

CSRDCF_BD: Adaptive Spatial Mask Refinement for Visual Tracking via Bhattacharyya Distance

Jung-Min Song*

*Researcher, Agency for Defense Development, Daejeon, Korea

[Abstract]

This paper proposes CSRDCF_BD, a novel extension of the CSRDCF tracker that addresses the limitations of spatial reliability modeling through adaptive filtering based on the Bhattacharyya Distance. The proposed method dynamically adjusts the weighting of binary masks by measuring histogram similarity between the foreground and background regions. This enables more robust filter learning under challenging visual conditions such as occlusion, background clutter, and illumination variation. Quantitative evaluations on the OTB100 and VisDrone benchmarks demonstrate that CSRDCF_BD consistently outperforms the baseline CSRDCF and other state-of-the-art trackers in terms of both precision and success rate. Furthermore, CSRDCF_BD exhibits superior temporal stability, as evidenced by its performance across attribute-specific and sequence-level analyses. These results confirm that refined spatial reliability modeling combined with adaptive filtering strategies can significantly enhance the robustness and generalization capability of correlation filter-based object tracking systems.

▶ **Key words:** Object Tracking, Correlation Filter, Spatial Mask Reliability, Bhattacharyya Distance, Adaptive Filtering

[요 약]

본 논문에서는 기존 CSRDCF 추적기의 공간적 신뢰도 모델에 대한 한계를 보완하기 위해, 마스크의 정확도와 신뢰도를 정량적으로 조절하는 Bhattacharyya Distance 기반의 적응형 필터링 기법, CSRDCF_BD를 제안한다. 제안된 방법은 foreground-background 간의 히스토그램 유사도에 기반하여 마스크에 대한 가중치를 동적으로 조정하며, 가림, 배경 혼잡, 조명 변화 등 복잡한 시각 조건에서도 필터 학습의 안정성과 추적 성능을 향상시킨다. OTB100 및 VisDrone 벤치마크에서 수행한 정량 실험 결과, CSRDCF_BD는 기존 CSRDCF 및 주요 추적기 대비 높은 정밀도와 성공률을 기록하였으며, 시간 안정성 측면에서도 우수한 일관성을 보였다. 본 연구는 공간 신뢰도 모델의 정교화 및 적응형 필터링 전략이 객체 추적기의 강건성 향상에 실질적으로 기여할 수 있음을 실증적으로 보여준다.

▶ **주제어:** 객체 추적, 상관 필터, 공간 마스크 신뢰도, 바타차리야 거리(Bhattacharyya Distance), 적응형 필터링

-
- First Author: Jung-Min Song, Corresponding Author: Jung-Min Song
 - *Jung-Min Song (jungmin_song@add.re.kr), Agency for Defense Development
 - Received: 2025. 05. 15, Revised: 2025. 06. 09, Accepted: 2025. 06. 09.

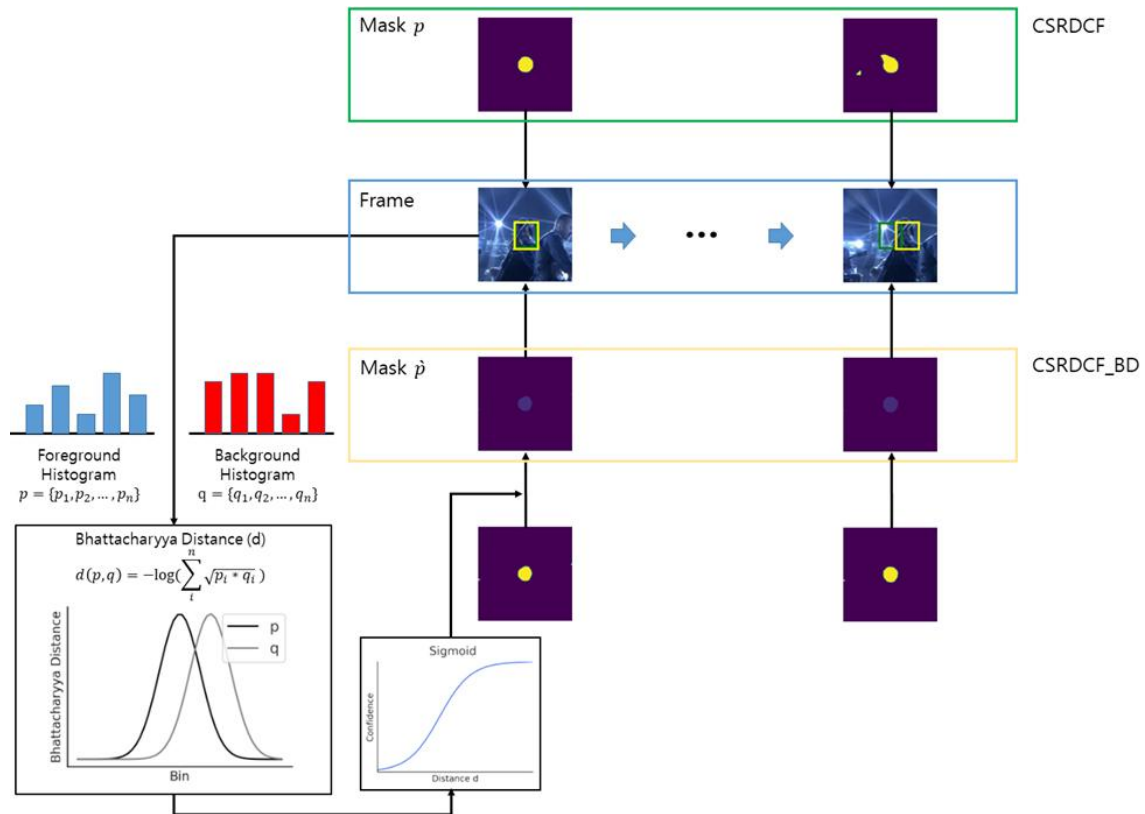


Fig. 1. Overview of CSRDCF_BD approach.

I. Introduction

객체 추적(Object Tracking)은 컴퓨터 비전 분야에서 자율주행, 영상 감시, 증강 현실, 로봇 비전 등 핵심적인 역할을 한다. 이 중 단일 객체 추적(Single Object Tracking, SOT)은 초기 프레임에 지정된 객체의 정보를 바탕으로, 이후 프레임에 따라 해당 객체의 위치를 지속적으로 추적하는 문제로 정의된다. 다양한 시각적 변화(외형, 조명, 가림 등) 속에서도 객체 추적의 안정성을 유지하는 것을 목표로 한다.

