였다. 또한, 타겟 객체의 가림 현상을 스스로 인지하는 가림 판단 및 재포착

www.kci.go.kr

Fig. 2. Bounding box instability during a tracking failure

모듈을 통해, 추적 대상이 일시적으로 시야에서 사라지는 예외 상황에서 바운딩 박스의 지터링을 억제할 수 있음을 시퀀스 별 IoU 비교를 통해 확인할 수 있었다. 더불어 추론 속도 측정 실험 결과, 제안된 기법들의 추가적인 연산 오버헤드는 베이스라인 모델 대비 프레임 당 약 2.4ms 정도에 불과하였다. 이는 기존 모델의 실시간성을 저해하지 않으면서도, 실제 온디바이스 환경에서 강건한 추적 시스템으로 고도화할 수 있는 매우 효율적인 방법임을 입증하였다.

본 연구의 주요 기여는 다음 세 가지로 요약한다.

(1) 다중 센서 및 동적 복합 시나리오에서 요구되는 방대한 데이터 수집 비용을 절감하기 위해, 기 연구된 RandConv 기법을 짐벌 기반 추적 모델의 파인튜닝 과정에 도입하였다. 해당 일반화 기법이 객체 추적 모델의 텍스처 편향을 완화하고 성능을 향상하는 데에도 효과적으로 작용함을 실험적으로 확인하였다.

(2) 인위적 마스킹을 통한 가시성 학습 헤드를 도입하여, 프레임 이탈 및 가림 상황 시 발생하는 바운딩 박스의 지터링 현상을 차단하고 추적 실패 후에도 타겟을 안정적으로 재포착할 수 있는 매커니즘을 설계하였다.

(3) 모델의 일반화 성능과 예외 상황 대응 능력을 향상하면서도, 추가적인 연산 시간을 최소화하여 연속적인 구동이 필수적인 짐벌 기반 추적 시스템에 즉각적으로 적용 가능한 실용성을 확보하였다.

II. Preliminaries

1. Single Object Tracking

단일 객체 추적(Single Object Tracking, SOT) 기술은 주어진 비디오 시퀀스의 첫 프레임에서 지정된 타겟의 초기 위치와 크기 정보를 바탕으로, 이후 연속되는 프레임에서 해당 객체의 상태를 지속적으로 추정하는 컴퓨터 비전의 핵심 분야이다. 최근 몇 년간 딥러닝 기술이 비약적으로 발전함에 따라, 객체 추적 알고리즘은 특징 추출 및 매칭 방식에 있어 큰 패러다임 변화를 겪어왔다.

CNN-based Tracker

초기의 심층 신경망 기반 객체 추적 연구는 주로 합성곱 신경망(CNN)의 강력한 공간적 특징 추출 능력에 의존하였다. 특히 샴 네트워크(Siamese Network) 구조의 도입은 객체 추적의 성능과 속도를 획기적으로 향상시켰다. 대표적으로 SiamFC[2]는 타겟 템플릿과 탐색 영역 이미지를 공유 CNN 백본에 통과시킨 후, 두 특징 맵 간의 상호 상관관계(Cross-correlation)를 계산하여 객체의 위치를 찾는 효율적인 프레임워크를 제시하였다. 이후 Li 등[3]은 객체 탐지 분야의 Region Proposal Network(RPN)를 샴 네트워크에 결합한 SiamRPN을 제안하여, 크기 변화 및 비율 변화에 강건하고 정확한 바운딩 박스를 예측할 수 있게 하였다. ResNet과같이 더 깊은 신경망을 백본으로 채택한 후속 연구들 또한 뛰어난 성능을 달성하였으나, CNN 구조 특성상 국소적인 수용 영역(Local Receptive Field)에 의존하기 때문에 타겟의 전역적인 문맥 정보나 장기적인 시공간적 변화를 포착하는 데에는 구조적인 한계가 존재했다.

Transformer-based Tracker

최근 컴퓨터 비전 분야에서 트랜스포머(Transformer) 아키텍처가 뛰어난 전역적 문맥 포착 능력을 입증함에 따라, 객체 추적 분야 역시 이를 적극적으로 활용하였다. 초기 연구에선 주로 템플릿과 탐색 영역 간의 고도화된 정보 교환을 위해 어텐션(Attention) 메커니즘을 활용하였다. 대표적으로 TrnsT[4]는 기존의 단순한 상호 상관관계 연산을 대체하는 어텐션 기반의 특징 융합 모듈을 도입하여 매칭 성능을 크게 향상시켰다. 이어 STARK[5]는 트랜스포머를 시공간적 특징 융합 모듈로 채택하여, 타겟의 공간적 정보와 동적 템플릿을 통한 시간적 상태 변화를 동시에 모델링하며 성능을 향상시켰다.

이후 연구 흐름은 특징 추출과 정보 융합을 분리하던 기존의 파이프라인에서 벗어나, 단일 비전 트랜스포머(ViT) 내에서 두 과정을 동시에 수행하는 통합형 모델로 진화하였다. MixFormer[6]는 혼합 어텐션 모듈을 통해 다중 스케일 특징 추출과 타겟-탐색 영역 간의 정보 융합을 End-to-End로 결합함으로써, 구조적 복잡성을 줄이는 동시에 강력한 식별력을 확보하였다. 더 나아가, 객체 탐지 모델인 DETR에서 영감을 받아 타겟 쿼리(Target query)

를 도입하거나, ARTrack[7]과 같이 시간적 맥락을 자기회귀(Autoregressive) 방식으로 학습하는 등, 비디오 시퀀스 내의 복잡한 시공간적 관계를 탐색하며 최첨단 성능을 갱신하고 있다.

2. Occlusion-aware Tracker

단일 객체 추적에서 타겟이 배경 구조물이나 다른 객체에 의해 가려지는 상황(Occlusion)은 타겟의 외형 정보를 훼손하고 모델의 특징 매칭을 방해하여 추적 실패를 유발하는 주요 원인이다. 이를 해결하기 위해 다양한 가려짐 인식 및 복구 메커니즘이 연구되어왔다.

