

Failure-State Assessment and Transition Estimation for Tracking-by-Detection Trackers under Composite Detection-Output Perturbations

Young-Jun Oh*

*CTO, LK ROBOTICS Inc., Daejeon, Korea

[Abstract]

In this paper, the Failure-State Assessment and Transition Estimation Scheme(FATES) for Tracking-by-Detection is proposed. FATES transforms degradation in MOT metrics into multi-label failure compositions and perturbation-space transition intervals using four labels: recall loss, fragmentation, identity-association failure, and localization instability. Rules selected from development data are evaluated against predefined operational event references using 1,260 synthetic runs and 45 window-tracker samples from MOT17, MOT20, and DanceTrack. On synthetic data, agreement with these references yielded 96.8% Macro-F1, 96.9% Micro-F1, and 91.1% exact match. External evaluation was limited to supported failure states and did not assess transitions. An offline application improved mean IDF1 from 64.0% to 67.8% and MOTA from 53.9% to 69.9% on 630 synthetic outputs. On 45 external samples, the gains were limited: IDF1 increased from 55.4% to 55.6% and MOTA from 50.9% to 51.3%.

▶ **Key words:** Multi-object tracking, Tracking-by-detection, Detection perturbation, Tracking failure analysis, Robustness evaluation

[요 약]

본 논문에서는 검출 기반 추적기의 성능지표 악화를 다중 실패 구성과 교란 공간의 상태 전이 구간으로 변환하는 Failure-State Assessment and Transition Estimation Scheme(FATES)를 제안한다. FATES는 네 가지 실패 라벨을 정의하고 개발 자료에서 결정한 규칙을 고정하여 평가한다. 1,260 개 합성 표본에서 사전 정의한 운용상 사건 참조 기준과의 일치도는 Macro-F1 96.8%, Micro-F1 96.9%, 활성 집합 완전 일치율 91.1%였다. 실제 영상 45개 표본의 Macro-F1은 65.6%였으나 지원 표본이 존재한 상태만 평가했으며 상태 전이는 검증하지 않았다. 오프라인 보정에서 합성 출력의 IDF1은 64.0%에서 67.8%, MOTA는 53.9%에서 69.9%로 향상되었다. 외부 표본에서는 각각 55.4%에서 55.6%, 50.9%에서 51.3%로 소폭 증가하여 실제 영상의 보정 효과는 제한적으로 나타났다.

▶ **주제어:** 다중 객체 추적, 검출 기반 추적, 검출 결과 교란, 추적 실패 분석, 강건성 평가

• First Author: Young-Jun Oh, Corresponding Author: Young-Jun Oh
*Young-Jun Oh (yjoh@lkrobotics.co.kr), LK ROBOTICS Inc.
• Received: 2026. 08. 04, Revised: 2026. 09. 29, Accepted: 2026. 09. 30.

I. Introduction

다중 객체 추적(Multiple Object Tracking, MOT)은 연속 영상에서 여러 객체의 위치와 식별정보를 유지하며 이동 궤적을 생성하는 기술이다. 검출 기반 추적(Tracking-by-Detection, TBD)은 프레임별 검출 결과와 기존 궤적 간 데이터 연관을 수행하며, SORT, ByteTrack 및 OC-SORT가 대표적인 방법이다[1-3]. 영상 관제 및 이동 로봇 환경에서는 궤적 단절이나 식별정보 전환(Identity Switch, IDSW)이 발생할 경우 이동 경로와 객체별 행동에 대한 해석이 달라질 수 있다. 따라서 추적 성능의 저하 정도뿐 아니라 어떠한 실패 양상이 발생했는지를 함께 분석할 필요가 있다.

Multiple Object Tracking Accuracy(MOTA)와 Identification F1-score(IDF1)는 추적 오류와 식별정보 일치도를 정량적으로 평가하는 대표 지표이다[4-5]. 그러나 개별 지표만으로는 여러 지표가 동시에 악화될 때 나타나는 실패 상태의 구성을 파악하거나, 검출 오류 강도에 따라 실패 상태가 변화하는 구간을 직접 확인하기 어렵다. 예를 들어 IDF1 감소만으로는 반복적인 궤적 단절과 ID 연관 실패를 명확히 구분하기 어렵다. 여러 추적기와 복합 오류 조건을 동일한 기준에서 해석하기 위해서는 지표 악화를 실패 상태와 전이 구간으로 변환하는 별도의 진단 절차가 필요하다.