고전적 SOT 방식으로는 상관 필터(Correlation Filter, CF) 기반 추적이 널리 활용되어 왔다. CF 기반 추적기는 단순한 목적 함수와 주기적 경계 조건을 활용한 FFT 최적화를 통해 실시간 속도와 낮은 연산 비용의 이점을 가진다. 대표적인 CF 기반 추적기는 MOSSE[1], KCF[2], DSST[3], SRDCF[4] 등이 있으며, UAV, 임베디드 시스템, 실시간 로봇 비전 환경 등에서 그 효율성이 입증되었다.

반면, 딥러닝 기반 추적기는 SOT 성능 향상의 새로운 패러다임을 제시하고 있다. SiamRPN 계열은 region proposal 방식을 통해 정확한 바운딩 박스를 생성하고[5], TransT는 self-attention 기반의 특징 융합을 통해 복잡한 시각적 변화에 강인한 성능을 보였다[6]. 그러나, 이들

방식은 대규모의 학습 데이터와 고성능 GPU 자원을 요구하며, 실시간 처리 및 경량 시스템 적용에는 여전히 제약이 따른다. 이를 보완하기 위해 few-shot 및 zero-shot 학습 기법이 고려되고 있으나 각각 데이터 품질 및 외부 지식 의존성 문제로 인해 실용화에 한계가 존재한다[7].

이러한 배경에서 CF 기반 추적기는 여전히 다양한 환경에서 실용적인 대안으로 고려되고 있다. CF 기반 추적기 CSRDCF(Discriminative Correlation Filter with Channel and Spatial Reliability)[4]는 채널별 중요도(channel reliability)와 공간적 신뢰도(spatial reliability)를 결합한 구조로, 기존 CF 기반 추적이 배경 간섭에 취약하고 공간적 정보 활용이 제한적이라는 한계를 극복하고 향상된 추적 성능을 보인다. CSRDCF는 색상 히스토그램 기반의 foreground/background 분할을 수행하고, 이를 이진화하여 공간 신뢰도 마스크를 생성함으로써 background 간섭을 억제하고 foreground 중심의 응답 학습을 유도한다.

하지만 이런 이진 마스크 방식은 분할 품질에 민감하며, foreground와 background 간 색상 분포가 유사한 경우 잘못된 마스크가 필터 학습에 직접적인 오류로 전이되는 문제가 발생할 수 있다. 본 논문은 이 한계를 극복하기 위해, foreground 및 background 분포 간의 유사도를

Bhattacharyya distance(BD)[10]로 정량화하고, 이를 신뢰도로 변환하여 마스크에 가중치 형태로 적용한다. 이는 기존 CSRDCF의 구조를 유지하면서도, 분할 품질의 불확실성에 따라 학습 제약 강도를 동적으로 조정할 수 있는 점에서 기존 이진 마스크 방식 대비 높은 유연성과 안정성을 제공한다.

본 논문의 기여는 다음과 같다. 첫째, BD를 이용해 분할 마스크의 신뢰도를 정량화하고 이를 필터 학습에 반영하는 기법을 제안한다. 둘째, OTB100[12]와 VisDrone2019-SOT[18] 벤치마크 실험을 통해 CPU 환경에서도 구동 가능한 CF 기반 추적기들과의 비교를 통해 제안 기법의 성능 우수성과 실용성을 입증한다.

제안 기법은 공개된 추적기 코드[17]를 기반으로 구현되었으며, 핵심 수정 사항은 BD 기반 적응형 분할 마스크 구조이다. Fig 1은 마스크에 따른 CSRDCF와 CSRDCF_BD의 기본 구조를 비교하여 나타낸다.

II. Related Works

1. Correlation Filter-Based Tracker

CF 기반 추적기는 높은 연산 효율로 실시간 성능을 만족시킬 수 있어, 다양한 저전력 환경에서 널리 활용되어 왔다. MOSSE는 고속 처리를 구현한 CF 기반 구조이며, KCF는 푸리에 도메인에서의 고속 처리와 다채널 특징을 통합함으로써 정확도와 효율을 동시에 향상시켰다. 이후 SRDCF는 공간 정규화를 통해 필터의 경계 왜곡 문제와 배경 간섭을 완화하였다. 이러한 발전을 기반으로 제안된 CSRDCF는 CF 필터의 표현력을 향상시키기 위해 채널별 가중치와 공간 신뢰도를 도입하였다.

2. Overview of CSRDCF

CSRDCF는 분할 기반의 공간 마스크 p 를 CF 필터 학습에 포함시킴으로써, foreground 중심의 필터 응답을 유도하고 background 간섭을 줄이는 방식으로 동작한다.

3. Extensions of CSRDCF Method

CSRDCF를 바탕으로 SOT의 성능 개선을 위해 다양한 시도가 있었다. Adaptive Spatial-Temporal Regularization (ASTR)[10]은 시간 축에서 필터의 변화량을 정규화하여 외형 및 스케일 변화에 견디는 필터 안정화 구조를 설계하였다. 또한, Auto-Learning Correlation Filter (ALCF)[11]는 응답 맵의 통계 분포를 기반으로 필

터 업데이트 강도를 자동 조정하여, 외부 변화에 적응하는 유연성을 부여하였다. 이들 방법은 주로 필터의 적응성과 학습 안정성에 초점을 맞추고 있으며, 분할 마스크의 품질에 대한 정량적 평가에 기반한 가중 조정 방식은 연구된 바 없다.

4. Challenges in Spatial Reliability

CSRDCF는 분할 마스크가 필터 학습에서 foreground의 응답을 강조하고 배경의 영향을 억제하는 핵심 요소로 작용한다. 그러나 대부분의 CSRDCF 기반 연구는 공간 마스크를 고정된 이진 형태로 적용하며, 해당 마스크의 품질을 동적으로 평가하거나 필터 학습에 반영하는 설계는 부족하다. 결과적으로 색상 분포가 유사하거나 조명 변화가 큰 환경에서는 분할 마스크의 오류가 필터 학습에 직접 전이되어 성능 저하를 초래할 수 있다.

본 논문은 이러한 구조적 한계를 해결하고자 분할 마스크의 신뢰도를 정량적으로 평가하고 가중 조정하는 새로운 방식을 제안한다. 이를 위해 foreground와 background 간의 색상 분포 유사도를 측정하는 통계적 거리 척도인 BD를 도입하고, 이를 필터 학습에 직접 반영하는 구조를 제안하였다.

5. BD in Image Processing

BD는 확률 분포 간의 중첩 정도를 기반으로 거리 값을 산출하는 통계적 척도로, 영역 분리[9], 배경 제거[19] 등 다양한 영상 처리 응용에서 활용되어 왔다. 커널 기반 추적기에서 색상 히스토그램의 유사도 계산에 이를 도입하였으며[10], 이후 객체 인식, 모션 분석, 영상 분할 등에서 foreground와 background 모델의 분리도를 평가하는 수단으로 널리 사용되었다[8, 9].