먼저 타겟 주변의 문맥이나 배경 정보를 적극적으로 활용하는 접근법이 등장했다. 대표적으로 Bhat 등[8]의 KYS(Knowing Your Surroundings)는 추적 과정에서 타겟 주변의 배경 구조물과 방해 요소(Distractors)들의 상태를 모델링하여, 타겟이 일시적으로 가려지더라도 주변 환경 정보를 바탕으로 타겟의 위치를 예측할 수 있도록 설계되었다. 최근에는 트랜스포머의 전역 문맥 파악 능력을 활용하여, 가려짐 상황에 본질적으로 강건한 모델들이 연구되기도 한다. 비전 트랜스포머는 타겟의 일부 패치가 가려지더라도 나머지 보이는 패치 간의 상호작용을 유도함으로써 객체의 전체적인 의미 정보를 유지할 수 있다. 특히 OSTrack[9]은 특징 추출과 관계 모델링을 단일 스트림으로 결합함과 동시에, 훈련 과정에서 의도적으로 타겟 패치의 일부를 마스킹하는 기법을 적용했다. 마스킹 기반 학습 전략을 더욱 고도화한 연구도 등장했다. ORTrack[10]은 단순한 무작위 마스킹의 한계를 극복하고자 공간적 콕스 과정(Spatial Cox Process)을 도입하여 실제 환경에서 발생하는 불규칙적이고 군집화된 가려짐 패턴을 정교하게 시뮬레이션하였다.

3. Domain Generalization

도메인 일반화(Domain Generalization, DG)는 타겟 도메인의 데이터에 접근할 수 없는 상황에서 학습된 모델이 미지의 도메인(Unseen Domain)에서도 높은 성능을 유지하도록 만드는 것을 목표로 한다. 사전에 타겟 도메인을 특정하고 데이터에 접근 가능한 도메인 적응(Domain Adaptation)과 달리, 도메인 일반화는 타겟 도메인에 대한 어떠한 사전 정보도 주어지지 않는 미지의 환경을 전제로 한다. 따라서 오직 소스 도메인만으로 도메인 변화에 강건한 모델을 구축하는 데 집중하며 다음과 같은 접근법들로 연구되어 왔다.

가장 직관적인 접근법은 여러 소스 도메인 간에 공통적으로 존재하는 도메인 불변적(Domain-invariant)인 특징을 추출하는 것이다. Muandet등[11]은 커널 기반 최적화 기법을 도입하여 도메인 간의 분포 차이를 최소화하는 동시에 클래스 분류 성능을 유지하는 특징 표현 학습 프레임워크를 제안하며 도메인 일반화 분야의 기초를 마련했다. 이후에는 적대적 학습(Adversarial learning)을 활용하여, 특징 추출기가 도메인 특화 정보를 지우고 객체의 본질적인 의미 정보만 남기도록 유도하는 연구들이 주류를 이루었다. 또한, 모델이 특정 소스 도메인의 편향된 스타일에 과적합되는 것을 방지하기 위해, 다양한 증강 기법을 활용하는 연구도 활발히 진행되었다. Carlucci 등[12]은 직소 퍼즐 풀기와 같은 자가 지도 학습 보조 태스크를 도입하여, 모델이 이미지의 공간적 구조와 도메인에 독립적인 의미적 특징을 학습하도록 유도하였다. 또한 MixStyle[13]은 특징 맵 수준에서 서로 다른 도메인의 통계량을 혼합하여 새로운 스타일의 특징을 합성함으로써, 연산량 증가 없이 모델의 도메인 일반화 성능을 크게 향상시켰다.

III. The Proposed Scheme

본 장에서는 다중 센서 및 동적 복합 시나리오에서 제한된 학습데이터만으로도 강인한 객체 추적 모델로 고도화하기 위한 효율적인 파인튜닝 전략을 설명한다. 이를 위해 극복해야 할 핵심 도전과제는 크게 두 가지이다. 먼저, 다중 파장 대역 및 변동성이 큰 환경 시나리오가 혼재된 조건에서, 기존 소스 도메인에서의 추적 성능을 잃지 않으면서 미학습된 타겟 도메인에서 일반화 능력을 극대화하는 것이다. 다음으로, 타겟 객체의 가려짐과 같은 추적 실패 상황을 모델이 스스로 인지하게 함으로써 바운딩 박스의 비정상적인 발산 및 지터링 현상을 억제해야 한다. 이와 같은 과제들을 해결하기 위해 설계된 모델 튜닝 파이프라인은 다음의 세 가지 핵심 단계로 이루어진다.

첫 번째 단계는 네트워크 입력단에 RandConv[1] 기반의 전처리 기법을 도입하여 모델을 파인튜닝하는 것이다. 이는 특정 파장 대역이나 센서 고유의 텍스처 특성에 모델이 과적합 되는 것을 방지한다. 두 번째 단계는 객체 가려짐 상황에 능동적으로 대응하기 위한 가시성 예측 헤드를 추가하였다. 학습 과정에서 랜덤 마스킹을 적용하여 가려짐 정도에 대한 의사 라벨(Pseudo-label)을 자가 생성함

www.kci.go.kr

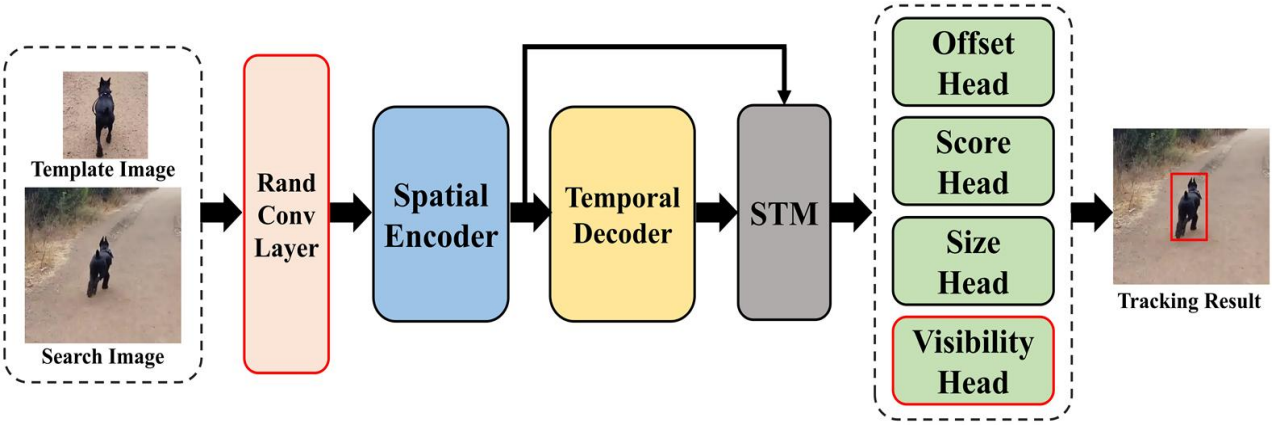

Fig. 3. Overall Architecture (the proposed modules are highlighted with red borders)