본 논문에서는 정상 조건 대비 MOT 지표의 악화량을 다중 실패 상태의 활성 집합과 교란 공간의 전이 구간으로 변환하는 Failure-State Assessment and Transition Estimation Scheme(FATES)를 제안한다. FATES는 추적 결과의 지표 변화를 기반으로 실패 구성을 판정하고, 사전에 고정한 운용상 사건 참조 기준과의 일치도로 판정 결과를 평가한다. 실제 영상 평가는 지원 표본이 존재한 실패 상태의 제한적 외부 평가로 수행하며, 통제된 교란 강도가 필요한 상태 전이 평가는 합성 환경으로 한정한다.

II. Related Work

1. MOT and association

SORT는 칼만 필터와 IoU 기반 데이터 연관을 이용하여 객체를 추적한다[1]. ByteTrack은 고·저 신뢰도 검출 결과를 단계적으로 연관하며, OC-SORT는 관측값을 활용하여 운동 추정 오류에 대응한다[2-3]. 이러한 데이터 연관 과정에서 객체와 추적 ID의 대응이 유지되지 않으면

IDSW나 궤적 단절이 발생하며, 이는 전체 궤적의 연속성과 식별정보 일관성에 영향을 미친다.

최근에는 검출 실패와 연관 오류에 대응하여 추적 안정성을 향상시키는 연구가 수행되고 있다. Building Unmatched trajectories Capitalizing on Attention(BUSCA)은 검출이 누락된 구간에서 문맥 정보를 활용하여 기존 궤적을 유지한다[6]. 이러한 연구가 추적 과정의 오류 감소에 초점을 두는 반면, FATES는 추적 결과의 지표 변화를 이용하여 실패 상태의 구성과 변화 구간을 분석한다.

2. Evaluation and Failure Diagnosis

CLEAR MOT는 오검출(False Positive, FP), 미검출(False Negative, FN) 및 IDSW를 기반으로 추적 성능을 평가하며[4], ID 지표는 정답 궤적과 추적 궤적 간의 식별정보 일치도를 평가한다[5]. Detection-aware 평가는 입력 검출 품질을 고려하여 추적 과정의 성능 기여도를 분석한다[7]. 이러한 방법은 오류의 총량을 제시하지만 복수 지표가 동시에 악화될 때의 실패 구성을 직접 판정하지 않는다.

추적 실패를 직접 분석하기 위한 연구에서는 시간에 따라 누적되는 관측 정보를 이용하여 실패 발생 여부를 판정하는 방법도 제안되었다[8]. FATES는 시간축의 실패 발생 시점을 검출하는 방식과 달리, 정상 조건 대비 복수 MOT 지표의 악화를 다중 실패 활성 집합으로 구성하고 인접한 교란 조건을 비교하여 상태 전이 구간을 산정한다.

3. Perturbation and Interpretable Assessment

MOT17, MOT20 및 DanceTrack과 같은 공개 벤치마크는 다양한 객체 움직임과 밀집 환경에서 추적 성능을 평가하는 데 활용된다[9-11]. 그러나 실제 영상에서는 미검출, 오검출, 위치 오차와 같은 개별 오류의 강도를 독립적으로 조절하기 어렵다. 이에 본 연구에서는 검출 삭제, 오검출 추가, 위치 잡음, 신뢰도 감소 및 연속 가림을 적용하여 단일 및 복합 교란 조건을 구성한다.

FATES는 교란 조건에 따른 MOT 지표의 악화량을 명시적인 판정 규칙으로 변환하여 다중 실패 구성을 산정하고, 인접 조건 간 상태 변화를 통해 전이 구간을 추정한다. 또한 단일 지표 기반 방법 및 학습 기반 분류기와 비교하고, 실제 영상 데이터에 동일한 판정 규칙을 적용하여 실패 상태 판정의 특성과 적용 가능성을 평가한다.

III. Proposed Method

1. Overview and notation

FATES는 정답 궤적 G , 이상 검출 D , 교란 조건 q 와 추적기 a 를 입력으로 받는다. 교란 검출에 대한 출력 Y 를 G 와 비교한 뒤, 정상 출력의 지표 $m(a,0)$ 에 대한 악화량을 계산한다. Fig. 1의 두 경로는 실패 판정과 사건 참조 생성이며, 하단은 판정에 조건부인 사후 보정이다.

