



일반논문 (Regular Paper)

방송공학회논문지 제31권 제4호, 2026년 7월 (JBE Vol.31, No.4, July 2026)

<https://doi.org/10.5909/JBE.2026.31.4.765>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

복잡한 상황에 대응 가능한 무인기 시각 객체 추적: 동적 가중치 제어 기반의 상관 필터 추적기 및 CNN 결합 확장

이정우^{a)}, 고대일^{b)}, 권구용^{c)}, 황재기^{c)}, 장현성^{c)}, 하남구^{c)}, 김선옥^{b)}, 김필은^{d)†}

Visual Object Tracking for UAV under Complex Scenarios: Dynamic Weight Control-based Correlation Filter Tracker and CNN-based Extension

Jeongwoo Lee^{a)}, Daeil Ko^{b)}, Kuyong Kwon^{c)}, Jaegi Hwang^{c)}, Hyunsung Jang^{c)}, Namkoo Ha^{c)}, Sunok Kim^{b)}, and Pileun Kim^{d)†}

요약

무인 항공기(UAV) 기반 시각 객체 추적(VOT) 시스템은 제한된 임베디드 자원에서 높은 추적 정확도와 강건성을 동시에 만족해야 한다. 상관 필터(CF) 기반 추적기는 푸리에 도메인 연산을 통해 낮은 연산 복잡도를 제공하지만 조명 변화, 표적의 급격한 외형 변화 등 복잡한 상황에서 필터 오염에 의한 모델 drift 현상에 취약하다. 딥러닝 기반 추적기는 강건한 의미론적 판별력을 제공하지만 연산 비용이 크다는 상호 보완적 한계를 갖는다. 이를 극복하기 위해 본 논문에서는 CF 추적기인 BiCF의 목적함수에 양방향 불일치 오류 기반의 동적 2차 시간 가중치 정규화 항을 도입한 BAWCF와 이를 CNN의 의미론적 특징 및 NLM 전처리와 결합하여 정확도를 우선시하는 운용 환경에 적합하도록 확장한 BAWCF_CN을 제안한다. 제안한 non-convex 목적함수는 ADMM을 적용하여 닫힌 형식 해를 갖는 네 개의 subproblem으로 분해된다. UAV123@10fps 및 DTB70 벤치마크의 속성별 정량 비교에서, 제안 모델은 BiCF 대비 향상된 추적 성능을 보였다.

Abstract

Visual object tracking (VOT) for unmanned aerial vehicles (UAVs) must simultaneously guarantee tracking accuracy and robustness under constrained computational resources. Correlation filter (CF)-based trackers achieve low computational complexity via Fourier-domain operations but are prone to model drift under abrupt appearance changes. Deep-learning-based trackers offer strong semantic discrimination at the cost of high computational overhead. To address these complementary limitations, this paper proposes two trackers built upon the BiCF baseline. The Bidirectional Adaptive Weight Correlation Filter (BAWCF) introduces a dynamic second-order temporal weight regularization term modulated by the bidirectional incongruity error, enabling filter inertia to adapt to instantaneous tracking reliability. BAWCF uses only hand-crafted features and serves as a CPU-friendly baseline for resource-constrained environments. The BAWCF_CN extends this framework by fusing pre-trained CNN semantic features with Non-Local Means(NLM)-denoised hand-crafted features, targeting accuracy-oriented applications with sufficient computational resources. The resulting non-convex objective is decoupled into four closed-form subproblems via the Alternating Direction Method of Multipliers (ADMM). Experiments on the UAV123@10fps and DTB70 benchmarks demonstrate consistent improvements over BiCF in both precision and success-rate metrics across challenging attributes.

Keyword : Visual Object Tracking (VOT), Discriminative Correlation Filter (DCF), ADMM Optimization, Dynamic Second-order Weight Regularization, CNN Feature Fusion

Copyright © 2026 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

1. 서론

무인 항공기(Unmanned Aerial Vehicle, UAV) 기술의 급격한 발전에 따라 재난 구조, 정찰, 감시, 물류 등 다양한 분야로 활용 범위가 확대되고 있으며, 이에 따라 UAV의 탑재 카메라로부터 관심 표적을 지속적으로 식별·추적하는 시각 객체 추적(Visual Object Tracking, VOT) 기술의 중요성이 부각되고 있다. 특히 UAV 플랫폼은 배터리와 탑재 중량의 제약으로 고성능 연산 장치 탑재가 어려운 임베디드 환경이므로, 낮은 연산 복잡도와 함께 조명 변화, 표적의 급격한 외형 변화 등 복잡한 상황에서도 안정적으로 표적 추적이 가능한 강건성이 동시에 보장되어야 한다.

이러한 요구를 충족하기 위한 두 가지 대표적인 패러다임은 상관 필터(Correlation Filter, CF) 기반 추적기와 딥러닝 기반 추적기이다. CF 기반 추적기는 푸리에 도메인에서 합성곱을 요소별 곱셈으로 치환하여 $O(N \log N)$ 연산 복잡도를 달성하므로 경량 환경에서도 안정적인 추적이 가능하다는 장점을 가지나, 복잡한 상황 발생 시 학습 필터가 오염되어 추적 위치가 점차 이탈하는 모델 drift 현상에 노출된다. 반면 딥러닝 기반 추적기는 사전학습된 합성곱 신경망(CNN)으로부터 의미론적(Semantic) 특징을 추출하여 강한 판별력을 가지나, 순전파에 따른 연산량 과부하로 자원이 제한된 환경에서는 적용이 어렵다는 상호 보완적 한계

를 가진다.

본 논문은 두 패러다임의 장·단점을 각각 고려하여, CF 추적기의 연산 효율성과 CNN의 강건성을 목적에 따라 선택적으로 활용할 수 있는 이중 구성 프레임워크를 제안한다.

먼저, CF 추적기의 모델 drift 문제를 완화하기 위해 양방향 일관성을 강제하는 BiCF(Bidirectional Incongruity-aware Correlation Filter)의 목적함수에 양방향 불일치 오류(Bidirectional Incongruity Error) 기반의 동적 2차 시간 가중치 정규화(Dynamic Second-Order Weight Regularization) 항을 결합한 BAWCF(Bidirectional Adaptive Weight Correlation Filter)를 제안한다. BAWCF는 hand-crafted feature만을 사용하는 CPU 친화적 기본 모델로, 연산 자원이 제한된 임베디드 환경에서의 강건한 추적을 목표로 설계되었다.

여기에 추가로, CNN의 의미론적 특징과 NLM(Non-Local Means) 기반 노이즈 억제 전처리를 결합한 확장 모델 BAWCF_CN을 제시한다. BAWCF_CN은 CNN 순전파에 따른 추가 연산 비용을 수반하므로, GPU 등 고성능 연산 자원이 가용한 환경 또는 추적 정확도를 최우선으로 하는 운용 환경에 적합하도록 설계되었다. 제안된 목적함수는 비볼록(non-convex) 형태이므로 ADMM(Alternating Direction Method of Multipliers) 최적화를 적용하여 닫힌 형식 해를 갖는 네 개의 subproblem으로 분해함으로써 복잡한 목적함수를 최적화하였다.

본 논문의 주요 기여는 다음과 같다. 첫째, BiCF의 목적함수에 양방향 불일치 오류 ϵ 를 입력으로 하는 동적 2차 시간 가중치 정규화 항을 도입하고, 정규화 계수 α 를 매 프레임 적응적으로 재추정하는 제어 메커니즘을 제안한다. 이는 BiCF와 같은 기존 모델들이 정규화 강도를 고정된 하이퍼파라미터로 사용하여 외형 변화 강도에 따른 모델 관성과 적응 속도를 능동적으로 조절하지 못하던 한계를 해소한다. 둘째, 제안한 non-convex 목적함수를 필터 f , 공간 가중치 w 및 보조변수 g, q 로 분리하는 ADMM을 유도하여, 닫힌 형식 해(closed-form solution)를 갖는 네 개의 subproblem으로 분해함으로써 효율적으로 최적화한다. 셋째, 정확도를 우선시하는 운용 환경을 위한 보완적 확장 모델 BAWCF_CN을 제시한다. 이는 사전학습된 CNN(VGG-

a) 한국항공대학교 항공전자정보공학부(Department of Electronics and Information Engineering, Korea Aerospace University)

b) 한국항공대학교 인공지능학과(Department of Artificial Intelligence, Korea Aerospace University)

c) LIG 디엔에이 광정보융합연구소(EO/IR Systems R&D, LIG Defense & Aerospace)

d) 한국항공대학교 AI자율주행시스템공학과(Department of Autonomous Vehicle Engineering, Korea Aerospace University)

‡ Corresponding Author : 김필은(Pileun Kim)

E-mail: pkim@kau.ac.kr

Tel: +82-2-300-0173

ORCID: <https://orcid.org/0000-0001-8750-4524>

※ 본 연구는 2025년 LIG D&A (주)의 용역과제 “실시간 표적추적 최적화” 지원을 받아 수행된 연구임.

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원·학석사연계 ICT핵심인재양성 지원을 받아 수행된 연구임(IITP-2024-RS-2024-00437857).

· Manuscript received June 1, 2026; Revised July 7, 2026; Accepted July 7, 2026.

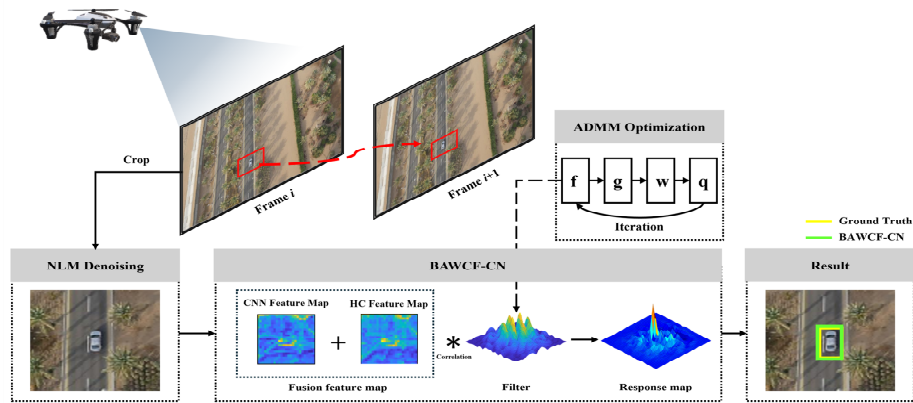


그림 1. 제안하는 추적기의 전체 플로우차트
Fig. 1. Overall Flowchart of the Proposed Tracker

M-2048)의 의미론적 특징과 NLM 전처리로 안정화된 hand-crafted features를 결합하여 복잡한 배경에서의 판별력을 강화한다. 마지막으로, UAV123@10fps 및 DTB70 벤치마크의 속성별 비교를 통해 BiCF 베이스라인 대비 성능 향상을 입증한다.