으로써, 별도의 추가 정답 데이터 없이도 모델 스스로 현재 타겟의 가시 상태를 학습하고 판단할 수 있도록 학습시킨다. 마지막 단계는 파인튜닝된 모델의 추론 결과에 규칙 기반의 후처리 알고리즘을 결합하는 것이다. 예측된 가시성 클래스와 바운딩 박스 궤적을 통합적으로 분석하여, 추적 실패 상황에서도 추적 모델의 신뢰도를 유지하고 최종 위치 예측 결과를 한층 더 정밀하게 보정한다.

1. Simple Application Scheme for Domain Generalization

다중 도메인 환경에서도 강건하게 동작하는 모델을 학습할 때 직면하는 제약 요인 중 하나는 도메인 간 학습데이터의 심각한 불균형이다. 다중 파장 대역 시나리오의 경우 가시광선(RGB) 대역의 객체 추적 데이터셋은 대규모로 구축되어있는 반면, 적외선(Infrared) 등 특수 센서 기반의 정제된 고품질 라벨링 데이터는 현실적으로 확보하기가 어렵다. 이러한 양적 불균형은 모델의 특징 추출 능력을 가시광선 도메인의 특정 시각적 스타일에 고착시키며, 타 도메인 입력이 주어졌을 때 성능 저하 가능성이 있다. 이와 같은 도메인 격차를 극복하고, 복잡한 구조적 변경 없이도 다수의 타겟 도메인에서 일관된 성능 향상을 유도하고자 도메인 일반화 기법인 Random Convolution 연산을 활용한다.

RandConv 연산은 네트워크 입력 전처리 단계에 위치하며, 학습 과정에서 매 반복(Iteration)마다 수식 (1)과 같이 가중치가 특정 분포로부터 무작위로 초기화되는 합성곱 계층으로 작동한다. 원본 입력 이미지가 이 계층을 통과하면, 객체의 전반적인 위치나 윤곽선과 같은 전역적인 구조적 정보는 유지된 채 지역적 질감은 새로운 형태로 변환된다. 이를 통해 모델은 센서 및 조명 종속적인 시각적 특성에 의존하는 대신, 형태 기반의 불변 특징을 학습하게 된다.

한편, 도메인 일반화의 핵심은 미지의 타겟 도메인에서의 성능 향상뿐만 아니라, 원본 소스 도메인에서의 성능 저하를 방지하는 것이다. 특히 단일 객체 추적 태스크는 타겟의 정확한 공간적 해상도를 잃지 않고 정밀하게 바운딩 박스를 예측해야 하므로, 과도한 영상 변형은 오히려 성능에 악영향을 미칠 수 있다. 따라서 공간 정보의 손실 및 좌표 왜곡을 방지하고자 RandConv의 커널 크기를 3×3로 제한하였으며, 원본 영상과 동일한 해상도 유지를 위하여 패딩을 적용하였다. 또한, 수식 (2)와 같이 원본 데이터가 가진 유효한 특징 맵을 보존하기 위해 변환 영상과 원본 영상 간의 선형 보간(Linear Interpolation)을 수행하였다. 이를 통해 모델이 새로운 도메인에 대한 적응력을 확보함과 동시에 원본 가시광선 도메인에서 추적 정확도를 안정적으로 유지할 수 있도록 적절한 균형을 유도하였다.

$$\Phi \sim N(0,\sigma) \tag{1}$$

$$I_{mix} = \alpha I + (1-\alpha)(I * \Phi) \tag{2}$$

2. Occlusion-Aware Head

베이스라인으로 선정한 객체 추적 모델은 타겟 중심점 점수, 바운딩 박스 오프셋, 그리고 크기를 추정하는 3개의 독립적인 헤드로 구성된다. 제안하는 파이프라인은 기존 구조에 더해 타겟 쿼리와 공간적 특징 맵 간의 크로스 어텐션을 수행하여 객체의 가시성을 총 4단계로 예측하는 분류 헤드를 추가한다. 이 헤드는 현재 프레임에서 타겟 객체의 가려짐 정도를 모델 스스로 인지하고, 후술 되는 후처리 알고리즘에서 추적 상태를 판단하는 기준값을 출력하는 역할을 한다. 해당 모듈을 학습시키기 위해 추가적인 정답 라벨링 없이도 동작 가능한 자가 지도 형태의 입력 전처리 과정을 도입하였다. 우선 탐색 영역 이미지에

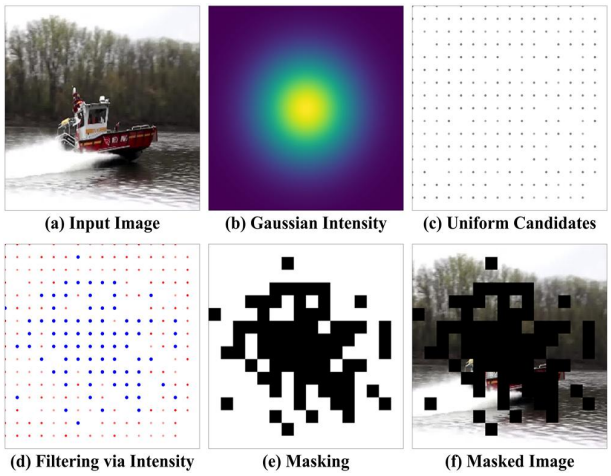

Fig. 4. The process of generating random masks via a spatial Cox process

ORTrack[10]에서 제안된 공간적 콕스 과정(Spatial Cox Process) 기반의 무작위 마스킹을 적용한다. 이후, 정답 바운딩 박스 내부에 포함되는 마스킹 픽셀의 비율을 계산하여 모델이 예측할 가시성 클래스 라벨을 자동으로 부여한다. Fig. 5에 도시된 바와 같이 타겟 영역 내 마스킹 비율에 따라 5% 미만, 5~30%, 30~60%, 60% 이상의 4개 구간으로 나누어 각각 부분 가려짐부터 완전한 가려짐까지 가시성 클래스를 세분화하여 학습을 수행하였다.