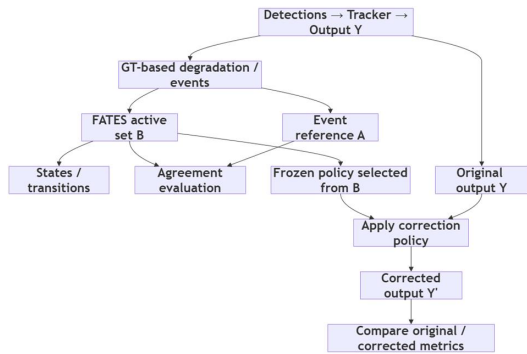


Fig. 1. FATES assessment and its offline correction application.

두 경로는 추적 출력 Y 와 정답 궤적 G 의 프레임 대응 정보만 공유한다. FATES 판정 경로는 정상 조건 대비 집계 MOT 지표의 상대적 악화량과 훈련·개발 검증에서 고정한 예측 임계값을 사용한다. 사건 참조 경로는 미매칭, 재획득 간격, IDSW, 오연관 및 IoU 저하를 프레임·사건 단위로 집계하고 Table 3의 사전 고정 기준을 적용한다. 시험 및 실제 영상에서는 어느 경로도 재조정하지 않았다.

Table 1. Main notation.

Symbol	Meaning
G, D, Y	GT, ideal detections, output
$q; h$	Detection perturbation; trajectory setting
d_j, θ_j, z_j	Degradation, threshold, normalized value
$K; B_a(q); F_a(q)$	Label set; active set; summary state
$A_a(q); E(Q); R_a$	Reference set; grid edges; transition intervals

2. Perturbations and degradation

검출 교란 벡터는 식 (1)과 같이 미검출 확률, 프레임당 오검출 수, 위치 잡음, 신뢰도 감소 및 연속 가림 길이로 정의한다. 단일 교란은 한 요소만, 복합 교란은 둘 이상의 요소를 변경한다. 객체 수와 속도는 궤적 환경 h 로 분리한다. 정답 경로를 바꾸는 객체 교차율은 검출 교란 벡터에 포함하지 않는다.

$$q = (p_{miss}, \lambda_{fp}, \sigma_{box}, r_{conf}, L_{occ}) \quad (1)$$

큰 값이 우수한 성능을 의미하는 지표 집합을 J^+ , 큰 값이 성능 저하를 의미하는 지표 집합을 J^- 로 정의하면, 식 (2)와 같이 기준 조건 대비 부호 있는 악화량을 산출할 수 있다. 정상 조건보다 성능이 향상된 경우 악화량은 음수로 나타나며, 성능이 저하된 경우에는 양수로 증가한다. 이후 각 지표의 악화량을 해당 임계값으로 정규화하여 임계값 대비 상대적 크기를 표현하며, 이 값은 확률적 의미를 갖지 않는다.

$$d_j(a, q) = \begin{cases} m_j(a, 0) - m_j(a, q) & j \in J^+ \\ m_j(a, q) - m_j(a, 0) & j \in J^- \end{cases} \quad (2)$$

식 (3)은 지표별 악화량 d_j 를 판정 임계값 θ_j 로 정규화한 값 z_j 를 나타낸다. $z_j \geq 1$ 이면 임계값 이상의 성능 저하를, $z_j \leq 1$ 이면 임계값 미만의 변화를 의미한다.

$$z_j(a, q) = \frac{d_j(a, q)}{\theta_j}, \quad \theta_j > 0 \quad (3)$$

3. Failure composition and rule selection

실패 상태 집합 K 는 재현율 저하(R), 궤적 단절(F), ID 연관 실패(I), 위치 불안정(L)으로 구성한다. 식 (4)는 교란 조건 q 에서 활성화된 실패 상태 집합 $B_a(q)$ 를 정의하며, 식 (5)는 이를 Normal, 단일 실패 상태 또는 Mixed로 구분한다.

$$B_a(q) = \{k \in K : C_k(z) = 1\} \quad (4)$$

$$F_a(q) = \begin{cases} N & |B_a(q)| = 0 \\ k & |B_a(q)| = 1 \\ \text{Mixed} & |B_a(q)| \geq 2 \end{cases} \quad (5)$$

상태별 후보 지표는 R에 recall drop, F에 fragmentation-rate increase, I에 IDSW-rate increase·IDF1 drop·FP-rate increase, L에 mean-IoU drop을 사용하였다. 각 지표의 임계값 후보는 훈련 자료의 0, 1/30, ..., 1 분위수에서 생성하고 중복값과 0 이하 값은 제외하였다. 단일 지표 조건과 서로 다른 지표 두 개의 AND·OR 조합을 개발 검증 세트에서 비교하여 클래스별 F1-score가 최대인 규칙을 선택하였다. 동률인 경우 조건 수가 적은 규칙을 우선하였으며, 최종 규칙과 임계값은 시험 전에 고정하였다.