본 논문의 구성은 다음과 같다. II장에서 관련 연구를, III장에서 제안 방법을, IV장에서 실험 결과를, V장에서 결론을 기술한다.

II. 관련 연구

1. DCF 기반 시각 객체 추적기

DCF(Discriminative Correlation Filter) 계열 추적기는 합성곱을 푸리에 도메인의 요소별 곱셈으로 환원하여 높은 연산 효율성을 달성한다. 본 절에서는 베이스라인 추적기 선정 근거를 명확히 하기 위해, 기존 DCF 추적기 발전 과정을 경계효과(boundary effect) 완화를 위한 공간 정규화 축, 모델 drift 억제를 위한 시간 정규화 축, 양방향 일관성 기반 자기 검증 축의 세 가지 관점에서 정리한다.

공간 정규화 축에서는 Bolme 등의 MOSSE^[1]가 최초로 합성곱 응답과 가우시안 형태의 목표 응답 간 MSE를 푸리에 도메인에서 최소화하는 방식을 최초로 제시하였으나, 학습 패치를 순환 행렬로 가정함에 따른 경계효과 문제를

안고 있었다. 이를 해결하기 위한 일련의 후속 연구로서 Danelljan 등의 SRDCF^[2]는 패치 경계로 갈수록 필터 가중치를 억제하는 공간 정규화를 도입하였고, Dai 등의 ASRCF^[3]는 표적 형상에 따라 공간 가중치 자체를 적응적으로 학습함으로써 정적 패널티의 한계를 보완하였다.

시간 정규화 축에서는 Li 등의 STRCF^[4]가 인접 프레임 간 필터 변화량 $\|f^t - f^{t-1}\|_2$ 을 직접 제약하는 1차 시간 정규화 항을 도입함으로써 급격한 모델 변화에 따른 drift를 완화하였다. 이러한 1차 차분 제약은 이후 2차 차분 형태로 확장되었는데, Zhang 등의 CACF^[5]는 인접 세 프레임 간 채널별 응답의 2차 차분을 활용해 변화의 가속도 정보를 반영하고 퍼지 C-평균 기반 채널 가중을 결합하였으며, Yu 등의 LDCF^[6]는 공간 정규화 가중치 자체에 2차 차분 형식을 적용하여 UAV의 불규칙한 기동에 대한 안정성을 입증하였다.

양방향 일관성 축에서는 Lin 등의 BiCF^[7]가 전방·후방 추적 응답의 불일치 오류를 직접 목적함수에 포함함으로써, 단방향 학습이 갖는 시간 비대칭 편향과 과적합 문제를 동시에 해결하였다. 최근에는 Wang 등의 FWRDCF^[8]과 같이 ADMM 기반 효율적 최적화를 활용한 적응형 정규화 계열의 연구가 활발히 진행되며 ARC(Aspect Ratio Change) 변화 등 어려운 속성에서의 강건성을 추가로 끌어올리고 있다.

이러한 연구 흐름 속에서 본 논문은 BiCF를 베이스라인

모델로 채택한다. BiCF의 목적함수는 양방향 불일치 오류 항과 1차 시간 정규화 형태를 취하고 있어, 본 논문에서 제안하는 동적 2차 시간 정규화 항으로 확장하기에 용이하다. 다만, BiCF 및 그 후속 모델들은 정규화 강도를 고정된 하이퍼파라미터로 사용하므로 외형 변화 강도에 따른 모델 관성 및 적응 속도의 능동적 조절이 불가능하다는 한계를 갖는다. 이에 본 논문에서는 양방향 불일치 오류 ε 를 입력으로 받아 정규화 계수 α 를 매 프레임 재추정하는 동적 메커니즘을 통해 이를 해소한다.

2. CNN 의미론적 특징과 상관 필터의 결합

DCF 기반 추적기에서 사용하는 특징은 hand-crafted feature(HC)와 CNN 기반 의미론적 특징으로 구분된다. HC 특징은 HOG, Color Names(CN), Grayscale, Intensity Contrast(IC) 등 사람이 설계한 알고리즘으로 엣지(Edge), 색상(Color), 질감(Texture) 등의 정보를 추출하며 연산량이 적어 원본 해상도를 유지하면서 효율적인 처리가 가능하다. 그러나 저수준(low-level) 통계량에 의존하므로 심한 형태 변형이나 배경이 복잡한 장면에서는 식별력이 저하된다.

CNN 기반 특징은 객체의 의미론적 정보를 내포하여 가림이나 형태 변화에 강건하지만 연산 비용이 크다는 trade-off를 가진다. CF 프레임워크와 CNN을 결합하려는 시도는 Ma 등의 HCF^[9]가 VGG-19의 다층 특징을 계층적

으로 활용한 것을 시작으로, 연속 도메인 보간으로 다중 해상도 특징을 통합한 C-COT^[10], ECO^[11], SRDCF의 공간 정규화에 단일 CNN 층을 결합한 DeepSRDCF^[12] 등의 연구로 발전해 왔다.

그러나 기존 연구는 CNN의 Spatial Stride에 의해 특징맵 해상도가 입력 대비 1/8~1/16로 감소하여 작은 표적의 위치 정밀도가 떨어지고, CNN 순전파 비용으로 인해 자원이 제한된 환경에서의 적용이 어렵다는 공통 한계를 갖는다. 이를 고려하여 본 논문에서는 CNN 결합 여부를 모델 선택으로 분리한다. BAWCF는 HC 특징만을 사용하는 기본 모델로 설계하여 연산 자원이 제한된 환경에 적합하도록 하고, BAWCF_CN은 CNN 특징을 결합하여 고성능 연산 환경에서 추적 정확도를 극대화하는 확장 모델로 제안한다.

III. 제안 방법

1. BAWCF 목적함수 설계

본 연구에서는 BiCF를 베이스라인으로 사용하였다. BiCF는 연속된 프레임 사이의 양방향 불일치 오류를 목적함수에 반영하여, 객체의 외관 변화와 배경 변화에 강건하게 대응하는 상관 필터 기반 추적기이다. 단순하면서 강건한 목적함수 구조를 가진다는 점에서 본 연구의 기반 모델로 선정하였으며 BiCF의 목적함수는 다음과 같다.

표 1. 제안 방법에 사용된 주요 기호 정의

Table 1. Definitions of the main symbols used in the proposed method

Symbol	Meaning	Symbol	Meaning
f	Correlation filter	ζ	Lagrange multiplier for the constraint $f = g$
w	Spatial regularization weight	u	Lagrange multiplier for the constraint $w = q$
g	Auxiliary variable of f	μ	Penalty parameter
q	Auxiliary variable of w	η	EMA learning rate
α	Dynamic inertia coefficient($\in [0.8, 1.0]$)	ε	Bidirectional inconsistency error
τ	Sigmoid threshold for α	k	Sigmoid slope for α
λ_1	Weight coefficient of the spatial regularization term	λ_2	Weight coefficient of the bidirectional inconsistency term
λ_3	Weight coefficient of the dynamic second-order temporal regularization term	μ_w	Penalty parameter for the constraint $w = q$
β	Penalty increase rate (> 1)	$\mu_{\max}$	Upper bound of the penalty parameter
D	Number of feature channels	$\odot$	Element-wise (Hadamard) product
x	Feature map	y	Desired Gaussian response
$*$	Correlation operation	$\sqrt{N \cdot F}$	Normalized discrete Fourier transform (DFT)

$$\varepsilon(f) = \min \|y - \sum_{d=1}^D x_d^t * f_d^t\|_2^2 + \lambda_1 \sum_{d=1}^D \|w \times f_d^t\|_2^2 + \lambda_2 \epsilon_d \quad (1)$$

$$\begin{aligned} \epsilon_d &= \|\Delta_F R - \Delta_B R\|_2^2 \\ &= \|(R_f^{x^t} - R_{f^{t-1}}^{x^t}) - (R_{f^{t-1}}^{x^{t-1}} - R_{f^t}^{x^{t-1}})\|_2^2 \\ &= \|(f^t * x^t - f^{t-1} * x^t) - (f^{t-1} * x^{t-1} - f^t * x^{t-1})\|_2^2 \\ &= \|(f_d^t - f_d^{t-1}) * (x_d^t + x_d^{t-1})\|_2^2 \end{aligned} \quad (2)$$

$$\begin{aligned} \varepsilon(f, w) &= \min \frac{1}{2} \|y - \sum_{d=1}^D x_d^t * f_d^t\|_2^2 + \frac{\lambda_1}{2} \sum_{d=1}^D \|w \odot f_d^t\|_2^2 \\ &\quad + \frac{\lambda_2}{2} \|(f_d^t - f_d^{t-1}) * (x_d^t + x_d^{t-1})\|_2^2 + \frac{\lambda_3}{2} \|(w^t - w^{t-1}) - \alpha(w^{t-1} - w^{t-2})\|_2^2 \end{aligned} \quad (3)$$

기존 BiCF의 목적함수는 응답 오차 항, 공간 정규화 항, 그리고 양방향 불일치 정규화 항으로 구성된다. 양방향 불일치 항은 현재 프레임과 이전 프레임 사이의 외관 변화 차이를 최소화하여, 급격한 조명 변화나 가려짐 상황에서도 안정적인 필터 업데이트가 가능하도록 설계되었다. 양방향 불일치 오류 ϵ_d 는 다음과 같다.

여기서 ε 는 양방향(전·후방) 추적 응답의 불일치 정도를 정량화한 스칼라 값으로, 값이 클수록 표적의 외형이 급격히 변하여 추적 신뢰도가 낮음을 의미한다. 본 논문에서는 이 ε 를 동적 관성 계수 α 의 입력으로 사용한다.