3. Rule-based Tracking Post-Processing Algorithm

앞선 기법들을 적용함으로써 추적 모델은 타 도메인에 대한 강건성을 확보함과 동시에 타겟의 가시성을 자체적으로 판단할 수 있게 되었다. 일반적인 짐벌 기반 추적 시스템 환경은 카메라 제어를 통해 타겟 객체가 영상 프레임의 중앙부에 안정적으로 유지된다는 특성을 지닌다. 따라서 추적 실패로 인한 바운딩 박스 지터링은 추적 모델의 자체적인 재포착이 불가능해질 뿐만 아니라 짐벌 시스템의 오작동과 시야 상실이라는 연쇄적인 시스템 실패를 야기한다. 따라서 본 절에서는 모델의 출력값을 기반으로 추적 실패 여부를 판단하고, 이상치 바운딩 박스의 출력을 억제하여 지터링을 방지하는 규칙 기반의 후처리 알고리즘을 제안한다. 바운딩 박스의 비정상적인 변화를 감지하기 위해 다음과 같은 요소들을 중점적으로 고려한다.

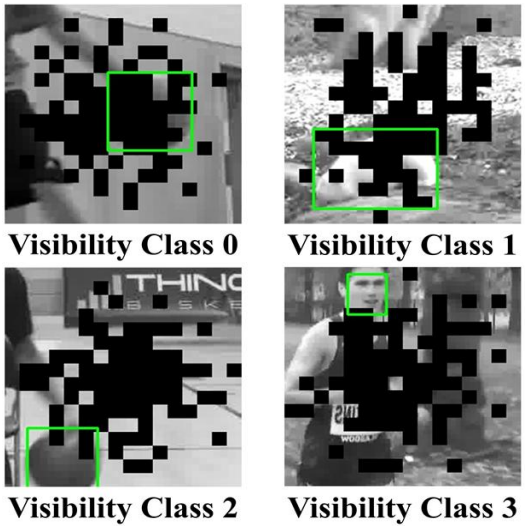

Fig. 5. Four classes of visibility labels determined by the masked pixel ratio

4. Prevention of Bounding Box Jittering via Adaptive Filtering

먼저, 객체의 최근 이동 궤적을 통계적으로 분석하여 동적 임계값(Threshold)을 산출하는 적응형 거리 필터(Adaptive Distance Filter)를 설계하였다. 이 필터의 역할은 객체 이탈 및 표류 현상으로 인한 비정상적인 위치 급변을 이상치로 식별하는 것이다. 산출된 이상치 여부는 가시성 예측 결과와 함께 최종적인 추적 실패를 판단하는 기준이 된다.

이러한 이상치를 판별하기 위하여, 통계적 이상치 탐지의 표준 방법론인 Z-score 기반의 알고리즘을 시계열 추적 환경에 맞게 슬라이딩 윈도우 방식으로 적용하였다. 구체적으로, 최근 N개의 프레임 동안 누적된 타겟 중심점의 이동 거리 데이터를 바탕으로 이동 평균(μ)과 표준편차(σ)를 계산한다. 현재 프레임에서 산출된 중심점 이동 거리를 d라고 할 때, 정상적인 이동 거리를 판별하기 위한 동적 임계값 T는 수식 (3)과 같이 정의된다.

$$T = \max(\mu + k\sigma, 2D_{\min}) \tag{3}$$

여기서 k는 Z-score의 이상치 판별 임계값에 해당하는 민감도 계수로, 값이 클수록 필터가 둔감하게 반응한다. $D_{\min}$은 카메라의 흔들림 등 미세한 노이즈로 인한 오작동을 방지하기 위한 최소 보장 거리이다.

이러한 동적 임계값과 더불어, 본 알고리즘은 앞선 기법에서 추정한 타겟 객체의 가시성 클래스를 필터 업데이트의 중요한 판단 기준으로 함께 활용한다. 계산된 현재 이

www.kci.go.kr

동 거리 d가 동적 임계값 T를 초과하는 경우뿐만 아니라, 추적 모델이 예측한 가시성 클래스가 0 또는 1로 도출되는 상황(30%이상 가려짐)에서 해당 프레임의 벡터 이동량을 비정상적인 이상치, 즉 추적 실패로 간주한다.

이처럼 이상치로 판단될 경우, 필터의 통계 데이터 오염을 막기 위해 현재 이동 거리를 누적 히스토리에 반영하지 않으며, 바운딩 박스 역시 직전 프레임의 정보를 그대로 유지한다.

5. Bounding Box Refinement using Response Map Statistics

앞서 제안된 가시성 예측 헤드는 자가 생성된 의사 라벨(Pseudo-label)을 기반으로 학습되었기 때문에, 예측의 불확실성을 내포한다. 이로인해 타겟의 특징이 여전히 관측 가능한 상황에서도, 모델이 극심한 폐색으로 오판하여 정상적인 추적을 실패로 간주하는 경우가 발생한다.

이러한 단일 지표의 한계를 보완하고 성급한 추적 실패 판정을 방지하기 위해, 추적 응답 맵의 통계적 지표인 PSR(Peak-to-Sidelobe Ratio)과 최대 예측 점수(Max Score)를 도입하였다. 또한 실시간 연산에 최적화 된 Welford 알고리즘을 적용하여 평균과 분산을 온라인으로 갱신하며, Z-score 기반의 이상치 탐지를 통해 현재 프레임의 PSR 및 Score 값이 과거의 정상적인 통계 분포에서 크게 벗어났는지를 평가한다.

해당 모듈의 핵심은 추적의 완벽성을 보장하는 것이 아니라, 명확한 추적 실패 상황만을 선별하여 보수적인 판정 기준을 완화하는 예외 처리 매커니즘으로 동작하는 것이다. 즉, 예측된 가시성 클래스가 0 또는 1이며, 응답 맵 신뢰도의 통계적 이상치가 동시에 관측될 때만 최종적인 추적 실패로 분류한다.