Table 2. Selected activation thresholds.

Label	Degradation $\geq$ threshold
R	Recall drop ≥ 0.102
F	Gaps/object increase ≥ 0.400
I	IDSW/matches increase ≥ 0.007500
L	Mean IoU drop ≥ 0.277428

4. Event reference and transition intervals

사건 참조 A 는 프레임별 정답과 추적 결과의 대응에서 매칭, 재획득 간격, IDSW, 오연관 및 IoU 저하를 집계하여 생성한다. 상태별 참조 라벨은 Table 3의 기준에 따라 부여하며, 장기 손실은 15프레임 이상, 낮은 IoU는 0.65 미만으로 정의한다. Table 3의 사건 참조 임계값은 시험 성능을 최대화하여 추정된 값이 아니라, 초기 개발 단계에서 실질적인 손실·단절·오연관 및 위치 저하를 구분하기 위한 운용 기준으로 사전 정의하였다. 이 기준은 시험 전에 고정하여 합성 시험과 실제 영상에 동일하게 적용하였으며, 통계적 최적값이나 독립적인 전문가 정답을 의미하지 않는다. 예측 임계값과 사건 참조 임계값의 영향을 혼합하지 않기 위해 두 집합을 각각 0.8~1.2배로 변화시켜 분석하였다.

Table 3. Operational event-reference conditions.

Label	Activation condition (OR unless noted)
R	Unmatched fraction > 0.15; long-loss object fraction ≥ 0.30
F	Reacquisition gaps/object ≥ 0.50
I	Consecutive ID changes/object ≥ 0.30 ; (wrong matches ≥ 5 AND impurity ≥ 0.02)
L	Matched > 0 AND (low-IoU fraction ≥ 0.25 OR mean IoU < 0.70)

전이 구간은 동일한 궤적, 난수 시드 및 추적기에서 다른 조건을 고정한 채 한 교란 축의 인접 지점 q_i 와 q_j 를 비교하여 산정한다. 두 지점의 요약 상태가 다르면 해당 간선을 전이로 정의한다. Mixed 상태 내부의 구성 변화는 활성 집합 B 를 통해 별도로 확인한다.

$$R_a = \{e_{ij} \in E(Q) : F_a(q_i) \neq F_a(q_j)\} \quad (6)$$

5. Failure-conditioned offline correction

사후 보정은 FATES의 활성 실패 상태에 따라 짧은 누락 구간 보간, 위치 평활 및 단기 궤적 제거를 선택적으로 적용한다. 보간은 동일 ID의 누락 구간이 8프레임 이하인 경우에 적용하고, 위치 평활은 현재 관측값 가중치 0.65의 지수 평활을 사용하며, 4프레임 미만의 짧은 궤적은 제거한다. 세 보정 연산의 조합과 무보정을 포함하여 총 8개 정책을 구성하며, 출력 Y 에만 적용한다. 정책은 훈련 세트에서 추적기와 활성 집합별 평균 $\frac{IDF1 + MOTA}{2}$ 가 최대가 되도록 선택한다. 표본 수가 3개 미만인 경우에는 무보정을 적용하며, 동일한 성능에서는 연산 수가 적은 정책을 우선한다. 개발 검증 세트에서 성능 개선이 확인된 정책만

유지하고, 최종 정책은 시험 전에 고정한다.

IV. Experiments and Results

1. Experimental setup and data split

합성 환경은 1000×1000 pixel, 300프레임으로 구성하였다. 표준 환경은 객체 10개와 속도 1.5~5 pixel/frame, 밀집·고속 환경은 객체 20개와 속도 3~8 pixel/frame으로 설정하였으며, 방향 잡음은 각각 0.12와 0.24를 적용하였다. 검출 교란은 미검출, 오검출, 위치 잡음, 신뢰도 감소 및 연속 가림으로 구성하였고, 정상 조건과 단일·복합 조건을 포함하였다.