본 연구에서는 기존 BiCF 목적함수에 동적 2차 가중치 정규화 항을 추가한 BAWCF를 제안한다. 특히 STRCF의 1차 시간 정규화가 인접 두 프레임 간 필터 변화량만을 제약하는 데 반해, 본 논문의 2차 정규화 항은 가중치 변화의 추세 즉, 변화의 가속도 정보까지 반영하여 급격한 외형 변화 구간에서의 과도한 필터 변동을 억제한다. 또한 CACF와 LDECF가 2차 차분을 각각 채널 응답과 공간 정규화 가중치에 적용한 것과 달리, 본 논문은 시간 영역의 가중치 변화에 2차 정규화를 적용하고 그 강도를 양방향 불일치 오류에 따라 적응적으로 제어한다는 점에서 차별화된다. 추가된 정규화 항은 이전 프레임의 가중치 변화 추세를 현재 업데이트에 반영하여, 시간적으로 연속적인 필터 변화를 유도한다. 제안하는 목적함수는 다음과 같다.

여기서 새롭게 추가된 동적 2차 가중치 정규화 항 $(w^t - w^{t-1}) - \alpha(w^{t-1} - w^{t-2})$ 은 공간 가중치 w 의 2차 차분(변화의 가속도)을 제약하며, λ_2 는 해당 항의 가중

계수, 위첨자 t 는 프레임 인덱스를 나타낸다. α 는 직전 변화 추세의 반영 강도를 결정하는 동적 관성 계수이다. 기존의 1차 정규화 방식은 현재와 이전 상태의 차이만 고려하기 때문에 급격한 외관 변화가 발생하는 경우 필터의 변동성이 커질 수 있다. 반면 제안 방식은 과거의 변화 추세를 함께 고려함으로써 보다 안정적인 업데이트를 수행한다.

또한 본 연구에서는 관성 계수 α 를 고정값으로 사용하지 않고, 양방향 불일치 오류 ϵ 의 크기에 따라 동적으로 조절되도록 설계하였다. 양방향 불일치 오류가 작은 경우에는 객체의 외관 변화가 비교적 안정적인 상태라고 판단하여 α 를 1에 가깝게 유지하고, 이전 프레임의 가중치 변화 추세를 현재 업데이트에 적극적으로 반영한다. 반면 가려짐(occlusion), 조명 변화, Motion Blur 등으로 인해 양방향 불일치 오류가 증가하는 경우에는 α 를 감소시켜 불확실한 과거 정보의 영향을 줄이도록 하였다. 이를 통해 잘못된 방향으로의 필터 업데이트를 억제하고 drift 현상을 완화할 수 있다.

특히 α 는 오류 크기에 선형적으로 반응하지 않도록 sigmoid 기반의 비선형 함수로 설계하였다. 안정적인 추적 구간에서는 α 가 1에 가깝게 유지되다가, 오류가 임계값 τ 를 초과하는 경우 감소하도록 하여 급격한 외관 변화 상황에 민감하게 대응할 수 있도록 하였다. 본 논문에서는 추적의 시간적 연속성을 유지하기 위해 α 의 범위를 $[0.8, 1.0]$ 으로 제한하였고 τ 는 양방향 불일치 오류 ε 가 안정 구간과 불안정 구간을 구분하는 임계점으로 사용된다. 실험을 통해 ε 의 변화 양상을 경험적으로 분석한 결과, ε 가 작은 구간에서는 전·후방 추적 결과가 비교적 일관되게 유지되는 반면, ε 가

약 3 이상으로 증가하는 경우에는 가림, motion blur, 조명 변화 등으로 인해 응답 불일치가 커지고 추적 신뢰도가 저하되는 경향이 관찰되었다. 이에 따라 본 논문에서는 $\tau=3$ 으로 설정하여, ε 가 해당 값을 초과하는 경우 α 를 점진적으로 감소시킴으로써 신뢰도가 낮은 과거 변화 추세가 현재 필터 업데이트에 과도하게 반영되는 것을 억제하였다.

또한 k 는 sigmoid 함수의 기울기를 결정하는 파라미터로, ε 변화에 따른 α 의 감소 속도를 제어한다. k 가 너무 작으면 ε 가 증가해도 α 의 변화가 미미하여 급격한 외형 변화에 충분히 대응하기 어렵고, 반대로 k 가 너무 크면 α 가 급격히 감소하여 시간적 연속성이 손상될 수 있다. 본 논문에서는 k 를 0.1부터 1.0까지 0.1 간격으로 변화시키며 UAV123@10fps 데이터셋에서 추적 성능을 비교하였고, 정밀도(DP)와 성공률(AUC)이 가장 높게 나타난 $k=0.5$ 를 최종 값으로 설정하였다.

$$\alpha_t = 0.8 + 0.2 \left(1 - \frac{1}{1 + \exp(-k(\varepsilon - \tau))} \right) \quad (4)$$

2. ADMM 최적화

제안하는 BAWCF 목적함수는 필터 f 와 공간 가중치 w

가 결합된 non-convex 문제이므로 두 변수를 동시에 최적화하기 어렵다. 이를 해결하기 위해 ADMM^[13]을 적용하여 목적함수를 단일 변수에 대해 convex한 여러 개의 subproblem으로 분해한다.

먼저 필터 f 와 공간 가중치 w 에 대한 보조변수 g 와 q 를 각각 도입하고 등식 제약 $f=g$, $w=q$ 를 부과한다.

보조변수를 도입하는 이유는, 원 목적함수에서 결합되어 있던 항들을 제약을 통해 분리함으로써 각 변수에 대한 subproblem이 독립적으로 닫힌 형식 해를 갖도록 하기 위함이다.

다음으로 Parseval's theorem을 이용해 목적함수를 주파수 영역으로 변환한다.

시간 영역에서의 상관 연산은 주파수 영역에서 요소별 곱셈으로 표현되므로, FFT(Fast Fourier Transform)를 통해 $O(N \log N)$ 의 연산량으로 효율적으로 계산할 수 있다. 이후 등식 제약을 증강 라그랑지안(augmented Lagrangian) 형태로 변환하여 라그랑주 승수 ζ , u 와 페널티 파라미터 μ 를 도입한다.

ADMM은 하나의 변수만 갱신하고 나머지는 고정하는 방식으로 반복 수행된다. 목적함수를 f , g , w , q 각각에 대한 subproblem으로 분해하면 식 (8)과 같다.

$$\begin{aligned} \varepsilon(f, g, w, q) = \min & \frac{1}{2} \left\| y - \sum_{d=1}^D x_d^t * g_d^t \right\|_2^2 + \frac{\lambda_1}{2} \sum_{d=1}^D \left\| w \odot f_d^t \right\|_2^2 \\ & + \frac{\lambda_2}{2} \left\| (g_d^t - f_d^{t-1}) * (x_d^t + x_d^{t-1}) \right\|_2^2 \\ & + \frac{\lambda_3}{2} \left\| (q^t - w^{t-1}) - \alpha(w^{t-1} - w^{t-2}) \right\|_2^2 + \frac{\mu_w}{2} \left\| q - w + u \right\|_2^2 \\ \text{s.t. } & g_d^t = f_d^t, q = w, d = 1, 2, \dots, D \end{aligned} \quad (5)$$

$$\begin{aligned} \varepsilon_d(f_d, \hat{g}_d, w, q) = \min & \frac{1}{2} \left\| \hat{y} - \sum_{d=1}^D \hat{x}_d^t \odot \hat{g}_d^t \right\|_2^2 + \frac{\lambda_1}{2} \sum_{d=1}^D \left\| w \odot f_d^t \right\|_2^2 \\ & + \frac{\lambda_2}{2} \sum_{d=1}^D \left\| (\hat{g}_d^t - f_d^{t-1}) \odot (\hat{x}_d^t + \hat{x}_d^{t-1}) \right\|_2^2 \\ & + \frac{\lambda_3}{2} \left\| (q^t - w^{t-1}) - \alpha(w^{t-1} - w^{t-2}) \right\|_2^2 + \frac{\mu_w}{2} \left\| q - w + u \right\|_2^2 \\ \text{s.t. } & \hat{g}_d^t = \sqrt{N} F f_d^t, q = w, d = 1, 2, \dots, D \end{aligned} \quad (6)$$

$$L(\hat{g}_d, f_d, \hat{\zeta}_d) = \frac{\mu}{2} \left\| \hat{g}_d^t - \sqrt{N} F f_d^t + \frac{1}{\mu} \hat{\zeta}_d \right\|_2^2 + \varepsilon_d(\hat{g}_d, f_d, w) \quad (7)$$

$$\begin{cases} \mathbf{f}_d^{(i+1)} = \arg \min \frac{\lambda_1}{2} \sum_{d=1}^D \|\mathbf{w} \odot \mathbf{f}_d^t\|_2^2 + \frac{\mu}{2} \|\hat{\mathbf{g}}_d^t - \sqrt{N} \mathbf{F} \mathbf{f}_d^t + \frac{1}{\mu} \hat{\zeta}_d\|_2^2 \\ \mathbf{g}_d^{(i+1)} = \arg \min \frac{1}{2} \|\hat{\mathbf{y}} - \sum_{d=1}^D \hat{\mathbf{x}}_d^t \odot \hat{\mathbf{g}}_d^t\|_2^2 + \frac{\lambda_2}{2} \sum_{d=1}^D \|(\hat{\mathbf{g}}_d^t - \hat{\mathbf{f}}_d^{t-1}) \odot (\hat{\mathbf{x}}_d^t + \hat{\mathbf{x}}_d^{t-1})\|_2^2 \\ \quad + \frac{\mu}{2} \|\hat{\mathbf{g}}_d^t - \sqrt{N} \mathbf{F} \mathbf{f}_d^t + \frac{1}{\mu} \hat{\zeta}_d\|_2^2 \\ \mathbf{w}^{(i+1)} = \arg \min \frac{\lambda_1}{2} \sum_{d=1}^D \|\mathbf{w} \odot \mathbf{f}_d^t\|_2^2 + \frac{\mu_w}{2} \|\mathbf{q} - \mathbf{w} + \mathbf{u}\|_2^2 \\ \mathbf{q}^{(i+1)} = \arg \min \frac{\lambda_3}{2} \|(\mathbf{q}^t - \mathbf{w}^{t-1}) - \alpha(\mathbf{w}^{t-1} - \mathbf{w}^{t-2})\|_2^2 + \frac{\mu_w}{2} \|\mathbf{q} - \mathbf{w} + \mathbf{u}\|_2^2 \end{cases} \quad (8)$$

각 subproblem은 해당 변수에 대해 볼록 이차식(convex quadratic) 형태이므로, 목적함수를 해당 변수로 미분하여 그 값이 0을 만족하는 해를 닫힌 형식으로 유도할 수 있다.