6. Dynamic Re-Detection Strategy

앞서 서술한 다중 검증 로직에 의해 추적 실패 상태가 장기화 될 경우를 대비하여, 본 알고리즘은 동적 재포착(Dynamic Re-detection) 전략을 포함한다. 시스템은 연속적인 추적 실패로 판정된 프레임의 누적 횟수를 지속적을 기록하며, 누적 카운트가 사전에 설정된 임계 기준을 초과할 경우 시스템이 타겟을 완전히 잃어버린 장기 추적 실패 상태로 간주한다. 이 상태에 진입하면 잘못된 탐색 영역에 고정되는 것을 막기 위해 바운딩 박스를 영상 프레임의 정중앙으로 강제 재배치한다. 이후 타겟 객체를 찾아 추적 성공 판단이 될 때까지 매 프레임 모델에 입력되는 탐색 영역의 크기를 점진적으로 확장시켜 나가며 추적 모델이 더 넓은 범위를 탐색하도록 유도한다. 이러한 일련의 과정을 통해, 제안하는 추적 파이프라인은 극단적인 환경에서도 강인하게 추적을 재개할 수 있다.

IV. Experiments

본 장에서는 제안하는 추적 모델 튜닝 기법의 유효성을 검증하기 위한 데이터 분석 및 실험 환경, 그리고 세부적인 평과 결과를 기술한다.

1. Datasets

본 연구의 도메인 일반화 성능을 검증하기 위해 다음과 같이 학습 및 평가 데이터셋을 구성하였다. 먼저, 추적 모델의 학습을 위해 다양한 실환경 RGB 이미지 시퀀스로 구성된 GOT-10k 데이터셋을 사용하였다. 해당 데이터셋은 1만개 이상의 풍부한 비디오 시퀀스와 광범위한 객체 클래스를 포함하고 있어 모델의 범용적인 특징 추출 능력을 학습하는 데 매우 적합하다. 그리고 학습된 모델의 강건성과 일반화 능력을 검증하기 위해, 3개의 테스트 데이터셋을 활용하였다. UAV123 데이터셋은 가시광선 도메인에서 항공 추적 성능을 평가하기 위한 표준 벤치마크 데이터셋이다. 무인기 플랫폼에서 촬영된 123개의 고해상도 비디오 시퀀스를 포함하며, 객체의 빠른 이동, 시점 변화, 종횡비 변화 등 실제 비행 환경에서 마주할 수 있는 도전 과제들을 포함하고 있어 가시광선 영역에서의 강건성을 정밀하게 측정할 수 있다. LSOTB-TIR 데이터셋은 열화상 카메라로 획득된 대규모 이미지 시퀀스들로 구성되어 있다. 조명 변화가 극심하거나 가시광선 영역의 정보를 활용하기 어려운 야간 환경 등에서 추적 모델이 얼마나 안정적으로 작동하는지 평가하는데 핵심적인 역할을 한다. NAT2021은 야간 환경에서 무인기 항공 추적에 특화된 데이터셋이다. 극도로 제한된 조도 및 다양한 시각적 노이즈가 존재하는 조건 하에서 모델이 타겟 객체를 성공적으로 추적하고 새로운 환경에 적응할 수 있는지를 검증한다.

AVisT 데이터셋은 악천후 및 시야 확보가 어려운 환경에 초점을 맞춘 데이터셋이다. 비, 안개, 눈과 같은 기상 악화나 화재, 연기 등 강한 시각적 방해 요소가 존재하는 다양한 시나리오를 포함하고 있어 극한의 환경에서 모델의 범용성을 테스트하는 데 사용된다.

www.kci.go.kr

Table 1. Domain generalization performance comparison based on precision score

	Model	UAV123	LSOTB-TIR	NAT2021	AVisT
(a)	Baseline	89.01%	82.44%	68.95%	51.50%
(b)	Fine-tuning w/o RandConv	88.91%	82.22%	69.69%	49.70%
(c)	Fine-tuning w/ RandConv	**89.33%**	**84.29%**	**70.54%**	**52.18%**

2. Implementation Details

본 연구에서는 메모리 구조와 자기회귀적(Autoregressive) 타겟 쿼리를 활용하는 최신 추적 모델인 AQATrack[14]을 베이스라인 모델로 선정하였다. 도메인 일반화 평가 프로토콜에 따라 타겟 도메인 데이터는 학습 과정에서 일절 사용하지 않고 오직 소스 도메인인 GOT-10k 데이터셋만을 활용하여 학습을 진행하였다. 모델의 가중치는 대규모 RGB 데이터로 사전 학습된 AQATrack-256 모델의 가중치로 초기화하였으며, 총 20 에폭(epochs=20) 동안 파인튜닝을 수행하였다. 학습 안정성을 위해 파인튜닝 시의 학습률(Learning Rate)은 기존 학습의 1/4 수준인 1e-4로 설정하였다. 또한, RandConv 모듈 적용 시 원본 특징 맵과 변환된 특징 맵 간의 특징 혼합 비율을 결정하는 하이퍼파라미터 α는 0.3으로 설정하였다. 추가적으로 콕스 과정 기반의 랜덤 마스킹 기법은 256×256 크기의 입력 영역에 적용된다. 마스킹 블록의 크기는 16×16으로 지정하였으며, 전체 면적 대비 목표 마스킹 비율은 0.5로 고정하였다. 마스킹 추출을 위한 분포의 스케일 파라미터는 16으로 설정하여 중앙 영역에 집중적인 패턴이 형성되도록 하였다.

제안하는 규칙 기반 후처리 알고리즘의 세부 하이퍼파라미터는 다음과 같이 설정되었다. 먼저 적응형 거리 필터의 슬라이딩 윈도우 크기 N은 100, 민감도 계수 k는 4.0, 최소 보장거리 $D_{\min}$은 1.0으로 설정하였다. 또한 Z-score 기반의 응답 맵 신뢰도 검증 모듈에서 PSR과 최대 예측 점수의 이상치 판별 임계값은 각각 1.2와 3.5로 경험적으로 설정하였다. 이 두 가지 이상치 판별 모듈은 추적 시작 후 초기 100프레임 동안 웜업(Warm-up) 단계를 거치며 누적 데이터의 통계적 분포를 안정화한다. 마지막으로 동적 재포착 전략을 가동하기 위한 장기 추적 실패 기준은 연속 20프레임 동안 추적 실패가 발생한 경우로 정의하였다. 모든 학습과 성능 평가 및 초당 프레임 수(FPS) 측정은 단일 NVIDIA RTX 4080 Laptop GPU 환경에서 진행되었다.