훈련에는 난수 시드 401, 409, 419를, 개발 검증에는 431, 433, 439를 사용하였다. 합성 평가는 1009, 1013, 1019, 1021, 1031, 1033의 6개 시드와 두 환경, 35개 조건, 세 추적기로 구성된 1,260개 실행 결과를 사용하였다. 실제 영상은 사후적인 유리 사례 선택을 방지하기 위해 기존 FATES 공개 분석에 사용하지 않은 가용 시퀀스 중 시퀀스 ID 순서가 앞선 MOT17-02/05-FRCNN, DanceTrack0004/0007 및 MOT20-03을 결과 확인 전에 고정하여 선정하였다. 각 시퀀스의 최초 300프레임을 서로 겹치지 않는 100프레임 세 구간으로 나누고 각 구간에서 추적기를 초기화하였다. 데이터셋별 표본은 18, 18, 9개이며, 추적기별로는 각각 15개로 총 45개이다. 공개 검출 결과로 추적 출력을 생성하고 GT 기반 ideal detection으로 정상 기준을 만들었으며, 합성 시험과 같은 사건 참조 규칙을 적용하였다.

비교 방법으로 Primary-only, Initial-rule, CART 및 Random Forest를 사용하였으며, 동일한 입력 지표와 데이터 분할을 적용하였다. 보정 실험은 별도의 난수 시드 2003, 2011, 2017을 사용한 630개 출력에서 수행하였으며, 무보정, 고정 보정, 무작위 보정 및 FATES 기반 선택 보정을 비교하였다. 보정 성능은 motmetrics 1.4를 이용하여 IDF1, MOTA, IDSW 및 궤적 단절로 평가하였다.

2. Tracking performance with and without FATES-conditioned correction

FATES 판정에 따른 보정 효과를 확인하기 위해 독립 시드 3개로 생성한 630개 신규 합성 출력에서 무보정, 고정 보정, 무작위 보정 및 FATES 기반 선택 보정을 비교하였다. Fig. 2와 Table 4는 시드별 평균과 표준편차를 함께 제시한다. FATES 기반 선택 보정은 IDF1 67.8±0.8%,

MOTA $69.9 \pm 1.4\%$ 로 가장 높았고, 궤적 단절 감소량도 115.5 ± 5.9 events/output으로 가장 컸다. $\Delta\text{Frag.}$ 는 무보정 단절 수에서 보정 후 단절 수를 뺀 값이며 양수일수록 개선을 의미한다.

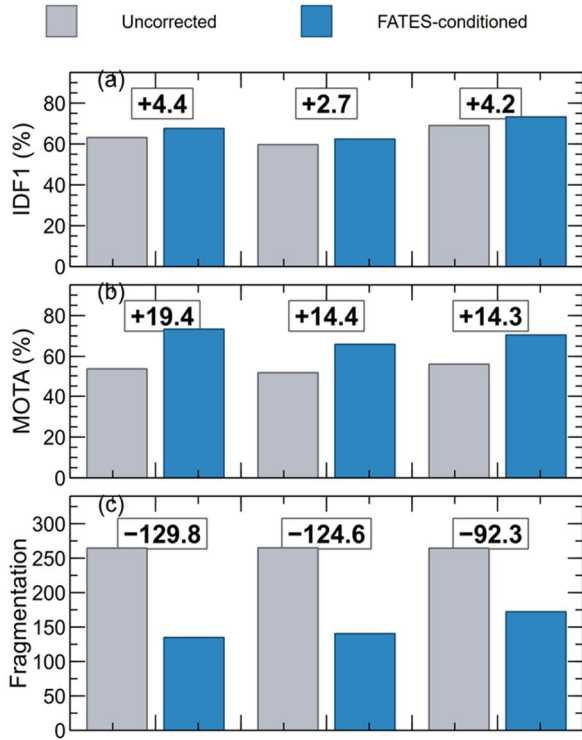


Fig. 2. Correction results by tracker (x-axis): (a) IDF1 (%), (b) MOTA (%), and (c) fragmentation events.

Table 4. Tracking performance by policy (IDF1/MOTA: %, $\Delta\text{Frag.}$: events/output; mean $\pm$ SD).