최근에는 의료 영상의 조직 분할 평가, 열화상 기반 추적, 색상 기반 탐지 등에 BD가 적용되어 왔지만, 대부분은 후처리 단계의 평가 기준이나 threshold 조정 지표로 사용되었다. 반면, 본 논문은 BD를 활용하여 분할 마스크의 신뢰도를 계산하고, 이를 필터 학습에 직접적으로 반영하는 구조를 제안한다는 점에서 기존 연구와 본질적으로 차별화된 접근을 수행한다.

III. The Proposed Scheme

본 장에서는 기존 CSRDCF의 기본 구조를 간략히 소개한 뒤, 제안하는 BD 기반 마스크 신뢰도 조정 기법을 수식과

함께 설명한다. 마지막으로, 신뢰도 기반 적응형 마스크가 CF 필터 학습에 어떤 영향을 미치는지에 대해 설명한다.

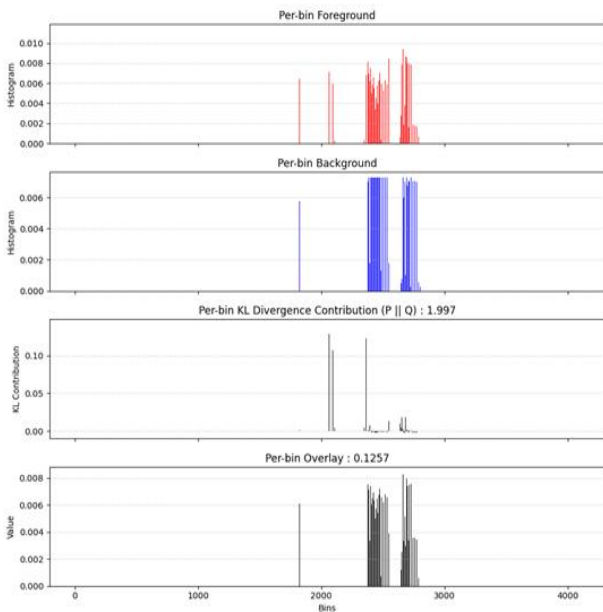


Fig. 2. BD vs KL with OTB100 Shaking

1. CSRDCF

CSRDCF는 각 채널 c 에 대해 다음과 같은 목적 함수를 최소화하여 필터를 학습한다:

$$\min_h \sum_c \|h_c * f_c - y\|^2 + \lambda * \|p \odot h_c\|^2$$

여기서 f_c 는 입력 특징 맵, h_c 는 CF 학습 필터, y 는 타겟 응답 라벨, p 는 공간 마스크이며, \odot 는 Hadamard 곱을 나타낸다. 자세한 구조 및 최적화는 원 논문 [4]을 따른다.

2. Foreground-Background Distribution Modeling for Confidence Adjustment

기존 CSRDCF의 이진 마스크 적용은 분할 품질에 관계 없이 고정된 학습 제약을 유도한다. 본 절에서는 색상 분포 간 유사도에 기반한 신뢰도 정량화를 통해 이를 보완하는 방식을 설명한다.

Foreground와 background 각각에 대해 색상 분포를 히스토그램으로 모델링하고, 두 분포 간의 유사도를 기반으로 마스크의 신뢰도를 계산한다. 이때 거리 함수로

Histogram Intersection(HI), KL Divergence(KL), BD를 비교하였으며, 다음과 같은 분석을 통해 BD를 최종 채택하였다:

- HI는 계산이 간단하지만, 분포의 구조적 차이에 둔감하며, 대칭성 부족으로 해석력에 한계가 있다.
- KL은 민감도는 높지만, 일부 bin에 과도하게 영향을 받는 불안정성을 가진다.
- BD는 분포 간 중첩 정도를 기반으로 안정적인 거리 값을 제공하며, 구조적 유사성에도 민감하게 반응한다.

Fig 3은 foreground와 background가 유사한 상황임에도 KL값이 커지는 것을 확인할 수 있다. 반면, BD는 낮은 값을 가진다. 본 연구에서는 BD를 마스크 신뢰도 평가 지표로 채택하고 필터 학습에 반영한다[8, 9].

3. BD-Based Confidence Adjustment

BD를 이용해 마스크 신뢰도를 계산하고 이를 필터 학습에 반영하는 과정에 대해 구체적으로 설명한다.

3.1 Foreground and Background Histogram Modeling

Foreground Ω_f , background Ω_b 에 대해 색상 히스토그램 p_i, q_i 를 다음과 같이 정의한다:

$$p_i = (1/Z_f) * \sum_{x \in \Omega_f} w_f(x) * \delta(b(x) = i)$$

$$q_i = (1/Z_b) * \sum_{x \in \Omega_b} w_b(x) * \delta(b(x) = i)$$

여기서 w_f, w_b 는 Epanechnikov 커널 기반 가중치, $b(x)$ 는 픽셀 x 의 색상 bin 인덱스를 나타낸다.

3.2 BD Computation

Bhattacharyya 계수 BC 는 다음과 같이 정의되며, 이로부터 거리 d 를 계산한다:

$$BC(p, q) = \sum_i \sqrt{p_i * q_i}$$

$$d(p, q) = -\log(BC(p, q) + \epsilon)$$

3.3 Confidence Transformation

거리 d 는 sigmoid를 통해 신뢰도값으로 변환된다[13]:

$$confidence = 1 / (1 + \exp(-\alpha * (d - \theta)))$$

α 는 민감도 조절 계수이며, θ 는 기준 임계값이다. 본 논문에서는 각각 5, 0.5로 설정하였다.

3.4 Mask Adjustment

최종 분할 마스크 \hat{p} 는 기존 이진 마스크 p 에 신뢰도를 곱하여 조정한다:

$$\hat{p} = confidence * p$$

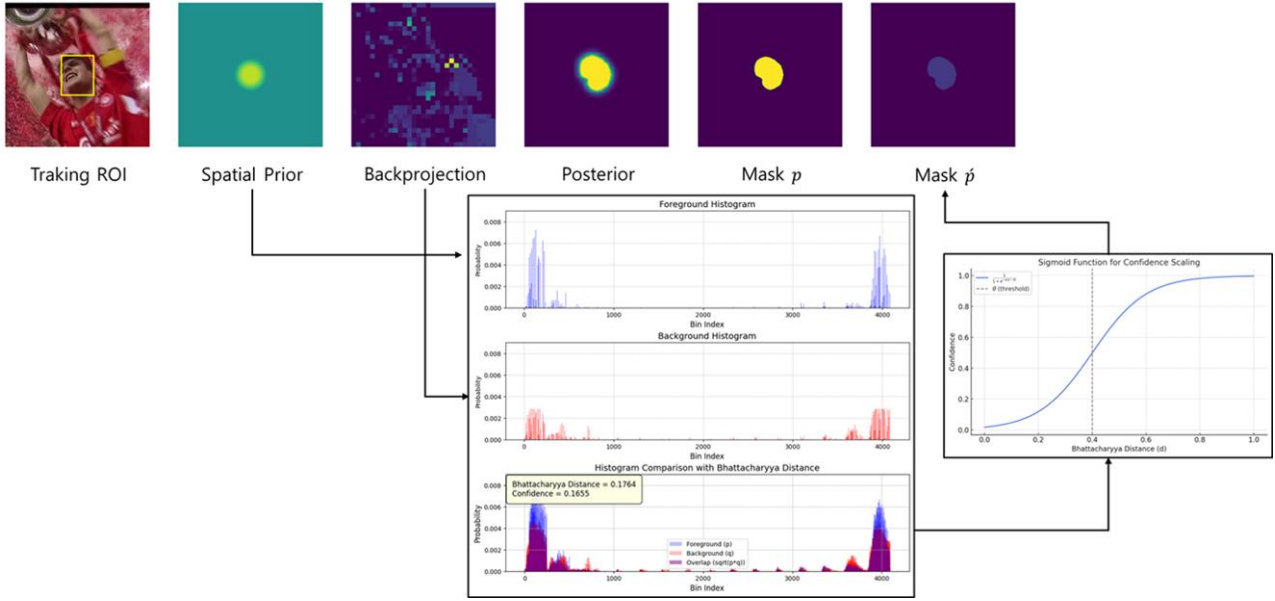


Fig. 3. Adaptive Mask based on Bhattacharyya Distance

4. Adaptive Confidence-Based Masking

제안 기법의 핵심은 분할 품질에 따라 필터 학습의 제약 강도를 조절하는 적응형 마스크 학습을 실현하는 데 있다.

- BD가 작을 때 (유사 분포): 낮은 신뢰도로 마스크 제약을 완화하여 필터가 더 유연하게 갱신된다. 잘못된 분할로 인한 CF 필터 왜곡을 방지한다.

- BD가 클 때 (명확한 분리): 높은 신뢰도를 부여해 foreground 중심 학습을 강화한다.

이러한 적응형 구조는 다양한 환경 변화에 대응하는 안정적인 CF 필터 학습을 가능하게 하며, 과적합이나 강제 일반화 문제를 동시에 완화한다. 실험 결과에서도 다양한 조건에서 추적 성능의 향상을 확인할 수 있다.

Fig. 3는 BD를 기반으로 한 신뢰도 계산부터 마스크 조정까지의 전체 과정을 시각적으로 요약한다.

IV. Experiment

본 논문에서는 제안하는 CSRDCF_BD 트래커의 성능을 정량적으로 평가하기 위해 대표적인 SOT 벤치마크인 OTB100, VisDrone 데이터셋을 사용하였다.

1. Setup

1.1 OTB100

OTB100은 다양한 난이도와 환경 변화를 고려한 100개의 시퀀스로 구성된 SOT 벤치마크로, 객체 추적 성능 비교에 널리 사용된다.

1.2 VisDrone

VisDrone 2019 SOT 데이터셋은 드론 기반 SOT를 위한 벤치마크로, 다양한 조명, 배경 복잡성, 시점 변화, 객체 스케일 변화 등 실제 환경의 도전적인 조건을 반영한다. 본 논문에서는 VisDrone 2019의 test-dev(이하 VisDrone)를 활용하여 추적기의 성능을 평가하였다.

Table. 1은 OTB100과 VisDrone의 주요 추적 속성 정의 및 포함 여부를 요약하며, VisDrone 고유의 시점 변화 요소도 포함한다.

1.3 Evaluation Metrics

추적기는 Precision, Success로 성능을 비교한다. Precision은 추적 결과의 중심점 위치 정확도를 측정한다. 전체 프레임 수를 N 이라 하고 각 프레임에서 추적 결과의 중심점 (x_t, y_t) 과 정답 중심점 (x_g, y_g) 간의 유클리드 거리 d_t 라 하면 Precision은 다음과 같이 정의한다.

$$d_t = \sqrt{[(x_t - x_g)^2 + (y_t - y_g)^2]}$$

$$Precision(\tau) = \frac{1}{N} \sum_{t=1}^N \mathbf{1}_A(d_t \leq \tau)$$

임계값 τ 는 20픽셀로 하며, $\mathbf{1}_A$ 는 지시 함수다.

Success는 추적 결과 B_t 와 정답 박스 B_g 간의 영역 중첩 정도를 기반으로 한다.

$$IoU_t = \frac{|B_t \cap B_g|}{|B_t \cup B_g|}$$

$$Success(\theta) = \frac{1}{N} \sum_{t=1}^N \mathbf{1}(IoU_t \geq \theta)$$

Table 1. Attribute Definitions and Coverage in OTB100 and VisDrone Benchmarks

Attribute	Definition	OTB100	VisDrone
IV	Illumination changes affecting appearance	Illumination Variation	
SV	Scale changes due to distance or zoom	Scale Variation	
OCC	Partial or full occlusion by objects	Occlusion	Occlusion (Full/Partial)
FM	Fast target motion	Fast Motion	
LR	Low spatial resolution	Low Resolution	
OV	Target leaves camera view	Out-of-View	
BC	Cluttered or similar background	B. Clutter*	B. Clutter*/ Sim. Obj.*
DEF	Non-rigid deformation or shape change	Deformation	Aspect Ratio Change
IPR	In-/Out-of-plane rotation of the target	In-/Out Rotation	Viewpoint Change
MB	Blur from motion of target or camera	Motion Blur	-
CM	Viewpoint shift due to camera	-	camera motion