3. Evaluation of Domain Generalization

Table 1은 적용한 도메인 일반화 기법의 유효성을 검증하기 위해, 대규모 RGB 데이터로 사전 학습된 베이스라인 (a), 회색조 파인튜닝 모델 (b), 그리고 RandConv 기반 최종 모델 (c)의 정밀도(Precision Score)를 4개의 이종 도메인 환경에서 비교한 결과이다. 먼저, 소스 도메인과 시각적 특성이 유사한 UAV123 데이터셋에서 베이스라인 (a)는 회색조 입력에 따른 모달리티 갭에도 불구하고 89.01%를 기록했다. 회색조 영상으로 직접 파인튜닝을 진행한 모델 (b)는 88.91%로 소폭 하락했으나, RandConv를 결합한 모델 (c)는 89.33%를 달성하였다. 이는 모델 (b)가 파인튜닝 과정에서 기존 대규모 RGB 사전학습으로 얻었던 범용적 특징 추출 능력을 일부 상실했음을 시사하며, RandConv 기법 도입이 데이터 증강을 통한 모델의 정규화에 기여하여 강건성이 향상됨을 보여준다.

가시광선 정보가 완전히 배제되고 열방사 실루엣만으로 객체가 표현되는 LSOTB-TIR 데이터셋에서 모델 (b)는 82.22%로 베이스라인 대비 다소 하락했다. 이는 단순히 그레이스케일 이미지로 학습하는 것만으로는 이질적인 열화상 실루엣 도메인으로의 전이를 보장할 수 없음을 의미한다. 반면 모델 (c)는 84.29%를 기록하며 베이스라인 대비 +1.85%의 뚜렷한 성능 향상을 이뤄냈다. 이는 RandConv를 통한 텍스처 교란이 가시광선-열화상 센서 두 모달리티 간의 시각적 간극을 효과적으로 극복했음을 보여준다.

형태적 단서가 희박한 NAT2021 데이터셋에서는 모델 (b)가 69.69%를 기록하며 예외적으로 베이스라인(68.95%)보다 우수한 성능을 보였다. 이에 더해 모델 (c)는 70.54%를 달성하며 야간 환경에서도 최고 성능을 기록하였다.

비, 안개 등 시각적 방해 요소가 극심한 AVisT 데이터셋에서 베이스라인 (a)는 51.50%를 기록한 반면, 모델 (b)는 49.70%로 전 도메인 중 가장 큰 하락세(-1.80%)를 보였다. 반면 모델 (c)는 52.18%를 달성하며 하락폭을 완전 복구함과 동시에 베이스라인의 성능을 뛰어넘었다.

평가 결과를 종합하면, 단순 회색조 파인튜닝 (b)의 한계점과 RandConv 기법 (c)의 간단하지만 강력한 정규화

www.kci.go.kr

Fig. 6. Qualitative tracking results comparing the baseline and the proposed model – Sequence (a)

Fig. 7. Qualitative tracking results comparing the baseline and the proposed model – Sequence (b)

효과를 명확히 확인할 수 있다. 평가 환경이 그레이스케일 이미지로 통일되었음에도 파인튜닝 적용 시, 저조도 환경 NAT2021을 제외하고는 모두 성능 하락을 겪었다. 이는 단순히 동일한 파인튜닝이 모든 복합적인 악조건에 대해 일관되고 범용적인 성능 향상을 보장하지는 않는다. 반면 RandConv를 추가한 기법 (c)는 4개의 모든 도메인에서 예외 없이 일관된 최고 성능을 달성하였다. 무작위 가중치 필터는 입력 이미지의 텍스처와 명도 분포를 지속적으로 교란함으로써, 모델은 조도나 환경에 따라 쉽게 변질되는 표면적 질감에 의존하지 않게 된다. 결과적으로 도메인이 극단적으로 변하거나 심각한 명암비 훼손이 발생하더라도 객체의 본질적인 도메인 불변적 구조에 집중하도록 학습을 강제하여, 교차 도메인 추적 환경에서 탁월한 강건성을 확보할 수 있음을 확인하였다.

4. Performance Evaluation of Occlusion Awareness and Re-Detection Algorithm

본 절에서는 객체 추적 중 발생하는 가림(Occlusion) 등의 추적 실패 상황에서 제안하는 알고리즘의 효용성을 검증한다. 이를 위해 Fig. 6-8과 같이 추적 알고리즘의 정성적 평가 그리고, 추적 결과 바운딩 박스와 정답 박스(Ground Truth)간의 IoU(Intersection over Union) 점수 변화를 프레임 단위로 비교 분석하였다. 실험은 열화상 데이터인 LSOTB-TIR에서 가림 현상이 빈번하게 발생하는 두 가지 시퀀스에 대해 수행되었으며,

Fig. 8-(a)는 첫 번째 시퀀스에서 프레임 별 IoU 비교 그래프를 나타낸다. 추적 초기부터 약 600프레임 구간까지는 알고리즘 적용 전후 모두 GT와 0.8 이상의 높은 IoU를 유지하며 안정적인 추적 성능을 보였다. 그러나 약 650 프레임 부근에서 객체가 완전히 가려지면서 두 모델 모두

www.kci.go.kr

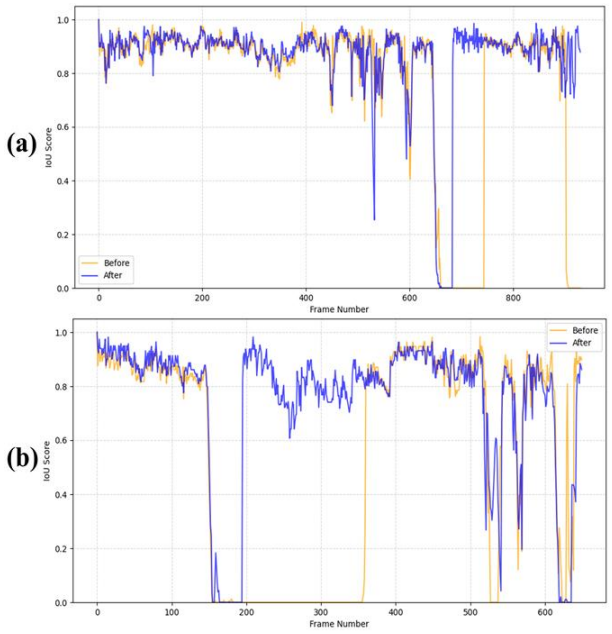

Fig. 8. IoU comparison between the baseline and the proposed model (Orange:Baseline, Blue:Ours)