Policy	IDF1	MOTA	$\Delta\text{Frag.}$
None	64.0 $\pm$ 0.9	53.9 $\pm$ 2.4	0.0 $\pm$ 0.0
Fixed	66.8 $\pm$ 0.8	68.5 $\pm$ 1.6	58.2 $\pm$ 3.1
Random	65.8 $\pm$ 0.9	57.4 $\pm$ 1.7	83.1 $\pm$ 12.9
FATES	67.8 $\pm$ 0.8	69.9 $\pm$ 1.4	115.5 $\pm$ 5.9

Fig. 3은 위치 잡음(pixel)을 x축으로 하여 보정 전·후 MOTA와 ΔMOTA 를 분리해 제시한다. 위치 잡음 10 및 20 pixel에서 세 추적기 모두 양의 ΔMOTA 를 보여, FATES 기반 선택 보정이 해당 교란에서 성능 저하를 완화함을 확인하였다.

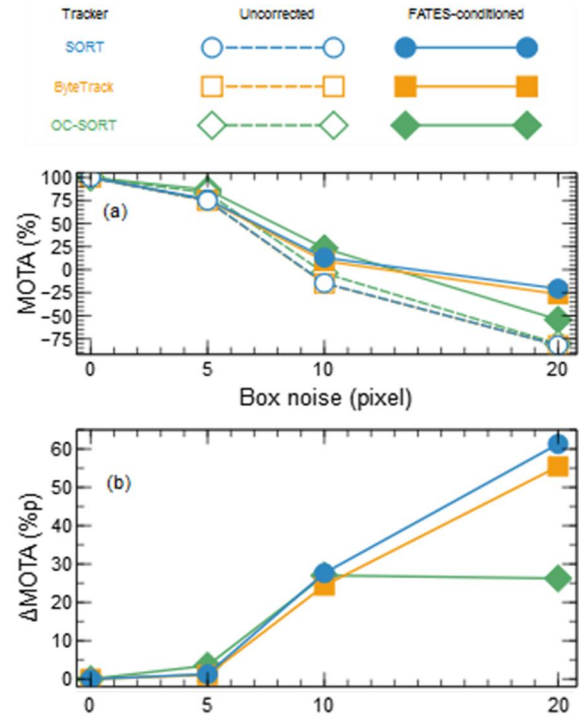


Fig. 3. Bounding-box noise (x-axis, pixels): (a) MOTA (%) before/after correction and (b) ΔMOTA (%p).

3. Failure-State Assessment and Baseline Comparison

실패 상태 판정은 사전에 고정한 운용상 사건 참조 기준과의 일치도로 평가하였다. 1,260개 합성 표본에서 FATES는 Macro-F1 96.8%, Micro-F1 96.9%, EM 91.1%, transition F1 89.1%를 기록하였다. 시드별 평균 $\pm$ 표준편차는 Macro-F1 $96.7 \pm 1.0\%$, EM $91.1 \pm 1.4\%$, transition F1 $89.1 \pm 2.3\%$ 였다. 상태별 F1은 재현율 저하 90.9%, 궤적 단절 99.6%, ID 연관 실패 99.8%, 위치 불안정 97.0%였다. 이 수치는 독립 전문가 정답에 대한 인식 정확도가 아니라 운용상 사건 기준과의 일치도를 의미한다.

6개 시험 시드에서 CART와 Random Forest의 Macro-F1은 각각 $97.8 \pm 2.2\%$ 와 $97.6 \pm 2.0\%$, transition F1은 $100.0 \pm 0.0\%$ 와 $99.8 \pm 0.4\%$ 로 FATES보다 높았다 (Table 5). FATES의 장점은 최고 분류성능이 아니라 상태별 입력 지표, 임계값 및 활성 근거를 명시적으로 추적할 수 있다는 점이다. 동일 시험 시드에서 산출한 평균과 표준편차를 함께 제시하여 방법 간 성능 차이의 안정성을 비교하였다.

Table 5. Synthetic comparison across six test seeds (mean $\pm$ SD; %).

Method	Macro-F1	Micro-F1	EM	Tr-F1
FATES	96.7 ± 1.0	96.9 ± 0.7	91.1 ± 1.4	89.1 ± 2.3
CART	97.8 ± 2.2	98.0 ± 1.8	94.8 ± 4.1	100.0 ± 0.0
Random Forest	97.6 ± 2.0	97.8 ± 1.6	94.2 ± 3.5	99.8 ± 0.4
Initial rule	82.4 ± 2.2	83.5 ± 2.3	61.4 ± 0.8	85.2 ± 4.1
Primary only	88.9 ± 2.2	88.5 ± 1.8	64.0 ± 3.1	67.9 ± 3.1

4. Threshold Sensitivity and Transition Estimation

예측 임계값과 사건 참조 임계값을 별도로 0.8, 0.9, 1.0, 1.1, 1.2배 변화시켰다(Table 6). 예측 임계값 변화의 Macro-F1은 95.0, 95.3, 96.8, 93.4, 93.7%였고, 사건 참조 임계값 변화에서는 96.0, 96.2, 96.8, 98.0, 96.3%였다. 기준값에서 최고 또는 이에 근접한 성능을 보였으나 예측 임계값 1.1배 이상에서는 성능 저하가 나타나므로 데이터셋 전이 시 재검증이 필요하다.