* B. Clutter: Background Clutter / * Sim. Obj.: Similar Object

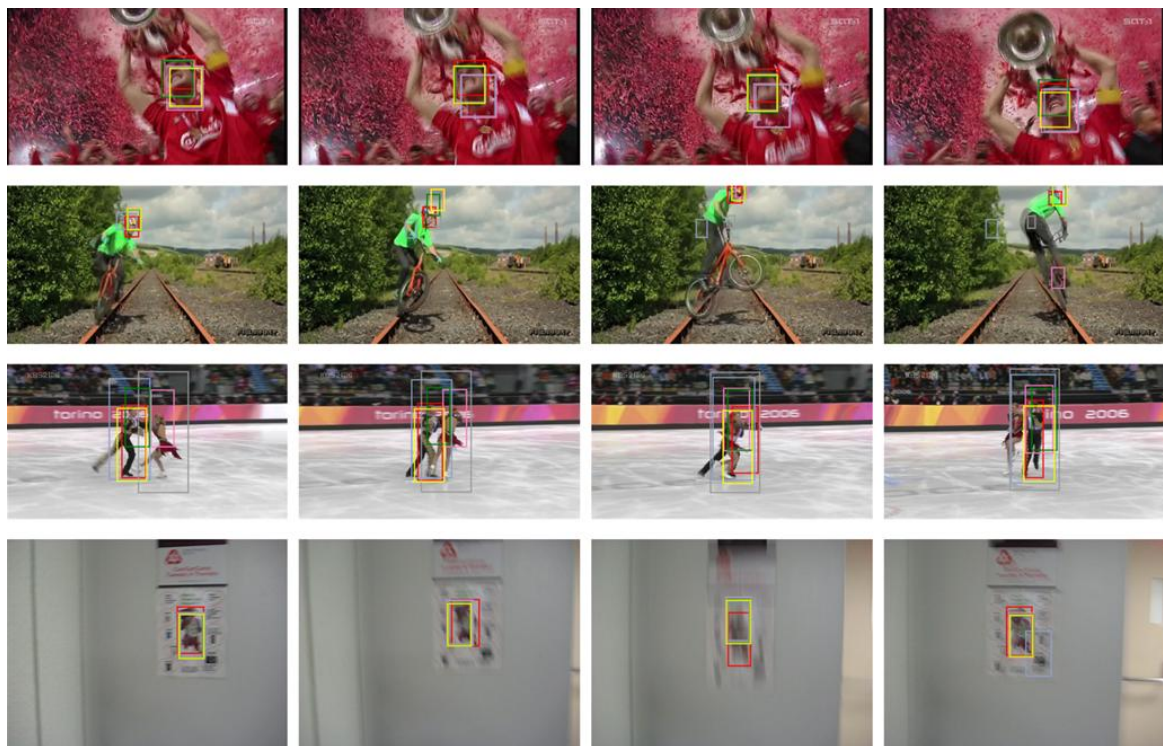


Fig. 4. OTB100 Dataset Track Result of 4 video sequence (i.e. Soccer, Biker, Skating2-2, BlurOwl) with GT, BACF, LDES, SAMF, CSRDCF, CSRDCF_BD

2. Results

제안하는 CSRDCF_BD는 OTB100 전체 시퀀스를 대상으로 한 평가에서 기존 CSRDCF 및 주요 추적기들 대비 우수한 성능을 보였으며, 특히 Success 지표에서 가장 높은 값인 0.5641을 기록하였다. Fig 5와 Table 2는 OTB100 전체 시퀀스에서의 평균 추적 성능을 비교한 결과를 보여준다.

제안 기법의 일반화 성능을 검증하기 위해, OTB100에서 Precision, Success가 우수한 CF 기반 추적기들을 VisDrone에 추가 실험 대상으로 선정하였다. 상대적으로 높은 성능을 보인 추적기들(CSRDCF, BACF, SAMF, LDES, Staple)을 대상으로 VisDrone 실험을 추가로 수행

하였다. VisDrone에서도 CSRDCF_BD는 Precision 0.7859로 모든 비교 추적기 중 가장 높은 값을 기록하였으며, Success는 0.5437로 CSRDCF(0.5586), BACF(0.5454)와 유사한 수준을 유지하였다. Fig 6와 Table 2는 VisDrone 전체 시퀀스에 대해 평균 추적 성능 비교를 나타낸다. 이는 드론 기반의 복잡한 환경에서도 CSRDCF_BD가 높은 정밀도와 안정적인 성능을 유지함을 나타낸다.

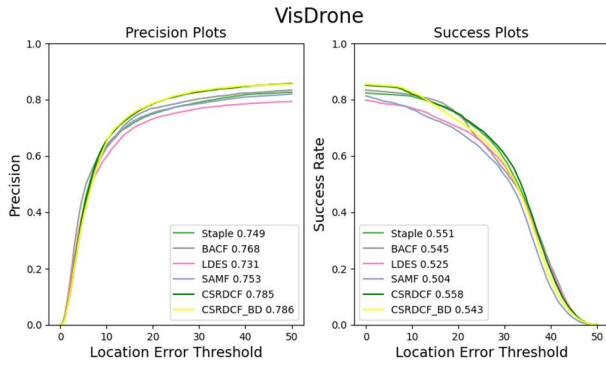


Fig. 5. VisDrone Precision/Success

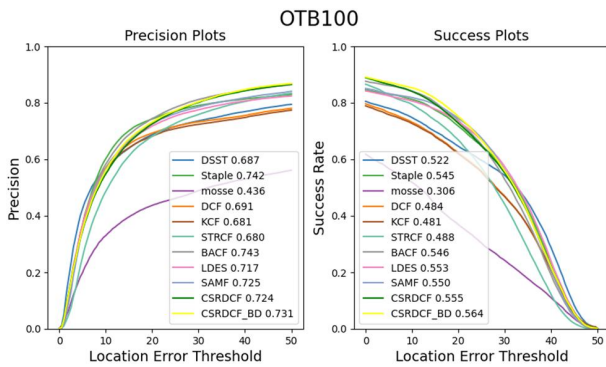


Fig. 6. OTB100 Precision/Success

Table 2. OTB100 and Vis Precision/Success

Tracker	OTB100		VisDrone	
	Pre*	Suc*	Pre*	Suc*
CSRDCF_BD	0.731	0.564	0.786	0.544
CSRDCF	0.725	0.555	0.785	0.558
LDES[15]	0.718	0.553	0.731	0.525
SAMF[13]	0.726	0.551	0.752	0.504
BACF[14]	0.744	0.546	0.768	0.545

* Pre: Precision / * Suc: Success

V. Discussion

1. Overall Performance on Benchmark Datasets

본 논문에서 제안한 CSRDCF_BD 트래커는 대표적인 벤치마크 데이터셋인 OTB100과 VisDrone을 기반으로 성능을 정량적으로 평가하였다. OTB100 실험 결과, CSRDCF_BD는 Precision 0.731, Success 0.564를 기록하여 기존 CSRDCF 및 비교 대상 추적기들 대비 가장 우수한 성공률을 나타냈다. 이는 BD 기반의 공간 신뢰도 모델이 다양한 환경 변화 속에서도 필터의 안정성과 정밀도를 유지하는 데 효과적임을 의미한다.