IoU가 0으로 급락하는 현상이 발생한다. 이때 제안하는 모델은 추적 실패라 판단하고, 짐벌의 안정성을 위해 가장 최근의 추적 성공 프레임의 바운딩 박스 결과를 유지한다. 이후 객체가 다시 시야에 나타나는 약 680프레임 부근에서 뚜렷한 성능 차이가 확인된다. 제안하는 알고리즘을 적용한 후 파란색 실선은 객체가 재등장함과 동시에 신속하게 타겟을 재포착하여 즉시 0.9 수준의 높은 IoU를 회복하였다. 반면 알고리즘 적용 전 주황색 실선의 경우 약 740프레임에 도달할 때까지 객체를 다시 추적하지 못하고, IoU가 0에 머무르며 장기적인 추적 실패로 이어짐을 확인할 수 있다. 또한 해당 시퀀스의 후반부인 900프레임 부근을 살펴보면, 타겟 주변에 유사 객체가 등장했을 때, 베이스라인 모델의 바운딩 박스가 유사 객체로 표류하는 추적 실패 현상을 보였다. 반면 제안하는 모델은 원래 타겟을 안정적으로 추적하여 IoU의 급격한 하락 없이 높은 수치를 일관되게 유지하였다.

두 번째 시퀀스 결과인 Fig. 8-(b)에서 제안하는 재포착 알고리즘의 효과가 더욱 극명하게 나타난다. 추적 초기인 약 150프레임부터 190프레임 구간에서 심각한 가림 현상으로 인해 두 기법 모두 IoU가 0으로 하락한다. 가려짐이 발생하는 동안 제안 모델은 추적 실패 상황을 인지하고 바운딩 박스의 비정상적인 이동을 억제하였고, 타겟이 다시 나타났을 때 신속하게 재포착함을 확인할 수 있다. 반면 베이스라인 모델의 경우, 한번 타겟을 상실한 이후 지터링과 표류 현상으로 인해 약 170프레임 동안 객체를 찾지 못하는 모습을 볼 수 있다. 결과적으로 기존 추적 모델 자체는 재포착 능력이 있지만, 객체가 완전히 가려지는 상황이나, 잘못된 객체를 잡게 될 경우 복구하기까지 어려움을 보이며, 이는 전체 추적 시스템의 신뢰도 저하로 직결된다. 반면 제안하는 가림 판단 및 재포착 알고리즘을 적용할 경우, 가림 발생 여부 및 결과 바운딩 박스의 큰 변동성을 억제하여, 객체를 빠르게 재포착할 수 있다. 이를 통해 심각한 추적 실패를 방지하고 추적의 강건성과 연속성이 크게 향상됨을 실험적으로 입증하였다.

Table 2. Comparison of inference FPS and latency

Model	GPU	FPS	Latency (ms)
Baseline	RTX 4080	95.4	10.5
Ours	RTX 4080	77.5	12.9
	RTX A4500	45.7	21.9
	Jetson Orin	30.3	33.0

5. Inference Speed Analysis

제안하는 알고리즘이 실제 산업 현장, 특히 컴퓨팅 자원이 제한적인 온디바이스 짐벌 시스템에 탑재 가능한지를 검증하기 위해 추론 속도(FPS) 비교 실험을 진행하였다. 실험은 단일 NVIDIA RTX 4080 Laptop GPU, NVIDIA RTX A4500 Embedded GPU, Jetson AGX Orin 64GB Developer kit 환경에서 수행되었으며, 모델의 추론 및 후처리 연산 시간을 측정하여 평균 FPS를 산출하였다.

실험 결과 후처리 알고리즘을 적용하지 않은 모델은 RTX 4080 환경에서 평균 95.4 FPS를 기록하였으며, 가시성 헤드와 모든 규칙 기반 후처리 모듈이 결합된 제안 모델은 평균 77.5 FPS를 달성하였다. 제안 모델은 추가적인 후처리 판단 로직이 도입되었음에도 베이스라인 대비 약 2.4ms의 연산 오버헤드만이 추가되며, 높은 연산 효율성을 보여주었다. 또한 RTX A4500 Embedded 환경에서는 평균 45.7 FPS를, Jetson AGX Orin에서는 통해 30.3 FPS를 달성하며, 온디바이스 환경에서도 충분한 실시간 처리가 가능함을 입증하였다.

결론적으로, 본 연구에서 제안하는 파인튜닝 및 후처리 파이프라인은 연산량의 희생 없이도 극단적인 교차 도메인 환경에서의 추적 강건성을 비약적으로 향상시킨다. 이는 복잡한 타겟 도메인 재학습이나 무거운 모델 구조의 변경 없이도, 효율적이고 간단한 튜닝만으로 실시간 온디바이스 추적 시스템을 구축할 수 있는 매우 실용적이 방법론임을 시사한다.

www.kci.go.kr

V. Conclusions

본 논문은 실제 환경에서 빈번하게 발생하는 가려짐 상황과 미지의 도메인 변화 속에서도 안정적으로 객체를 추적할 수 있는 강건한 단일 객체 추적 모델의 튜닝 방법을 제안하였다. 우선, 타겟 도메인에 대한 사전 정보 없이도 모델이 환경 변화에 적응할 수 있도록 도메인 일반화 기법 RandConv를 도입하여 다중 도메인 변화에 대한 본질적인 식별력을 확보하였다. 또한, 모델이 스스로 가려짐 상황을 인지할 수 있도록 콕스 과정 기반 랜덤 마스킹 및 의사 라벨을 활용해 가림 판단 헤드를 학습시켰다. 이와 더불어, 해당 헤드의 예측값과 결합하여 추적 실패 여부를 정밀하게 진단하는 규칙 기반의 후처리 알고리즘을 설계하였다. 이 알고리즘은 Z-score를 기반으로 바운딩 박스 중심점 위치 변화의 이상치를 탐지하는 적응형 필터를 적용하고, 응답 맵의 통계량인 PSR과 최대 점수의 이상치 역시 Z-score로 함께 평가함으로써 딥러닝 기반 가려짐 예측값이 가질 수 있는 불확실성을 효과적으로 보완하였다. 실험 결과 제안된 기법은 다양한 벤치마크에서 베이스라인 대비 강건성과 추적 성공률을 보였다. 특히 정성적 평가를 통해 가려짐이나 유사 객체 간섭이 발생하는 복잡한 시나리오에서도 안정적인 결과를 출력함을 확인하였으며, 이를 바탕으로 실제 짐벌 기반의 추적 시스템에서도 강건하게 동작할 수 있음을 입증하였다. 이러한 복합적인 동적 후처리 모듈의 도입으로 인한 추가 연산 오버헤드는 프레임당 약 2.4ms 수준에 불과하여 RTX A4500 Embedded 환경에서 77.5 FPS의 효율적인 실시간 추론 속도를 유지할 수 있었다. 또한 대표적인 온디바이스 환경인 Jetson AGX Orin에서도 30.3 FPS의 실시간 연산이 가능함을 입증하였다. 결론적으로 본 연구는 제한된 온디바이스 자원 환경에서도 시스템의 실시간성을 저해하지 않으며, 신뢰도 높은 추적 시스템을 구축할 수 있는 실용적이고 효율적인 고도화 방안을 제시하였다.