감소하면 수렴한 것으로 간주하여 반복을 종료한다.

$$\mu^{(i+1)} = \min(\mu_{\max}, \beta \mu^{(i)}) \quad (12)$$

$$\begin{cases} \mathbf{f}_d^{(i+1)} = \frac{\mathbf{F}^{-1}(\mu \hat{\mathbf{g}}_d^t + \hat{\zeta}_d)}{\frac{\lambda_1}{N}(\mathbf{w} \odot \mathbf{w}) + \mu} \\ \hat{\mathbf{g}}_d^{(i+1)} = \frac{\hat{\mathbf{x}}_d^t \mathbf{y}^T + \lambda_2(\hat{\mathbf{x}}_d^t + \hat{\mathbf{x}}_d^{t-1})^2 \hat{\mathbf{f}}_d^{t-1} + \mu \hat{\mathbf{f}}_d^t - \hat{\zeta}_d}{(\hat{\mathbf{x}}_d^t)^T \hat{\mathbf{x}}_d^t + \lambda_2(\hat{\mathbf{x}}_d^t + \hat{\mathbf{x}}_d^{t-1})^2 + \mu} \\ \mathbf{w}^{(i+1)} = \frac{\mu_w(\mathbf{q} - \mathbf{u})}{\lambda_1(\mathbf{f}_d \odot \mathbf{f}_d) + \mu_w} \\ \mathbf{q}^{(i+1)} = \frac{\lambda_3(\mathbf{w}^{t-1} + \alpha(\mathbf{w}^{t-1} - \mathbf{w}^{t-2})) + \mu_w(\mathbf{w} + \mathbf{u})}{\lambda_3 + \mu_w} \end{cases} \quad (9)$$

라그랑주 승수 ζ 와 $\mathbf{u}$ 는 현재 반복 단계에서 발생한 제약 조건 오차를 누적하며, 이후 반복 과정에서 해당 오차를 줄이는 방향으로 업데이트된다.

$$\hat{\zeta}_d^{(i+1)} = \hat{\zeta}_d^{(i)} + \mu(\hat{\mathbf{g}}_d^{(i+1)} - \hat{\mathbf{f}}_d^{(i+1)}) \quad (10)$$

$$\hat{\mathbf{u}}_d^{(i+1)} = \hat{\mathbf{u}}_d^{(i)} + \mu_w(\hat{\mathbf{w}}_d^{(i+1)} - \hat{\mathbf{q}}_d^{(i+1)}) \quad (11)$$

여기서 $(\hat{\mathbf{g}}_d^{(i+1)} - \hat{\mathbf{f}}_d^{(i+1)})$, $(\hat{\mathbf{w}}_d^{(i+1)} - \hat{\mathbf{q}}_d^{(i+1)})$ 는 각각 필터 $\mathbf{f}$ 와 보조변수 $\mathbf{g}$, 공간 가중치 $\mathbf{w}$ 와 보조변수 $\mathbf{q}$ 사이의 제약 위반량을 의미한다. 반복이 진행될수록 라그랑주 승수는 이 차이를 감소시키는 방향으로 갱신된다. 또한 식 (12)에서 패널티 파라미터 μ 는 고정값 대신 반복마다 점진적으로 증가시켜, 초기에는 과도한 패널티로 인한 불안정성을 방지하고 후반에는 제약을 강하게 만족시켜 수렴을 가속한다. 원시 잔차 $\|\mathbf{f} - \mathbf{g}\|$, $\|\mathbf{w} - \mathbf{q}\|$ 가 사전 설정된 임계값 이하로

최적화가 완료된 이후에는 실시간 환경 변화에 적응하기 위해 지수 이동 평균(EMA, Exponential Moving Average) 기반의 온라인 모델 업데이트를 수행한다.

$$\hat{\mathbf{x}}_{d, \text{model}} = (1 - \eta) \hat{\mathbf{x}}_{d, \text{model}}^{t-1} + \eta \hat{\mathbf{x}}_d \quad (13)$$

여기서 η 는 학습률(learning rate)로, 이전 모델과 현재 프레임의 특징 정보를 혼합하는 비율을 결정한다. η 가 작은 경우 이전 모델의 정보가 더 크게 유지되어 안정적인 업데이트가 가능하며, 노이즈와 순간적인 외란 변화에 강건한 특성을 가진다. 반면 η 가 큰 경우 현재 프레임의 정보가 빠르게 반영되어 급격한 외란 변화나 스케일 변화에 보다 민감하게 적응할 수 있다. 이를 통해 제안한 BAWCF는 시간적으로 안정적인 모델 업데이트와 환경 변화에 대한 적응성을 동시에 확보할 수 있다.

3. NLM 기반 영상 전처리

BAWCF_CN은 정확도를 우선시하는 운용 환경을 위한 보완적 확장 모델로, 특징 추출 전 단계에서 NLM 전처리를 수행한다. NLM은 영상 내에 유사한 패턴이 반복적으로 존재한다는 자기유사성(self-similarity)에 기반한 denoising 기법이다. 인접한 픽셀만을 고려하는 Gaussian 필터와 달리, NLM은 넓은 탐색 영역(non-local) 내에서 현재 픽셀 p 주변의 패치와 유사한 형태를 가진 패치들을 찾아 가중 평

균을 수행하므로, 엣지 정보와 구조적 특징을 보존하면서 노이즈를 선택적으로 제거한다.

NLM은 세 단계로 동작한다. 먼저 픽셀 p 와 q 를 각각 중심으로 하는 이웃 패치 $B(p, f)$, $B(q, f)$ 사이의 거리를 다음과 같이 계산한다.

$$d^2(B(p, f), B(q, f)) = \frac{1}{(2f+1)^2} \times \sum_{j \in B(0, f)} (I_i(p+j) - I_i(q+j)) \quad (14)$$

여기서 f 는 이웃 영역의 크기를 결정하는 파라미터, j 는 이웃 영역 $B(0, f)$ 내의 임의의 픽셀 위치, I_i 는 I 번째 채널의 밝기 값이다. 두 패치 내 대응 픽셀들의 밝기 차이 제곱합을 패치 크기 $(2f+1)^2$ 으로 정규화한 값으로, 두 패치가 유사할수록 작은 값을 가진다. 이 패치 거리를 기반으로 가중치 $w(p, q)$ 를 다음과 같이 산출한다.

$$w(p, q) = \exp\left(-\frac{\max(d^2 - 2\sigma^2, 0)}{h^2}\right) \quad (15)$$

σ 는 영상의 노이즈 표준편차, h 는 필터링 강도를 결정하는 매개변수이다. $\max(d^2 - 2\sigma^2, 0)$ 항은 노이즈 자체에 의한 패치 거리 과대 추정을 보정하여 순수한 구조적 유사도만을 가중치에 반영하도록 한다. h 값이 크면 넓은 범위의 픽셀이 복원에 기여하여 평활화 효과가 강해지고, h 값이 작으면 매우 유사한 패치만 선택적으로 반영된다. 마지

막으로 **denoising** 된 픽셀값 $I_{\text{denoised}}(p)$ 는 탐색 영역 S 내 모든 픽셀의 정규화된 가중 평균으로 다음과 같이 계산한다.

$$I_{\text{denoised}}(p) = \frac{\sum_{q \in S} w(p, q) \cdot I(q)}{\sum_{q \in S} w(p, q)} \quad (16)$$

$w(p, q)$ 는 p 와 q 의 이웃 패치가 유사할수록 큰 값을 가지므로, 유사한 패턴을 갖는 픽셀들이 복원에 더 많이 기여한다.

본 논문에서는 필터링 매개변수 $h = 0.15$ 로 설정하였으며, 해당 값은 UAV123@10fps 데이터셋의 특성에 맞춰 실험적으로 결정하였다. NLM은 별도의 학습 과정이나 GPU 연산을 요구하지 않는 비학습(training-free) 기반 기법으로, 딥러닝 기반 **denoiser** 대비 연산 부담이 작다는 특징이 있다. 그림 2에서는 원본 이미지와 **denoising** 적용 이미지에 생성된 특징 맵 및 응답 맵을 비교한 결과를 나타낸다. 원본 이미지의 경우 응답 에너지가 주변 영역으로 분산되어 비교적 넓은 **peak** 형태를 보이는 반면, **denoising**이 적용된 경우 표적 위치에서 더욱 집중된 **peak**가 형성되는 것을 확인할 수 있다. 이는 노이즈로 인한 불필요한 응답이 감소하면서 필터가 표적의 주요 구조적 특징에 보다 집중할 수 있기 때문이며, 결과적으로 표적 위치 추정의 정확도와 추적 안정성을 향상시킨다. 이러한 응답 맵의 선명한 **peak**는 **drift** 억제와 강건성 향상에 보완적으로 기여한다.

