DOI: https://doi.org/10.15276/ict.03.2026.16
UDC 004.932

Methodology for evaluating the robustness of multi-object tracking algorithms to input video stream degradation

Oleh V. Streltsov[1)]
ORCID: https://orcid.org/0000-0002-4691-5703; streltsov.o.v@op.edu.ua. Scopus Author ID: 57210373473
Maksym V. Polishchuk[1)]
ORCID: https://orcid.org/0009-0004-4425-9564; 4107624@op.edu.ua;
Maksym A. Hrynyov[1)]
ORCID: https://orcid.org/0009-0007-7189-4485; grinevvv2002@stud.op.edu.ua
[1)] Odesa Polytechnic National University, 1, Shevchenko Av. Odesa, 65044, Ukraine

ABSTRACT

The paper presents an extended methodology for evaluating the robustness of multi-object tracking algorithms under degradation of video streams in intelligent video surveillance systems. The study is relevant because, in real conditions, video data are affected by spatial, photometric, temporal and observability-related distortions, such as blur, noise, compression artefacts, reduced resolution, frame loss, lower frame rate and occlusions. These degradations reduce detector stability, increase missed detections and false alarms, fragment trajectories and cause identity switches. The methodology uses a modular experimental pipeline comprising video-data preparation, controlled degradation modelling, object detection, tracker execution, ground-truth adaptation and quality-metric calculation. Fair comparison is ensured through factor isolation: in each run, one degradation parameter is changed, while other conditions remain fixed. The method combines multi-object tracking metrics with robustness indicators. Tracking effectiveness is assessed by Multiple Object Tracking Accuracy, Higher Order Tracking Accuracy, Identification F1 Score, Identity Switches, False Positives, False Negatives, trajectory fragmentation and processing speed in frames per second. For quantitative robustness analysis, the paper introduces the Robustness Score and Breaking Point indicators, which characterize absolute tracking quality and the rate of its decline as distortion intensity increases. The methodology supports comparative analysis of trackers, including Simple Online and Realtime Tracking, Deep Simple Online and Realtime Tracking, ByteTrack, Observation-Centric Simple Online and Realtime Tracking and CenterTrack, and helps select solutions for intelligent video surveillance systems under limited video-stream quality.

Keywords: Multi-object tracking; video stream degradation; intelligent video surveillance

For citation: **Streltsov O. V., Polishchuk M. V., Hrynyov M. A. "Methodology for evaluating the robustness of multi-object tracking algorithms to input video stream degradation".** ***Informatics. Culture. Technology*****. 2026; Vol. 3 No. 1(3): 189–199. DOI: https://doi.org/10.15276/ict.03.2026.16**

Relevance. At the same time, the operating conditions of real video surveillance systems differ significantly from laboratory assumptions. A video stream can be degraded by insufficient illumination, adverse weather phenomena, camera vibration, motion or defocus blur, network-bandwidth limitations, aggressive compression, reduced resolution and frame loss. As a result, tracking algorithms may lose objects, create false trajectories, incorrectly switch object identifiers or produce fragmented tracks. For safety-critical systems, such failures directly reduce reliability and may lead to incorrect automated decisions.

Traditional evaluation of algorithms on benchmark datasets under relatively stable video-quality conditions does not fully answer the question of their applicability in real environments. An algorithm that demonstrates high accuracy on clean video may lose operational suitability when the input stream quality deteriorates. Therefore, it is necessary to evaluate not only baseline tracking quality but also the ability of an algorithm to preserve acceptable performance under controlled spatial and temporal degradation of the input video stream.

Thus, the scientific and practical problem consists in developing a reproducible methodology that formalizes the influence of video-stream degradation on multi-object tracking quality, compares algorithms by their robustness profiles and determines the limiting operating conditions under which a tracker remains practically usable. The purpose of the paper is to develop such a methodology for intelligent video surveillance systems.

The paper is structured as follows: the related works section reviews the current state of multi-object tracking and evaluation metrics; the research methodology section formalizes the proposed degradation-aware evaluation pipeline; the next sections describe the experimental protocol, robustness indicators, practical significance and limitations; the final section summarizes the conclusions.

One of the best-known baseline algorithms is SORT, which combines the Kalman filter for object-state prediction with the Hungarian algorithm for matching detections to active tracks [1].

The main advantage of SORT is its high processing speed and simple integration. However, the algorithm is vulnerable to occlusions, missed frames and visually similar objects, which frequently results in identity switches.