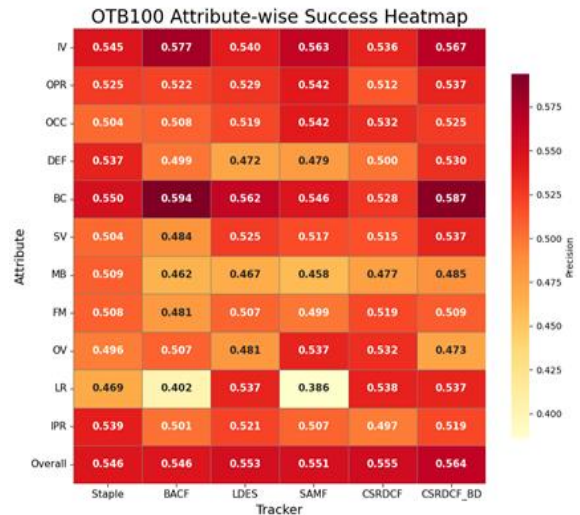
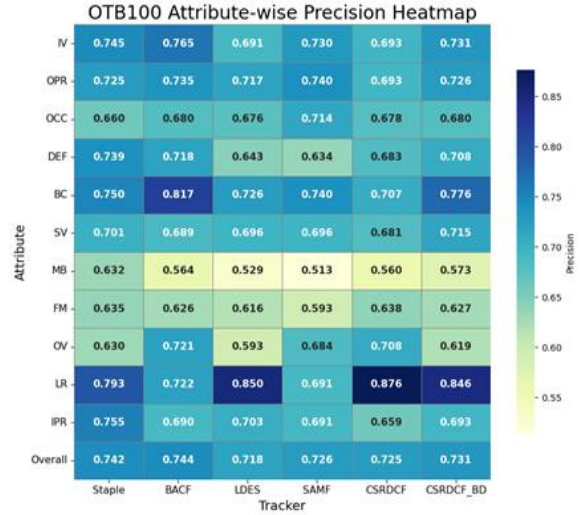


Fig. 7. OTB100 Attribute Precision/Success Heatmap

VisDrone 데이터셋은 드론 기반 영상의 시점 변화, 배경 복잡도, 조명 및 스케일 변화 등 보다 복잡한 조건을 포함하며, 실제 환경을 반영한 일반화 성능 평가에 적합하다. 해당 데이터셋에서 CSRDCF_BD는 Precision 0.786, Success 0.544를 기록하며, 정밀도 측면에서는 가장 뛰어난 성능을 보였고, 성공률 역시 다른 추적기와 비교하여 경쟁력 있는 수준을 유지하였다. 이는 제안한 방법이 정형화된 벤치마크뿐만 아니라 실제 환경에 가까운 조건에서도 높은 추적 안정성을 확보함을 의미한다.

두 벤치마크에 걸친 결과는 CSRDCF_BD가 다양한 환경 조건에서도 일관된 성능을 발휘하며, 일반화 가능성 측면에서 기존 방법 대비 우수한 특성을 지님을 보여준다.

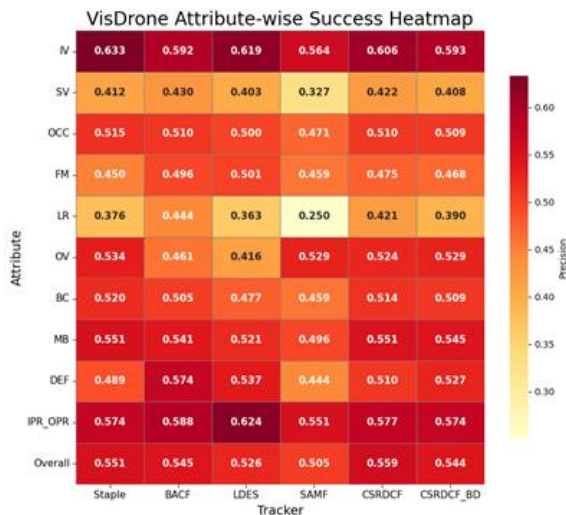
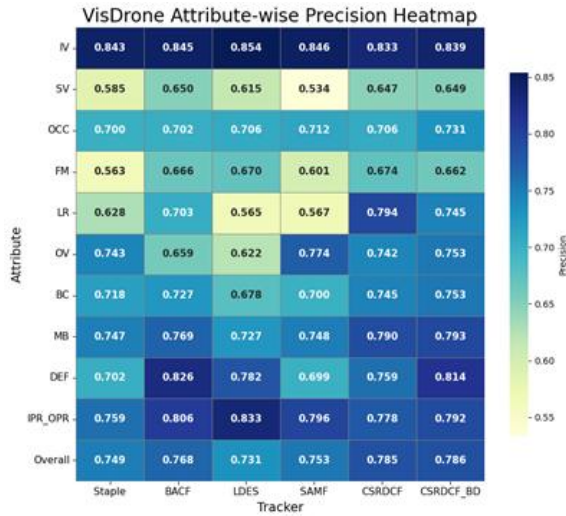


Fig. 8. VisDrone Attribute Precision/Success Heatmap

2. Attribute-Level Performance Analysis

OTB100와 VisDrone은 다양한 환경적 속성에 따라 시퀀스를 분류하고 있어, 속성 기반 성능 분석은 추적기의 세부적인 강점과 약점을 규명하는 데 중요한 기준이 된다.

속성 기반 및 시간 안정성 분석은 CSRDCF_BD의 구조적 강점을 다각도로 입증한다. 먼저, OTB100과 VisDrone 데이터셋 모두에서 CSRDCF_BD는 다양한 속성에 대해 Precision 및 Success 양 측면에서 높은 성능을 유지하였다. OTB100의 히트맵 분석 Fig 7에 따르면 CSRDCF_BD는 IV, BC, LR 속성에서 Top-3 이상의 일관된 성능을 보였고, VisDrone에서도 Fig 8을 통해 DEF, MB, OV 등 복잡한 시각 조건 하에서도 상위권 성능을 유지하는 것을 확인할 수 있다.

이는 특히 foreground-background 구분이 모호하거나 마스크 품질이 저하될 수 있는 조건에서 BD 기반 마스크 신뢰도 조절 기법이 오탐을 줄이고 CF 필터 학습을 유



Fig. 9. CSRDCF_BD Example OTB100 Human3

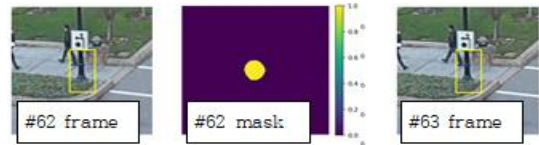


Fig. 10. CSRDCF Example OTB100 Human3

연하게 조정해주는 효과를 제공했기 때문이다. Fig 9와 Fig 10은 OTB100의 Human3의 탐지 결과이다. BC, OCC 상황으로 객체가 일시적으로 탐지되지 않고 마스크는 중심점 근처의 작은 응답만을 포함하게 된다. 기존 CSRDCF는 이진 마스크를 그대로 신뢰하고 강한 제약을 적용하여 추적에 실패한다. 반면 CSRDCF_BD는 해당 마스크의 신뢰도를 낮게 평가하고, 필터 학습 시 마스크 외곽까지 주변 배경 정보를 포함하여 탐색 공간을 유연하게 확장한다. 이후 #63프레임에서 객체가 다시 시야에 등장했을 때, 강건하게 응답을 회복하고 추적을 이어가는 데 성공한다.