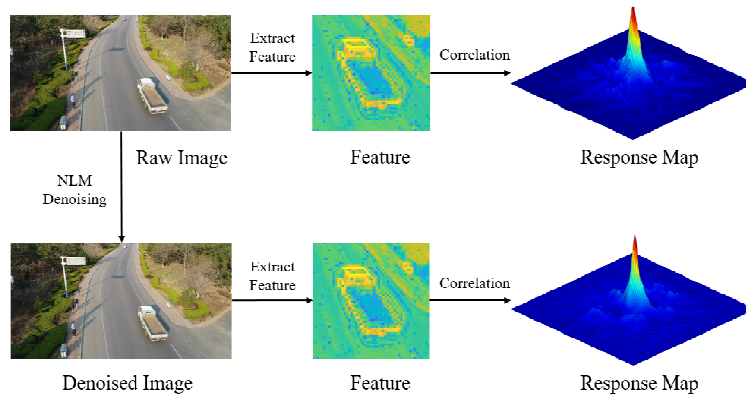


그림 2. NLM denoising 적용에 따른 특징 맵 및 응답 맵 비교

Fig. 2. Comparison of feature and response maps with and without NLM denoising

4. CNN 의미론적 특징 추출 및 결합

BAWCF는 HOG, Color Names 등 HC 특징만을 사용하므로 연산 효율은 높으나, 표적과 색상·질감이 유사한 배경이 존재하거나 표적의 외형이 의미론적으로 크게 변하는 상황에서는 판별력이 부족할 수 있다. 이를 보완하기 위해 본 논문에서는 정확도를 우선시하는 운용 환경을 위한 보완적 확장 모델 BAWCF_CN을 기술한다. BAWCF_CN은 앞서 제안한 핵심 모델 BAWCF에 HC 특징의 정밀한 위치 추정 능력과 CNN 의미론적 특징을 결합하여, 복잡한 배경에서의 판별력을 보완한다. BAWCF_CN은 GPU 등 고성능 연산 자원이 가용한 환경 또는 추적 정확도를 최우선으로 하는 운용 환경을 위한 확장 모델이다. 전체 네트워크를 재학습하지 않고 사전학습(pre-trained)된 모델을 특징 추출기로만 활용함으로써, CNN 도입에 따른 연산 비용 증가를 최소화하면서 복잡한 상황에서의 추적 강인성을 확보한다.

특징 추출을 위한 backbone network로 VGG-M-2048^[14] 모델을 선정하였다. ResNet-50과 같이 더 깊은 네트워크가 존재하나, 본 논문에서는 추가 연산 비용을 최소화하면서도 객체 추적에 필요한 의미론적 정보를 보존하기 위해 상대적으로 파라미터 수가 적고 구조가 단순한 VGG-M-2048을 채택하였다.

CNN의 각 레이어의 특징에 따라 추적 정확도와 속도가 달라진다. 얇은 레이어인 Conv1은 공간 해상도가 높아 위치 추정에 유리하나 의미론적 정보가 부족하고, 깊은 레이어인 Conv3은 의미론적 정보가 풍부하나 풀링(pooling) 과정에서 공간 해상도가 크게 저하된다. 본 논문에서는 BAWCF_CN의 연산 오버헤드를 줄이기 위해 Conv2 단일

레이어만을 사용하였다. VGG-M-2048 Conv2에서 추출한 의미론적 특징과 NLM 전처리를 거친 HC 특징을 연결(concatenate)하여 최종 특징을 구성한다.

그림 3에서는 UAV123@10fps 데이터셋의 Truck2 시퀀스에 대해 HC 특징 기반 응답 맵과 HC+CNN 결합 응답 맵을 비교하였다. HC만 사용한 경우 표적 위치 주변으로 응답이 비교적 넓게 분포하며 주변 배경에서도 불필요한 응답이 발생하는 것을 확인할 수 있다. 반면 HC와 CNN 특징을 결합한 경우 표적 위치에서 더욱 뾰족한 peak가 형성되었다. 이는 CNN의 의미론적 특징이 표적과 배경을 보다 효과적으로 구분하여 HC 특징의 한계를 보완하기 때문으로 해석된다. 특히 조명 변화 및 형태 변화가 존재하는 UAV 환경에서 HC+CNN 결합 방식은 응답 맵의 peak를 뾰족하게 하고 drift 발생 가능성을 감소시켜 보다 안정적인 추적 성능을 제공한다. 다만 일부 프레임에서는 CNN 특징이 배경 정보를 함께 강조하여 응답이 확산되는 경우도 관찰되었으나, 전체적으로는 HC 단독 사용 대비 우수한 추적 성능을 보였다.

IV. 실험

본 연구에서는 제안하는 추적 알고리즘의 성능을 검증하기 위해 UAV 환경의 특성(고도, 시점, 조명, 가림 등)을 반영하는 UAV123@10fps^[15]와 DTB70^[16] 데이터셋을 사용하였다. 실험은 MATLAB R2024a 환경과 AMD Ryzen 9 9950X CPU, 32GB RAM가 구축된 데스크탑에서 진행하였다.

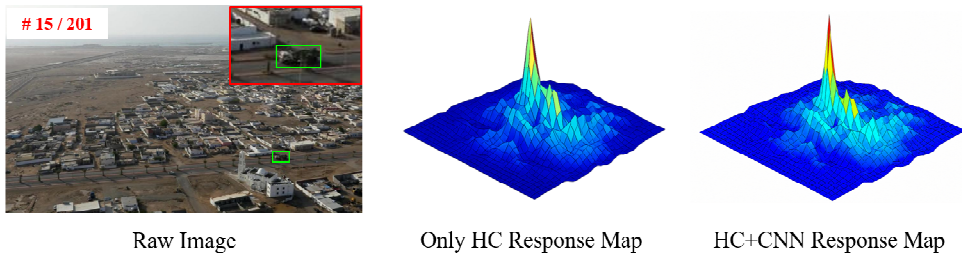


그림 3. UAV123@10fps 데이터셋의 Truck2 시퀀스에서 HC 및 HC+CNN 응답 맵 비교
Fig. 3. Comparison of HC and HC+CNN response maps on the Truck2 sequence from the UAV123@10fps dataset

1. Dataset

본 논문에서는 제안하는 동적 시간 정규화 메커니즘의 추적 성능 및 강건성을 평가하기 위해 두 가지 벤치마크 데이터셋을 활용한다. 첫 번째 데이터셋인 UAV123@10fps는 UAV의 저고도 시점에서 촬영된 123개의 비디오 시퀀스로 구성되어 있다. 해당 데이터셋은 원본 약 110,000 프레임 규모의 30fps 영상을 약 36,000~40,000 프레임 분량의 10fps 환경으로 다운샘플링(down-sampling)하여 구축되었다. UAV123@10fps의 속성으로는 중형비 변화(Aspect Ratio Change, ARC), 배경 간섭(Background Clutter, BC), 카메라 움직임(Camera Motion, CM), 빠른 움직임(Fast Motion, FM), 완전 가림(Full Occlusion, FOC), 조명 변화(Illumination Variation, IV), 저해상도(Low Resolution, LR), 시야 이탈(Out-of-View, OV), 부분 가림(Partial Occlusion, POC), 유사 객체(Similar Object, SOB), 크기 변화(Scale Variation, SV), 시점 변화(Viewpoint Change, VC) 총 12가지가 존재한다. UAV123@10fps는 이러한 극한 환경 속성들을 다수 포함하고 있어, 주로 hand-crafted feature 기반 추적기들의 한계 성능을 비교 및 검증하는 데 주로 활용된다.

두 번째 데이터셋인 DTB70은 대학 캠퍼스 내 드론 촬영 영상과 웹상에서 수집된 영상을 혼합하여 구축된 70개의 비디오 시퀀스(총 15,777 프레임)로 구성된다. 해당 데이터셋은 높은 데이터 다양성과 낮은 편향성을 특징으로 하며, UAV123@10fps와 마찬가지로 hand-crafted feature 기반 추적기들의 성능 평가에 주로 활용된다. DTB70은 총 11가지의 속성을 포함하고 있으며, 이 중 조명 변화(Illumination Variation, IV), 시야 이탈(Out-of-View, OV), 크기 변화(Scale Variation, SV)를 비롯하여, 가림 현상(Occlusion, OCC), 고속 카메라 기동(Fast Camera Motion, FCM), 모션 블러(Motion Blur, MB), 중형비 변화(Aspect Ratio Variation, ARV), 주변 유사 객체 존재(Similar Object Around, SOA) 속성은 앞서 언급한 UAV123@10fps의 주요 속성들과 물리적 특성이 유사하다. 반면, 표적 고유의 형태가 변하는 외형 변형(Deformation, DEF), 표적이 2차원 평면 상에서 회전하는 평면 내 회전(In-plane Rotation, IPR) 및 3차원 공간 상에서 회전하여 시점이 급변하는 평면

외 회전(Out-of-plane Rotation, OPR) 등을 추가로 포함하고 있어 추적 알고리즘의 강건성을 검증하기에 적합하다.

2. Metric

본 논문에서는 추적 알고리즘을 평가하는데 정밀도(Precision)와 성공률(Success) 지표를 사용하였으며, VOT 벤치마크의 표준인 OPE(One-Pass Evaluation) 방식을 적용하여 각각에 대한 값을 측정하였다.

정밀도는 표적의 실제 위치(ground truth)의 중심점과 추적기가 추적한 표적의 중심점 사이의 거리를 중심점 오차(Center Location Error, CLE) 기반으로 산출하여 나타낸 것으로, 표적의 위치를 추적기가 얼마나 정확하게 추정했는지 수치화하여 나타낸 지표이다. 정밀도는 0에서 50픽셀 사이의 임계값 τ 에 대하여, 중심점 위치 오차가 해당 임계값 이하($CLE \leq \tau$)를 만족하는 프레임의 비율로 정의한다.

$$CLE = \sqrt{(x_{tr} - x_{gt})^2 + (y_{tr} - y_{gt})^2} \quad (17)$$

성공률은 추적 결과에 대한 bounding box와 실제 정답지에 대한 bounding box 간의 중첩 영역 비율(Intersection Over Union, IOU)을 기반으로 정의되며, 표적의 크기 및 위치에 대한 정확도를 수치화하여 나타낸 지표이다. 성공률은 임계값 τ ($\tau \in [0, 1]$)의 변화에 따라 $IoU > \tau$ 조건을 충족하는 영상 프레임의 상대적 비율을 연속적으로 도출하여 산출한다. 이를 바탕으로 본 연구에서는 가변적인 비율 분포가 형성하는 곡선의 하위 면적(AUC, Area Under Curve)을 측정함으로써, 제안하는 알고리즘의 종합적인 추적 안정성을 검증하였다.

$$IoU = \frac{|R_{tr} \cap R_{gt}|}{|R_{tr} \cup R_{gt}|} \quad (18)$$

3. 실험 결과

본 절에서는 동적 2차 시간 정규화 항의 외관 변화 및 복잡한 환경 변화에 대한 효과를 분석하고, 추가된 CNN 의미론적 특징이 추적 성능 향상에 미치는 영향을 평가하였다.