Table 6. Macro-F1 sensitivity to prediction/reference thresholds.

Scale	Prediction	Reference	Unit
0.8	95.0	96.0	%
0.9	95.3	96.2	%
1.0	96.8	96.8	%
1.1	93.4	98.0	%
1.2	93.7	96.3	%

5. Generalization and Computational Cost

고정된 FATES 규칙을 실제 영상 45개 표본에 적용한 결과는 Table 7과 같다. 전체 Macro-F1은 65.6%였으나 DanceTrack과 MOT20은 각각 48.1%와 25.0%였다. 상태별로는 재현을 저하, 꺾적 단절 및 ID 연관 실패의 양성 표본이 각각 27, 21, 17개였고, 위치 불안정은 양성 표본이 없어 검증하지 못하였다. 실제 영상은 교란 강도 순서가 아니므로 상태 전이는 평가하지 않았다.

외부 영상에서 FATES 기반 보정 후 IDF1은 55.4%에서 55.6%, MOTA는 50.9%에서 51.3%로 각각 0.2%p와 0.4%p 증가하는 데 그쳤다. 45개 중 25개(55.6%)에서 무보정 정책이 선택되어 현재 실제 영상의 보정 효과는 제한적이다. 이는 실제 오류가 단일 정책으로 제거되지 않거나 보수적 선택 규칙이 개입을 제한했기 때문이다. 분류 처리 시간은 1,260개 표본 기준 0.012 ms였으나 본 평가는 GT에 의존하는 오프라인 진단이다.

Table 7. External results (data: N/Macro-F1; state: Pos./Neg./F1).

Type	Item	N/Pos.	Neg.	Score%
Data	MOT17	18	-	66.9
Data	DanceTrack	18	-	48.1
Data	MOT20	9	-	25.0
Data	Overall	45	-	65.6
State	Recall	27	18	94.7

V. Discussion

합성 시험의 Macro-F1은 96.8%였으나 실제 영상은 65.6%(DanceTrack 48.1%, MOT20 25.0%)였고 위치 불안정 양성 표본도 없었다. 따라서 외부 결과는 지원 표본이 있는 재현을 저하·단절·ID 연관 상태의 제한적 검증이며 네 상태 전체의 일반화를 의미하지 않는다.

CART와 RF는 합성 성능이 더 높았지만 FATES는 활성 지표와 임계값을 공개한다. Detection-aware evaluation은 검출 품질 기여, sequential hypothesis testing은 시간축 실패 시점을 출력하므로 FATES의 다중 활성 집합 전이 구간과 직접 정량 비교하기 어렵다. FATES의 기여는 성능 우위가 아니라 해석 가능성과 전이 분석이다.

대표 사례(dense_fast, seed 1013, SORT, box noise 5 pixels)에서 사건 참조와 FATES는 F+I, CART와 RF는 F+I+L을 출력하였다. FATES는 fragmentation increase $14.25 \geq 0.400$ 과 IDSW increase $0.0985 \geq 0.0075$ 로 F+I를 활성화하고, mean-IoU drop $0.2719 < 0.2774$ 로 L을 제외하였다. 사전 고정 정책표에서 SORT|FI는 정책 0(무보정)에 연결되므로, 판정부터 불필요한 개입을 피한 보정 결정까지 추적할 수 있다.

FATES와 사건 참조는 Y·G 대응을 공유하므로 독립 외부 정답이 아니다. 다만 각각 정상 대비 집계 지표와 프레임·사건 비율을 사용하고 기준을 시험 전에 고정하였다. 현재 범위는 GT 기반 오프라인 진단이며 온라인 적용에는 proxy 신호와 독립 임계값 검증이 필요하다.