ACKNOWLEDGEMENT

This research was supported by the Korea Research Institute for Defense Technology (KRIT), funded by the Defense Acquisition Program Administration (DAPA), Republic of Korea, under Grant No. KRIT-CT- 22-073, in 2026.

www.kci.go.kr

REFERENCES

[1] Z. Xu, D. Liu, J. Yang, C. Raffel and M. Niethammer, “Robust and generalizable visual representation learning via random convolutions,” arXiv preprint arXiv:2007.13003, 2020, doi:10.48550/arXiv.2007.13003

[2] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi and P. H. S. Torr, “Fully-convolutional siamese networks for object tracking,” in Proc. of European Conf. on Computer Vision (ECCV), pp. 850-865, Cham, Switzerland, October 2016, doi:10.1007/978-3-319-48881-3_56.

[3] B. Li, J. Yan, W. Wu, Z. Zhu and X. Hu, "High Performance Visual Tracking with Siamese Region Proposal Network," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 8971-8980, doi: 10.1109/CVPR.2018.00935.

[4] X. Chen et al., "Transformer Tracking," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 8122-8131, doi: 10.1109/CVPR46437.2021.00803.

[5] B. Yan, H. Peng, J. Fu, D. Wang and H. Lu, "Learning Spatio-Temporal Transformer for Visual Tracking," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 10428-10437, doi: 10.1109/ICCV48922.2021.01028.

[6] Y. Cui, C. Jiang, G. Wu and L. Wang, "MixFormer: End-to-End Tracking With Iterative Mixed Attention," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 6, pp. 4129-4146, June 2024, doi: 10.1109/TPAMI.2024.3349519.

[7] X. Wei, Y. Bai, Y. Zheng, D. Shi and Y. Gong, "Autoregressive Visual Tracking," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 9697-9706, doi: 10.1109/CVPR52729.2023.00935.

[8] G. Bhat, M. Danelljan, L. Van Gool and R. Timofte, “Know your surroundings: Exploiting scene information for object tracking,” in Proc. of European Conf. on Computer Vision (ECCV), pp. 205-221, Cham, Switzerland, August 2020, doi:10.48550/arXiv.2003.11014.

[9] B. Ye, H. Chang, B. Ma, S. Shan and X. Chen, “Joint feature learning and relation modeling for tracking: A one-stream framework,” in Proc. of European Conf. on Computer Vision (ECCV), pp. 341-357, Cham, Switzerland, October 2022, doi:10.48550/arXiv.2203.11991.

[10] Y. Wu et al., "Learning Occlusion-Robust Vision Transformers for Real-Time UAV Tracking," 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2025, pp. 17103-17113, doi: 10.1109/CVPR52734.2025.01594.

[11] K. Muandet, D. Balduzzi and B. Schölkopf, “Domain

generalization via invariant feature representation," in Proc. of International Conf. on Machine Learning (ICML), pp. 10-18, February 2013, doi: 10.48550/arXiv.1301.2115.

[12] F. M. Carlucci, A. D'Innocente, S. Bucci, B. Caputo and T. Tommasi, "Domain Generalization by Solving Jigsaw Puzzles," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 2224-2233, doi: 10.1109/CVPR.2019.00233.

[13] K. Zhou, Y. Yang, Y. Qiao and T. Xiang, "Domain generalization with mixstyle," arXiv preprint arXiv:2104.02008, 2021, doi: 10.48550/arXiv.2104.02008.

[14] J. Xie et al., "Autoregressive Queries for Adaptive Tracking with Spatio-Temporal Transformers," 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2024, pp. 19300-19309, doi: 10.1109/CVPR52733.2024.01826.

Authors

Ji-Su Yoon received the M.S. degree from the Department of Artificial Intelligence In, Hanyang University, Seoul, South Korea, in 2025. She is currently working with the EO/IR Systems Research and Development Laboratory, LIG Defense&Aerospace, Pangyo, South Korea. Her research interests include domain adaptation, object detection, object tracking in EO/IR systems.

Jae-Yeob Kim received the M.S. degree from the Department of Electronic Engineering, Hanyang University, Seoul, South Korea, in 2022. He is currently working with the EO/IR Systems Research and Development Laboratory, LIG Defense&Aerospace, Pangyo, South Korea. His research interests include image deblurring, super-resolution, and object detection, object tracking in EO/IR systems.

Hyun-Sung Jang received the B.S. and M.S. degrees in electrical and computer engineering from Pusan National University, Busan, South Korea, in 2007 and 2009, respectively. He is currently pursuing the Ph.D. degree in electrical and electronic engineering with Yonsei University, Seoul, South Korea. He is a Chief Research Engineer with the EO/IR Systems Research and Development Laboratory, LIG Defense&Aerospace, Pangyo, South Korea. His research interests include computer vision and machine learning.

Seung-Wook Park received the B.S. degree in electrical and electronic engineering from Korea University, Seoul, South Korea, in 2007. He is currently with the EO/IR Systems Research and Development Laboratory, LIG Defense&Aerospace, Pangyo, South Korea. His research interests include EO/IR systems and large-aperture telescope systems.

www.kci.go.kr