표 2. DTB70 데이터셋의 11가지 속성에 대한 정밀도 및 성공률

Table 2. Precision and Success Rates on the DTB70 Dataset across 11 Attributes

Trackers	ARV	BC	DEF	FCM	IPR	MB	OCC	OPR	OV	SOA	SV
STRCF	0.473/0.330	0.621/0.355	0.536/0.413	0.716/0.548	0.581/0.411	0.694/0.506	0.617/0.436	0.291/0.188	0.670/0.399	0.677/0.528	0.550/0.467
SRDCF	0.357/0.244	0.404/0.245	0.311/0.217	0.558/0.464	0.431/0.341	0.458/0.360	0.478/0.347	0.281/0.191	0.573/0.341	0.569/0.428	0.483/0.399
BACF	0.408/0.266	0.509/0.354	0.478/0.337	0.625/0.520	0.549/0.411	0.622/0.471	0.515/0.421	0.334/0.157	0.587/0.368	0.624/0.495	0.558/0.466
CACF	0.545/0.352	0.643/0.483	0.484/0.357	0.768/0.648	0.615/0.494	0.720/0.613	0.633/0.536	0.396/0.225	0.722/0.479	0.729/0.640	0.639/0.517
LADCF	0.483/0.364	0.574/0.348	0.527/0.423	0.718/0.568	0.589/0.442	0.678/0.514	0.667/0.508	0.266/0.185	0.749/0.429	0.679/0.547	0.593/0.514
BiCF	0.544/0.361	0.602/0.427	0.563/0.428	0.699/0.591	0.615/0.487	0.649/0.528	0.566/0.430	0.311/0.229	0.637/0.426	0.685/0.524	0.597/0.536
IBRI	0.589/0.374	0.608/0.427	0.643/0.422	0.721/0.600	0.611/0.472	0.685/0.550	0.588/0.501	0.350/0.215	0.661/0.481	0.700/0.573	0.653/0.544
MSEFCF	0.527/0.341	0.667/0.474	0.598/0.449	0.738/0.628	0.619/0.483	0.749/0.619	0.594/0.487	0.342/0.202	0.702/0.461	0.731/0.607	0.599/0.511
LDECF	0.592/0.388	0.633/0.485	0.661/0.498	0.771/0.656	0.668/0.534	0.735/0.635	0.670/0.554	0.359/0.185	0.637/0.439	0.777/0.647	0.655/0.557
BAWCF	0.580/0.392	0.663/0.494	0.626/0.475	0.751/0.634	0.631/0.499	0.729/0.619	0.615/0.519	0.389/0.282	0.591/0.387	0.779/0.648	0.590/0.521
BAWCF_CN	0.588/0.391	0.624/0.475	0.695/0.522	0.747/0.634	0.649/0.518	0.738/0.619	0.605/0.504	0.421/0.247	0.603/0.409	0.746/0.621	0.623/0.542

표 2는 DTB70의 11가지 속성별 정밀도와 성공률을 나타낸다. 먼저 제안하는 BAWCF는 베이스라인 BiCF 대비 11개 속성 중 9개 속성에서 우위를 보였으며, 특히 DEF, OPR, MB 등 표적 외관이 크게 변화하는 속성에서 높은 성능 향상을 보였다. BAWCF_CN 역시 11개 속성 중 10개 속성에서 BiCF를 상회하였으며, DEF, SOA, MB 등에서 의미론적 특징의 판별력이 효과적으로 발휘되었다. 다만 OV 속성에서는 BAWCF_CN이 BiCF보다 낮은 성능을 보였는데, 이는 시야 이탈 후 재등장 시 CNN 특징의 해상도 손실이 위치 재탐지에 영향을 미친 것으로 분석된다.

한편 IBRI, MSEFCF, LDECF 등 최신 추적기와의 비교

에서 BAWCF는 SOA에서, BAWCF_CN은 DEF에서 모든 비교 추적기 중 가장 높은 정밀도와 성공률을 기록하였으며, OPR에서도 제안 모델이 가장 높은 성능을 보였다. 반면 FCM, IPR, SV 등에서는 LDECF가 우위를 보였다. 이로써 제안 기법은 외형 변화가 큰 속성에서 최신 기법과 대응하거나 우수한 성능을 보이면서, 베이스라인 BiCF 대비 일관된 향상을 유지함을 확인하였다.

그림 4는 UAV123@10fps 데이터셋에 대한 전체 OPE 평가 결과를 나타낸다. 제안한 BAWCF와 BAWCF_CN은 정밀도 및 성공률 지표 모두에서 기존 추적기 대비 우수한 성능을 보였으며, BAWCF_CN이 가장 높은 성능을 기록하였다.

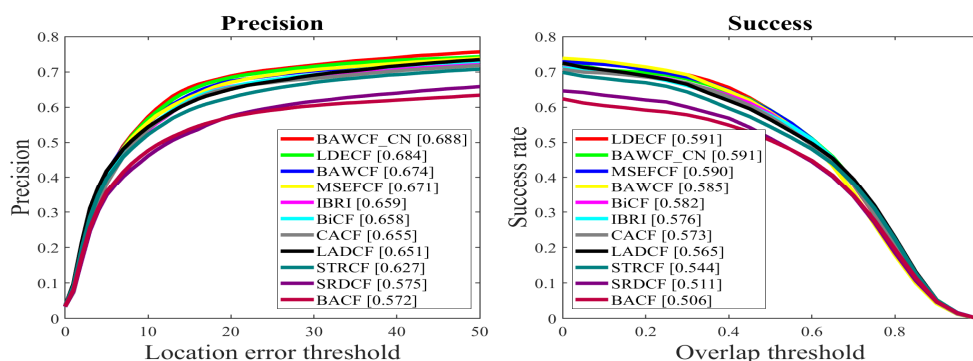


그림 4. UAV123@10fps 데이터셋에 대한 OPE(One-Pass Evaluation) 정밀도 및 성공률 곡선

Fig. 4. Precision and Success Rate Plots of OPE on the UAV123@10fps Dataset

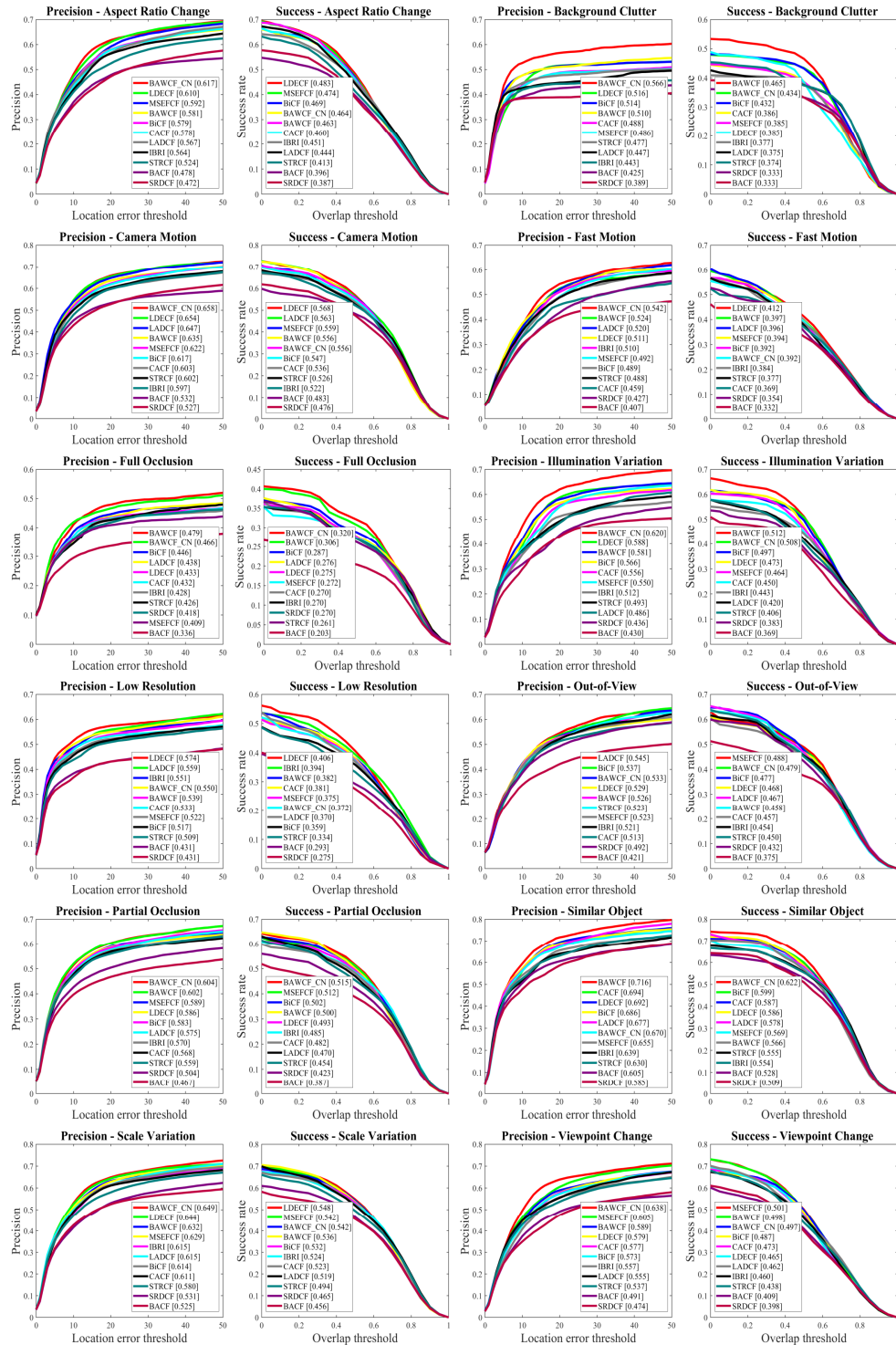


그림 5. UAV123@10fps 데이터셋의 12가지 속성에 대한 OPE(One-Pass Evaluation) 정밀도 및 성공률 곡선
Fig. 5. Precision and Success Rate Plots of OPE on the UAV123@10fps Dataset for 12 Attributes

그림 5는 UAV123@10fps의 12가지 속성별 정밀도와 성공률을 나타낸다. 먼저 제안하는 BAWCF는 베이스라인 BiCF 대비 정밀도에서 12개 속성 중 9개, 성공률에서 10개 속성에서 우위를 보였다. 특히 FM, POC, SOB 등 외관 변화가 큰 속성에서 항상 폭이 두드러졌는데, 이는 적응형 2차 시간 정규화 항이 외관 급변 구간에서 학습 관성을 유지하여 위치 오차의 누적을 차단하고, 안정 구간에서는 부드러운 갱신을 유지함으로써 모델 drift를 효과적으로 억제한 결과로 분석된다.