단순한 평균 성능만으로는 실제 환경에서의 신뢰도나 일관성을 충분히 설명하기 어렵다. 이를 보완하기 위해 논문은 Stability Deviation이라는 개념을 도입하였다. 이 지표는 속성별 시퀀스에서 프레임 단위로 계산된 정밀도 또는 성공률 곡선의 표준편차 곡선에 대한 면적(AUC)을 측정하여, 시간에 따른 성능의 일관성을 수치화한다. 값이 낮을수록 성능의 출렁임이 적고, 시퀀스 전반에 걸쳐 안정적이고 예측 가능한 추적이 이루어졌음을 의미한다. Fig 11과 Fig 12는 각각 데이터셋에 대한 Stability Deviation이다. CSRDCF_BD는 OTB100과 VisDrone 모두에서 가장 낮은 평균 AUC와 가장 작은 속성 간 편차를 기록하였으며, 이는 다양한 시각 조건 변화 속에서도 일관되고 강건한 추적이 가능함을 실증적으로 보여준다.

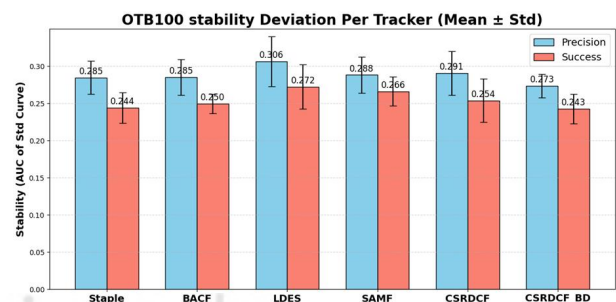


Fig. 11. OTB100 Stability Deviation

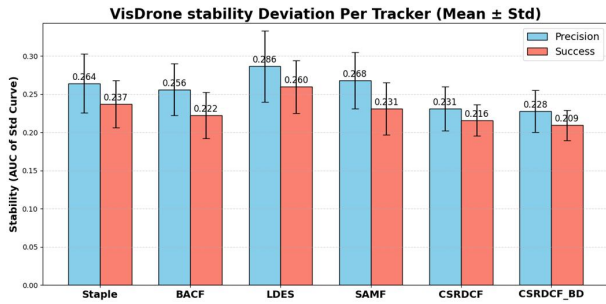


Fig. 12. VisDrone Stability Deviation

3. Limitations

제한된 CSRDCF_BD는 다양한 시각적 조건에서 견고한 성능을 보였지만, 다음과 같은 한계가 존재한다.

첫째, OCC 상황에서는 추적 성능이 급격히 저하될 수 있다. CSRDCF_BD는 foreground-background 간의 feature 유사도에 기반하여 마스크 신뢰도를 산정하고, 신뢰도가 낮은 경우 제약을 완화함으로써 필터가 주변 배경까지 포함하여 학습할 수 있도록 한다. 하지만 객체가 프레임 내에서 완전히 가려지거나 배경과 시각적으로 완전히 융합되는 경우, foreground와 background 간의 차이 자체가 사라지게 된다. 이 경우 마스크는 신뢰도를 거의 부여하지 못하며, 필터 학습을 위한 정보도 사실상 소실된다.

Fig 13은 CSRDCF_BD가 객체 완전 가림 이후 유효한 필터 업데이트에 실패하는 전형적인 사례를 제시한다. 초기 프레임에서는 비교적 뚜렷한 마스크와 객체 경계가 확인되지만, 객체가 점차적으로 가려지며 이후 프레임에서는 foreground 영역과 background가 시각적으로 거의 동일해지고, 마스크 confidence 또한 전반적으로 낮아진다. 이처럼 정보 자체가 소실된 경우, CSRDCF_BD는 더 이상

유효한 필터 업데이트를 수행할 수 없으며, 결과적으로 객체를 상실하게 된다.

둘째, 본 연구의 BD 기반 보정은 공간에서의 색상 히스토그램의 유사성만을 활용하므로, 시간적 예측이나 재진입 객체의 위치 복원 능력은 갖고 있지 않다. 따라서 장기 occlusion 이후 객체가 다시 등장하는 경우, 탐색 영역이 충분히 확보되지 않으면 복원이 어렵고, 성능이 급격히 저하됨을 확인할 수 있다.

이와 같이 CSRDCF_BD는 가림, 배경 간섭 등 복잡한 조건에서도 안정적인 성능을 유지하였으며, 그 한계 또한 명확히 드러남으로써 추후 개선 방향을 제시할 수 있는 기초를 제공한다.

4. Future Work

CSRDCF_BD는 전반적인 성능 향상뿐 아니라 다양한 환경에서의 견고한 동작을 통해 분할 기반 DCF 구조의 유효성을 다시금 부각시켰다. 그러나 마스크 품질이 본질적으로 낮거나 객체가 시야에서 완전히 소실되는 경우에는 신뢰도 기반 제약 구조만으로는 학습 오류를 방지하기에 충분하지 않다.

따라서 향후에는 다음과 같은 확장이 필요하다:

- 신뢰도 판단 기준의 다변화: BD 다양한 분포 유사도 척도의 도입 가능성 검토
- 시간적 일관성 기반 신뢰도 보정: optical flow, objectness score 등을 활용한 보조 신뢰도 지표 설계
- Fail-safe learning 구조: 분할 마스크 품질이 기준 이하일 경우 fallback strategy를 통해 DCF 필터 학습을 보완하는 전략 고려

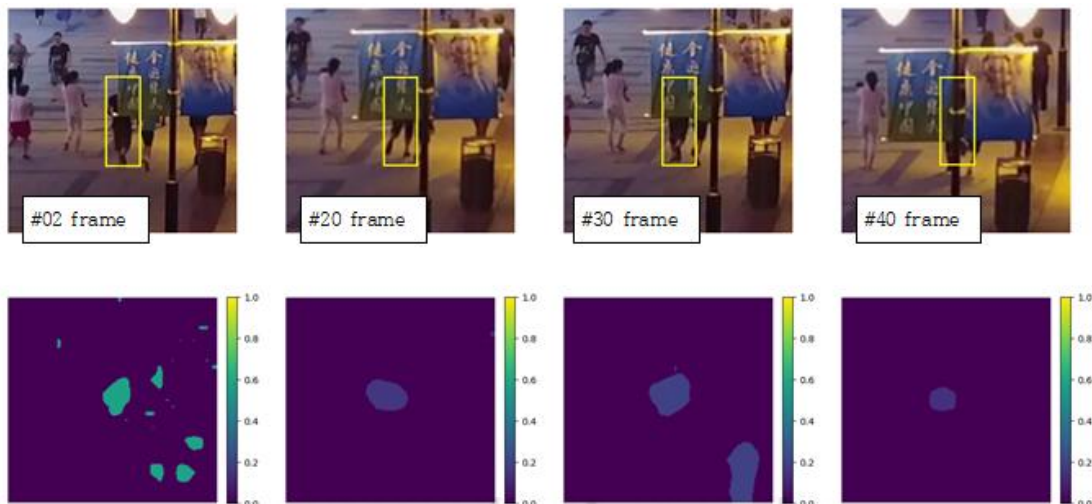


Fig. 13. CSRDCF Fail Example VisDrone uav000074_05712_s

이러한 방향은 CSRDCF_BD의 강점을 유지하면서도 그 약점을 보완하는 보다 강건한 분할 적응형 추적기 구조로의 발전을 가능하게 할 것이다.