VI. Conclusions

FATES는 정상 조건 대비 MOT 지표 악화를 재현을 저하, 꺾적 단절, ID 연관 실패 및 위치 불안정의 활성 집합으로 구성하고 교란 조건에 따른 상태 전이 구간을 산정한다. 합성 시험에서는 사전 운용 기준과 높은 일치도를 보

었다. 실제 영상에서는 지원 표본이 존재한 재현을 저하·
궤적 단절·ID 연관 실패에 대해 제한적인 외부 평가를 수
행했으며, 위치 불안정과 상태 전이의 일반화는 검증하지
못하였다.

FATES 판정 결과를 사후 보정 정책과 연계한 합성 시
험에서는 IDF1과 MOTA가 향상되어 진단 정보의 활용 가
능성을 확인하였다. 반면 실제 영상에서는 45개 중 25개
에서 무보정이 선택되었고, IDF1과 MOTA 증가폭도 각각
0.2%p와 0.4%p에 그쳐 현재 보정 방법의 실질적 효과는
제한적으로 나타났다.

향후에는 실제 데이터의 상태별 support를 확대하고
GT가 필요 없는 온라인 proxy와 적응형 임계값을 검증할
예정이다.

REFERENCES

- [1] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple Online and Realtime Tracking," Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 3464-3468, Phoenix, AZ, USA, Sep. 2016. DOI: 10.1109/ICIP.2016.7533003.
- [2] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," Proceedings of the European Conference on Computer Vision (ECCV), pp. 1-21, Tel Aviv, Israel, Oct. 2022. DOI: 10.1007/978-3-031-20047-2_1.
- [3] J. Cao, J. Pang, X. Weng, R. Khrodar, and K. Kitani, "Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9686-9696, Vancouver, Canada, Jun. 2023.
- [4] K. Bernardin and R. Stiefelagen, "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics," EURASIP Journal on Image and Video Processing, Vol. 2008, Article 246309, May 2008. DOI: 10.1155/2008/246309.
- [5] E. Ristani, F. Solera, R. S. Zou, R. Cucchiara, and C. Tomasi, "Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking," Proceedings of the European Conference on Computer Vision Workshops (ECCVW), pp. 17-35, Amsterdam, The Netherlands, Oct. 2016. DOI: 10.1007/978-3-319-48881-3_2.
- [6] L. Vaquero, Y. Xu, X. Alameda-Pineda, V. M. Brea, and M. Mucientes, "Lost and Found: Overcoming Detector Failures in Online Multi-Object Tracking," Proceedings of the European Conference on Computer Vision (ECCV), pp. 448-466, Milan, Italy, Sep.-Oct. 2024. DOI: 10.1007/978-3-031-73464-9_27.
- [7] J. C. SanMiguel, J. Muñoz, and F. Poiesi, "Detection-Aware Multi-Object Tracking Evaluation," Proceedings of the 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-8, Madrid, Spain, Nov.-Dec. 2022. DOI: 10.1109/AVSS56176.2022.9959412.
- [8] A. Monroy Muñoz, R. Verma, and A. Timans, "Detecting Object Tracking Failure via Sequential Hypothesis Testing," Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), pp. 171-180, Tucson, AZ, USA, Mar. 2026.
- [9] P. Dendorfer, A. Osep, A. Milan, K. Schindler, D. Cremers, I. Reid, S. Roth, and L. Leal-Taixé, "MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking," International Journal of Computer Vision, Vol. 129, pp. 845-881, Apr. 2021. DOI: 10.1007/s11263-020-01393-0.
- [10] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixé, "MOT20: A Benchmark for Multi Object Tracking in Crowded Scenes," arXiv preprint arXiv:2003.09003, Mar. 2020.
- [11] P. Sun, J. Cao, Y. Jiang, Z. Yuan, S. Bai, K. Kitani, and P. Luo, "DanceTrack: Multi-Object Tracking in Uniform Appearance and Diverse Motion," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 20993-21002, New Orleans, LA, USA, Jun. 2022. DOI: 10.1109/CVPR52688.2022.02032.

Authors



Young-Jun Oh received the B.S. degree in Information and Communication from Korea Nazarene University in 2009, and the M.S. degree in Electrical and Electronic Engineering and the Ph.D. degree in

Computer Engineering from KOREATECH in 2011 and 2016, respectively. Dr. Oh has worked in R&D and business planning and is currently the CTO and R&D Director at LK ROBOTICS Inc. His research interests include multiple object tracking, computer vision, sensor fusion, SLAM, autonomous driving, and robot perception.