CNN 의미론적 특징을 결합한 BAWCF_CN은 최신 추적기와 비교했을 때 특히 정밀도 측면에서 경쟁력을 보였다. BC, ARC, CM, VC 등 표적과 배경의 판별이 중요한 속성의 정밀도에서 최신 기법인 LDECf를 상회하였다. 이는 CNN의 의미론적 특징과 NLM 전처리가 색상·텍스처가 유사한 배경과 표적을 효과적으로 구분하여 중심 위치 추정 신뢰도를 높인 결과로 분석된다. 다만 성공률 측면에서는 LR을 비롯한 여러 속성에서 LDECf가 우위를 보였다. LDECf는 적응형 2차 차분을 공간 정규화 가중치에 적용하여 스케일 변화 대응을 강화한 반면, 본 논문은 2차 정규화를 시간 영역 가중치에 적용하여 외관 변화 대응에 초점을 둔다. 이러한 설계 차이가 성공률 격차와 관련된 것으로 보이며, 본 기법에 공간적 적응 요소를 결합하여 중첩·스케일 정합을 개선하는 것은 향후 과제로 남긴다.

종합하면, 제안 기법은 베이스라인 BiCF 대비 일관된 향상을 달성하는 동시에, 외관 변화·배경 간섭 속성의 정밀도에서 최신 기법과 대등하거나 우수한 성능을 보였다. 다만 BAWCF_CN은 CNN 순전파에 따른 추가 연산 비용을 수반하므로, 고성능 연산 자원이 가용한 환경에서 정확도 우선 운용을 목적으로 하는 보완적 확장 모델로서 적합하다.

또한, 본 논문에서는 VGG-M-2048의 저수준, 중간, 고수준 특징을 각각 대표하는 Conv1, Conv2, Conv3 레이어를 대상으로 단일 레이어 사용과 다중 레이어 결합 시의 성능을 비교하였다. 그 결과 다중 레이어 결합 방식은 강인성이 일부 향상되었으나 연산 비용이 증가하는 한계가 있었다. 반면 단일 레이어 기반 방식은 연산 오버헤드를 억제하면서도 안정적인 추적 성능을 유지할 수 있었으며, 특히 Conv2 레이어는 객체의 형태 정보와 의미론적 정보를 균형 있게 포함하여 가장 안정적인 성능을 보였다. 따라서 본 논

문에서는 추적 정확도와 연산 비용 간의 균형을 고려하여 Conv2를 단일 특징 레이어로 채택하였다. 또한 CNN 특징을 사용하지 않고 HC 특징만을 활용하는 BAWCF는 복잡한 배경 및 외관 변화 환경에서의 판별력이 상대적으로 제한되는 경향이 있으나, 연산 자원이 제한된 환경에서 안정적으로 동작하는 기본 모델로서의 역할을 수행한다.

4. Ablation Study

본 논문의 구성 요소별 기여를 정량적으로 검증하기 위해 UAV123@10fps 벤치마크에서 네 가지 실험을 수행하였다.

표 3. 구성 요소별 단계적 성능 분석
Table 3. Step-wise ablation of each component

Method	Dynamic 2nd-order	NLM	CNN	DP	AUC
BiCF(baseline)	✗	✗	✗	0.658	0.582
BAWCF	✓	✗	✗	0.674	0.585
BAWCF + NLM	✓	✓	✗	0.673	0.588
BAWCF + CNN	✓	✗	✓	0.688	0.586
BAWCF_CN	✓	✓	✓	0.688	0.591

표 3은 BiCF 베이스라인에 각 구성 요소를 단계적으로 누적 적용한 결과를 나타낸다. 먼저 동적 2차 시간 정규화 항을 도입한 BAWCF는 BiCF 대비 정밀도 0.658→0.674, 성공률 0.582→0.585를 기록하였으며, 이는 동적 정규화 항이 외관 급변 구간에서 학습 관성을 유지하여 필터 drift를 억제함으로써 중심 위치 추정의 정확도를 높인 결과로 분석된다. 이어 BAWCF에 NLM 전처리를 추가했을 때 성공률이 0.585→0.588로 향상되었고 이는 진동 blur·압축 노이즈로 인한 저수준 특징의 불안정을 완화하여 bounding box 정합도를 개선한 효과로 해석된다. 다만 정밀도는 0.674→0.673으로 감소했지만 사실상 큰 변화가 없었음을 알 수 있다.

한편 BAWCF에 CNN 의미론적 특징을 추가했을 때 정밀도가 0.674→0.688로 향상되었는데, 이는 색상·텍스처가 유사한 배경과 표적을 효과적으로 구분함으로써 판별력을 강화한 결과로 분석된다. 두 요소를 모두 적용한 BAWCF_

CN은 정밀도 0.688, 성공률 0.591로 가장 높은 성능을 달성하였다. 이로써 성능 향상의 핵심은 동적 2차 시간 정규화에 있으며, NLM과 CNN은 확장 모델 BAWCF_CN에서 각각 성공률과 정밀도를 추가로 보완하는 요소임을 정량적으로 확인하였다.

표 4는 CNN 특징 레이어 선택에 따른 추적 성능과 처리 속도를 비교한 결과를 나타낸다. Conv1+Conv3 조합이 가장 높은 정밀도(0.695)를 기록하였으나 FPS가 22로 감소하였다. 반면 Conv2 단일 레이어는 정밀도 0.692, 성공률 0.590, FPS 30을 달성하여 정확도와 연산 효율성 간의 균형이 가장 우수한 것으로 나타났다. 따라서 본 논문에서는 최종 CNN 특징으로 Conv2 레이어를 채택하였다.

한편 표 4의 처리 속도는 두 제안 모델의 역할 차이를 직접 보여준다. HC 특징만을 사용하는 BAWCF는 CNN 순전과 없이 60 FPS로 동작하여, CNN을 결합한 BAWCF_CN(HC+Conv2, 30 FPS)의 약 2배에 달하는 처리 속도를 보인다. 즉 BAWCF는 CNN 순전과를 포함하지 않아 연산 속도가 빠른 반면, BAWCF_CN은 CNN 의미론적 특징과 NLM 전처리를 추가하여 정확도를 높이는 대신 연산 속도가 감소한다. 이러한 구성·속도 차이가 두 모델의 역할 구분(연산 효율 우선/정확도 우선)의 근거가 된다.

표 4. CNN 특징 레이어 구성에 따른 정밀도, 성공률 및 처리 속도(FPS) 비교(AMD Ryzen 9 9950X, MATLAB R2024a 환경에서 측정)
Table 4. Comparison of tracking performance and FPS using different CNN feature layers (measured on AMD Ryzen 9 9950X, MATLAB R2024a)

CNN Layer	Precision	Success	FPS
Only HC	0.673	0.588	60
HC+Conv1	0.688	0.589	30
HC+Conv2	0.692	0.590	30
HC+Conv3	0.684	0.586	30
HC+Conv1+Conv3	0.695	0.588	22

또한 그림 6은 α 값 설정에 따른 정밀도와 성공률의 변화를 나타낸다. α 값의 범위가 [0.5, 1.0]인 경우 DP=0.663, AUC=0.584에서 시작하여 [0.8, 1.0] 구간에서 DP=0.673, AUC=0.593으로 최고 성능을 기록하였으며, α 가 1로 고정된 경우에는 성능이 소폭 감소하는 경향을 보였다. 이는 α 값을 고정적으로 사용하는 경우보다, 추적 상황에 따라

동적으로 조절하는 방식이 더 안정적인 성능을 제공함을 보여준다. 또한 α 값이 낮으면 외관 급변 시 시간적 연속성 정보가 손실되어 drift 억제 효과가 약화되고, 높으면 급격한 외관 변화에 대한 대응력이 감소하기 때문이다. 따라서 본 논문에서는 α 의 범위를 [0.8, 1.0]으로 설정하였다.

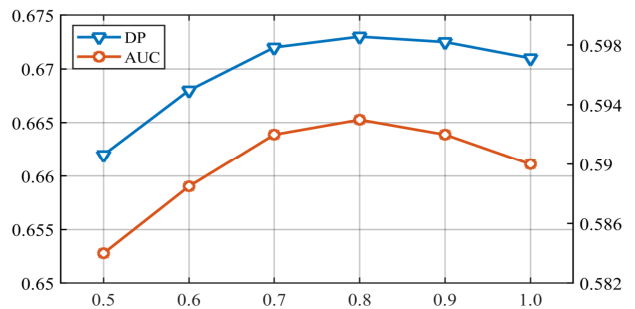


그림 6. α 값 범위의 설정 변화에 따른 정밀도(DP) 및 성공률(AUC) 변화
Fig. 6. Precision (DP) and success rate (AUC) variation according to changes in α value range setting

그림 7은 제안 방법의 주요 파라미터가 추적 성능에 미치는 영향을 확인하기 위해 λ_3 에 대한 민감도 분석을 수행하였다. λ_1 과 λ_2 는 베이스라인 모델인 BiCF의 설정값을 그대로 사용하였으며, 새롭게 도입한 정규화 항의 반영 강도를 조절하는 λ_3 만 별도로 분석하였다. 본 논문에서는 λ_3 를 0.05부터 0.15까지 변화시키며 성능을 비교하였고, 정밀도(DP)와 성공률(AUC)이 가장 높게 나타난 $\lambda_3=0.1$ 을 최종 값으로 설정하였다. λ_3 가 작으면 양방향 불일치 오류에 대한 제약이 약해져 drift 억제 효과가 감소하고, 반대로 크면 현재 프레임의 변화에 대한 적응성이 저하될 수 있다. 따라서 $\lambda_3=0.1$ 은

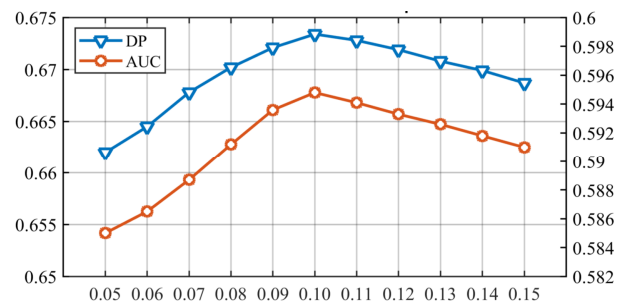


그림 7. λ_3 값 설정 변화에 따른 정밀도(DP) 및 성공률(AUC) 변화
Fig. 7. Precision (DP) and success rate (AUC) variation according to changes in λ_3 value setting

안정성과 적응성 간의 균형을 유지하는 설정이다.