VI. Conclusion

본 논문에서는 기존의 CSRDCF 추적기 구조를 확장하여, 마스크의 공간적 신뢰도를 보정하는 CSRDCF_BD를 제안하였다. 제안된 기법은 마스크의 품질이 저하되거나 foreground와 background 간 경계가 불분명한 상황에서도 필터 학습의 안정성을 유지할 수 있도록 설계되었다.

OTB100과 VisDrone 데이터셋을 대상으로 한 정량적 실험을 통해, CSRDCF_BD는 기존 CSRDCF 및 주요 CF 기반 추적기들 대비 높은 Precision와 Success를 기록하였으며, 특히 OCC, IV, BC과 같은 도전적인 조건에서 우수한 성능을 나타냈다. 또한 Stability Deviation 분석을 통해, CSRDCF_BD는 시간에 따른 성능의 일관성 측면에서도 가장 안정적인 추적 특성을 보이는 것으로 확인되었다.

한편, 완전한 가림이나 객체-배경 간 특성 차이가 극도로 약화되는 상황에서는 BD 기반 마스크 신뢰도 추정의 한계가 존재하였으며, 시간적 예측 능력의 부재로 인해 장기 OCC 상황에서의 복원 성능 또한 제한적이었다. 이와 같은 한계를 보완하기 위해, 향후 연구에서는 BD 기반 공간 필터링을 시간적 정보와 통합하거나, 더욱 정밀한 분포 기반 거리 척도의 안정적 적용 가능성을 모색할 필요가 있다.

결론적으로, CSRDCF_BD는 기존 CSRDCF의 장점을 유지하면서도 공간적 신뢰도에 대한 보다 정교한 제어를 가능하게 함으로써, 복잡한 시각 조건에서도 향상된 성능과 안정성을 제공할 수 있는 효과적인 확장 기법임을 실증적으로 확인하였다.

ACKNOWLEDGEMENT

This work was supported by Agency for Defense Development of Korea in 2025.

REFERENCES

- [1] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2544-2550, Jun. 2010. DOI: 10.1109/CVPR.2010.5539960
- [2] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 37, No. 3, pp. 583-596, Mar. 2015. DOI: 10.1109/TPAMI.2014.2345390
- [3] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," British Machine Vision Conference (BMVC), pp. 1-11, Sep. 2014.
- [4] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Discriminative scale space tracking," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 39, No. 8, pp. 1561-1575, Aug. 2017. DOI: 10.1109/TPAMI.2016.2590494
- [5] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "High performance visual tracking with Siamese region proposal network," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8971-8980, Jun. 2018. DOI: 10.1109/CVPR.2018.00935
- [6] X. Chen, B. Yan, J. Zhu, D. Wang, X. Yang, and H. Lu, "Transformer Tracking," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8126-8135, Jun. 2021. DOI: 10.1109/CVPR46437.2021.00804
- [7] S. Sahin, A. Djelouah, and C. Schroers, "Zero-Shot Object Tracking via Transferable Visual Prompting," arXiv preprint arXiv:2304.00590, Apr. 2023. [Online]. Available: <https://arxiv.org/abs/2304.00590>
- [8] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 5, pp. 564-577, May 2003. DOI: 10.1109/TPAMI.2003.1195991
- [9] M. Michailovich, Y. Rathi, and A. Tannenbaum, "Image segmentation using active contours driven by the Bhattacharyya gradient flow," IEEE Transactions on Image Processing, Vol. 16, No. 11, pp. 2787-2801, Nov. 2007. DOI: 10.1109/TIP.2007.906265
- [10] T. Xu, Z. Bian, J. Chen, L. Ma, and W. Cai, "Adaptive Spatio-Temporal Regularized Correlation Filters for UAV-Based Tracking," Applied Intelligence, Vol. 52, No. 5, pp. 5073-5088, 2022. DOI: 10.1007/s10489-021-02825-1
- [11] Z. Bian, T. Xu, J. Chen, L. Ma, W. Cai, and J. Li, "Auto-Learning Correlation-Filter-Based Target State Estimation for Real-Time UAV Tracking," Remote Sensing, Vol. 14, No. 21, Article 5299, 2022. DOI: 10.3390/rs14215299
- [12] Y. Wu, J. Lim, and M. Yang, "Object tracking benchmark," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 37, No. 9, pp. 1834-1848, 2015. DOI: 10.1109/TPAMI.2014.2388226
- [13] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker

- with feature integration," European Conference on Computer Vision Workshops (ECCV-W), pp. 254-265, 2014. DOI: 10.1007/978-3-319-16178-5_18
- [14] H. Kiani Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," IEEE International Conference on Computer Vision (ICCV), pp. 1135-1143, 2017. DOI: 10.1109/ICCV.2017.126
- [15] Y. Li, J. Zhu, and S. Hoi, "Reliable patch trackers: Robust visual tracking by exploiting reliable patches," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 353-361, 2015. DOI: 10.1109/CVPR.2015.7298634
- [16] M. Li, J. Zhu, and S. Hoi, "Target-aware deep tracking," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13625-13634, 2019. DOI: 10.1109/CVPR.2019.01394
- [17] Fengyang95 pyCFTrackers, <https://github.com/fengyang95/pyCFTrackers/tree/master>
- [18] P. Zhu, L. Wen, X. Bian, H. Ling, and Q. Hu, "VisDrone-VID: The Vision Meets Drone Video Detection and Tracking Challenge," International Journal of Computer Vision (IJCV), Vol. 129, No. 11, pp. 2895-2911, 2021. DOI: 10.1007/s11263-021-01516-4
- [19] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," IEEE Transactions on Image Processing, Vol. 2, No. 4, pp. 429-441, Oct. 1993. DOI: 10.1109/83.242361

Authors



Jung-Min Song received the B.S. degree in Electrical and Electronic Engineering from Kobe University and the M.S. degree in the same field from the University of Tokyo. He is currently a researcher at the Agency for

Defense Development. His research interests include EO/IR/SAR image processing and AI.