V. 결 론

본 논문에서는 UAV 시각 객체 추적 중 발생하는 복잡한 상황에서 표적을 강건하게 추적하기 위해, 양방향 불일치 오류 기반의 동적 2차 시간 가중치 정규화와 적응형 α 제어를 핵심으로 하는 BAWCF를 제안하였다. BAWCF는 hand-crafted feature만을 사용하는 CPU 지향 기본 모델로, non-convex 목적함수를 ADMM으로 분해하여 닫힌 형식해를 갖는 네 개의 subproblem으로 최적화한다. 나아가 정확도를 우선시하는 운용 환경을 위해, CNN 의미론적 특징과 NLM 전처리를 보완적으로 결합한 확장 모델 BAWCF_CN을 제시하였다.

UAV123@10fps 및 DTB70 데이터셋에서 속성별 정량 비교 실험 측정 결과, BAWCF와 BAWCF_CN은 BiCF 대비 정밀도와 성공률 모두에서 일관된 향상을 보였으며 특히 Fast Motion, Occlusion, Similar Object 속성에서 두드러진 효과를 입증하였다. 이는 제안한 동적 2차 시간 정규화 항과 sigmoid 기반 α 제어 메커니즘이 외관 급변 구간에서의 모델 drift를 효과적으로 억제하고, NLM 전처리와 CNN 의미론적 특징 결합이 복잡한 배경 환경에서의 판별력을 강화하였기 때문으로 분석된다.

앞서, BAWCF를 CPU 친화적 기본 모델, BAWCF_CN을 GPU 가용 환경 확장 모델로 구분한 것은 연산 자원 요구량에 따른 설계 지향을 의미하며, 향후 연구에서는 실제 드론에 탑재되는 Jetson Orin Nano와 같은 엣지 디바이스에 탑재하여 실측 성능을 검증하고자 한다. 또한, 공간 도메인에서의 정규화를 넘어 주파수 도메인에서의 구조적 제약을 활용하여, 표적의 극심한 외형 변화 속에서도 구조적 특징을 강건하게 유지하는 추적 기법을 연구할 계획이다.

참 고 문 헌 (References)

- [1] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual Object Tracking Using Adaptive Correlation Filters," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, USA, pp.2544-2550, June 2010.
doi: <https://doi.org/10.1109/CVPR.2010.5539960>
- [2] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning Spatially Regularized Correlation Filters for Visual Tracking," Proceedings of IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, pp.4310-4318, December 2015.
doi: <https://doi.org/10.1109/ICCV.2015.490>
- [3] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li, "Visual Tracking via Adaptive Spatially-Regularized Correlation Filters," Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, pp.4665-4674, June 2019.
doi: <https://doi.org/10.1109/CVPR.2019.00480>
- [4] F. Li, C. Tian, W. Zuo, L. Zhang, and M.-H. Yang, "Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, USA, pp.4904-4913, June 2018.
doi: <https://doi.org/10.1109/CVPR.2018.00515>
- [5] Y. Zhang, Y.-F. Yu, K.-K. Huang, and Y. Wang, "Channel Attentional Correlation Filters Learning with Second-Order Difference for UAV Tracking," IEEE Geoscience and Remote Sensing Letters, Vol.20, Art.no. 8001905, 2023.
doi: <https://doi.org/10.1109/LGRS.2023.3311441>
- [6] Y.-F. Yu, Z. Chen, Y. Zhang, C. Zhang, and W. Ding, "Learning Dynamic-Sensitivity Enhanced Correlation Filter with Adaptive Second-Order Difference Spatial Regularization for UAV Tracking," IEEE Transactions on Intelligent Transportation Systems, Vol.26, No.5, pp.7211-7230, May 2025.
doi: <https://doi.org/10.1109/TITS.2025.3533953>
- [7] F. Lin, C. Fu, Y. He, F. Guo, and Q. Tang, "BiCF: Learning Bidirectional Incongruity-Aware Correlation Filter for Efficient UAV Object Tracking," Proceedings of IEEE International Conference on Robotics and Automation (ICRA), Paris, France, pp.2365-2371, May 2020.
doi: <https://doi.org/10.1109/ICRA40945.2020.9196530>
- [8] S. Wang, H. Jiang, J. Yang, X. Ma, and J. Chen, "Learning Feature-Weighted Regularization Discriminative Correlation Filters for Real-Time UAV Tracking," Signal Processing, Vol.228, Art.no. 109765, March 2025.
doi: <https://doi.org/10.1016/j.sigpro.2024.109765>
- [9] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical Convolutional Features for Visual Tracking," Proceedings of IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, pp.3074-3082, December 2015.
doi: <https://doi.org/10.1109/ICCV.2015.352>
- [10] M. Danelljan, A. Robinson, F. Shahbaz Khan, and M. Felsberg, "Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking," Proceedings of European Conference on Computer Vision (ECCV), Amsterdam, Netherlands, pp.472-488, October 2016.
doi: https://doi.org/10.1007/978-3-319-46454-1_29
- [11] M. Danelljan, G. Bhat, F. Shahbaz Khan, and M. Felsberg, "ECO:

- Efficient Convolution Operators for Tracking,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, pp.6931-6939, July 2017.
doi: <https://doi.org/10.1109/CVPR.2017.733>
- [12] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, “Convolutional Features for Correlation Filter Based Visual Tracking,” Proceedings of IEEE International Conference on Computer Vision Workshops (ICCVW), Santiago, Chile, pp.621-629, December 2015.
doi: <https://doi.org/10.1109/ICCVW.2015.84>
- [13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Foundations and Trends in Machine Learning, Vol.3, No.1, pp.1-122, January 2011
doi: <https://doi.org/10.1561/22000000016>
- [14] .K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, “Return of the Devil in the Details: Delving Deep into Convolutional Nets,” Proceedings of British Machine Vision Conference (BMVC), Nottingham, UK, pp.6.1-6.12, September 2014.
doi: <https://doi.org/10.5244/C.28.6>
- [15] M. Mueller, N. Smith, and B. Ghanem, “A Benchmark and Simulator for UAV Tracking,” Proceedings of European Conference on Computer Vision (ECCV), Amsterdam, Netherlands, pp.445-461, October 2016.
doi: https://doi.org/10.1007/978-3-319-46448-0_27
- [16] S. Li and D.-Y. Yeung, “Visual Object Tracking for Unmanned Aerial Vehicles: A Benchmark and New Motion Models,” Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI), San Francisco, CA, USA, pp.4140-4146, February 2017.
doi: <https://doi.org/10.1609/aaai.v31i1.11205>

저 자 소 개



이 정 우

- 2021년 3월 ~ 현재 : 한국항공대학교 항공전자정보공학부 학사과정
- ORCID : <https://orcid.org/0009-0005-7247-4437>
- 주관심분야 : 딥러닝, 컴퓨터 비전, 객체 추적, 객체 탐지



고 대 일

- 2024년 2월 : 한서대학교 무인항공기학과(공학사)
- 2025년 3월 ~ 현재 : 한국항공대학교 인공지능학과 석사과정
- ORCID : <https://orcid.org/0009-0009-9644-1401>
- 주관심분야 : 컴퓨터 비전, 딥러닝, Edge AI, AI 기반 표적 추적



권 구 용

- 2014년 ~ 현재 : LIG 디앤에이 광정보융합연구소 선임연구원
- ORCID : <https://orcid.org/0000-0002-3687-6771>
- 주관심분야 : 컴퓨터 비전, Edge AI, FPGA 기반 영상처리, 딥러닝

저 자 소 개

황 재 기



- 2023년 ~ 현재 : LIG 디앤에이 광정보융합연구소 선임연구원
- ORCID : <https://orcid.org/0000-0003-3062-5530>
- 주관심분야 : MOT, 소형 객체 탐지, 딥러닝

장 현 성



- 2011년 ~ 현재 : LIG 디앤에이 광정보융합연구소 팀장
- ORCID : <https://orcid.org/0000-0002-5797-7264>
- 주관심분야 : 컴퓨터 비전, SFOD, 영상 융합, 소프트웨어 아키텍처

하 남 구



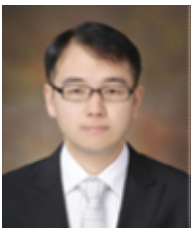
- 2009년 ~ 현재 : LIG 디앤에이 광정보융합연구소 팀장
- ORCID : <https://orcid.org/0000-0001-5146-4794>
- 주관심분야 : EO/IR, MUM-T, 컴퓨터 비전, 딥러닝

김 선 옥



- 2014년 : 연세대학교 전기전자공학과(공학사)
- 2019년 : 연세대학교 전기전자공학과 대학원(공학박사)
- 2021년 ~ 현재 : 한국항공대학교 인공지능학과 부교수
- ORCID : <https://orcid.org/0000-0002-9665-4214>
- 주관심분야 : 컴퓨터 비전, 딥러닝, 위성영상처리

김 필 은



- 2008년 : 인하대학교 기계공학부(공학사)
- 2017년 : Georgia Institute of Technology, Mechanical Engineering 대학원(공학석사)
- 2020년 : 동 대학원 Civil Engineering(공학박사)
- 2021년 ~ 현재 : 한국항공대학교 AI자율주행시스템공학과 조교수
- ORCID : <https://orcid.org/0000-0001-8750-4524>
- 주관심분야 : 자율 시스템, 모바일 로봇, 인공지능 기반 인식 및 제어