DeepSORT extends SORT by adding deep appearance features for object re-identification [2]. This improves identity preservation, especially in crowded scenes and during short-term occlusions. However, appearance features are sensitive to blur, compression artifacts, noise and low resolution, because such distortions reduce the discriminative quality of visual descriptors. Therefore, algorithms that perform well in standard benchmarks may become unstable when appearance descriptors are calculated from degraded image regions.

ByteTrack proposed an important modification of the association procedure by using not only high-confidence detections but also low-confidence detections [3]. This is particularly relevant under video degradation, when the detector confidence may decrease even for true objects. OC-SORT rethinks the motion component of SORT and introduces an observation-centric strategy to reduce error accumulation during occlusion and missing observations [4]. CenterTrack represents a neural approach in which object detection and tracking are treated jointly as point-based tracking [5].

Recent neural and transformer-based trackers, such as FairMOT, TrackFormer and MOTR, demonstrate that the field is moving from hand-crafted association rules to more integrated detection-association models [6], [7], [8]. However, more complex models are not automatically more reliable in surveillance practice. They may require more computational resources, larger training datasets and stable visual features. Therefore, a degradation-aware evaluation protocol is needed for both classical and neural trackers.

For evaluating multi-object tracking, the CLEAR MOT metrics, including MOTA and MOTP, are widely used [9]. However, these metrics do not always sufficiently reflect identity-preservation quality. Therefore, ID metrics such as IDF1 [10] and higher-order metrics such as HOTA [11] are increasingly applied. They allow the researcher to analyze detection quality and association quality more separately. Despite the development of benchmarks, many evaluations focus on results under predefined dataset conditions, while in real surveillance systems the video-stream quality may change dynamically. Hence, there is a need for a degradation-aware evaluation methodology that makes it possible to assess how tracking quality changes when input-video quality progressively deteriorates.

Purpose and objectives. The purpose of this paper is to develop a methodology for evaluating the robustness of multi-object tracking algorithms to input video stream degradation for intelligent video surveillance systems. To achieve this purpose, the following objectives are addressed: to formalize the impact of video-stream degradation on the tracking task; to determine the main degradation types typical of real video surveillance systems; to develop the structure of an experimental pipeline for controlled tracker testing; to define a set of metrics for evaluating accuracy, identity stability, trajectory continuity and computational efficiency; and to introduce integral robustness indicators that compare algorithms by the rate of quality decrease as degradation intensity increases.

Results. The essence of the proposed methodology lies in combining controlled one-factor degradation modelling, mandatory adaptation of ground-truth annotations to spatial and temporal transformations, joint analysis of detection, association and computational metrics, and two integral indicators that describe the complete degradation curve and the operational failure boundary. Unlike conventional benchmark evaluation on a fixed video stream, the proposed protocol compares trackers by robustness profiles while keeping the detector, source sequence, preprocessing and evaluation settings unchanged.

Let the input video stream be represented as a sequence of frames V = {I1, I2, ..., IT}, where T is the number of frames in the fragment. For each frame, the detector forms a set of observations Zt = {zt1, zt2, ..., ztn}, where each observation contains the bounding-box coordinates, confidence score and, if required, additional attributes. The tracker uses these observations to construct a set of trajectories Tr = {tau1, tau2, ..., tauN}, where each trajectory corresponds to one physical object and preserves its identifier over time.

In the case of input-stream degradation, each frame It is transformed by an operator Gtheta parameterized by a vector theta: I't = Gtheta(It). The parameter theta can represent noise variance, blur-kernel size, compression factor, frame-rate reduction coefficient, frame-loss probability or occlusion duration. If several degradations are applied simultaneously, the resulting transformation is a composition of degradation operators. This representation emphasizes that degradation affects the input data before detection and therefore influences both detection accuracy and the subsequent temporal association process.

Within the proposed robustness-evaluation methodology, the general experimental pipeline can be written as Y = E(T(D(Gtheta(V))), GT), where D is the object detector, T is the tracking algorithm, E is the evaluation module, GT is the ground-truth annotation and Y is the vector of quality metrics. The expression describes a complete reproducible experiment: the same source video is degraded in a controlled way, then processed by a fixed detector and a selected tracker, and finally compared with the synchronized ground truth.

The methodology is not limited to one detector or one dataset. However, for a fair comparison of tracking-by-detection algorithms, the detector must be fixed within one experimental series. Otherwise, final differences in MOTA, IDF1 or HOTA may be caused by detection quality rather than by the tracker. Therefore, all trackers should receive the same bounding boxes, confidence values and class labels obtained with the same preprocessing and non-maximum suppression settings.

The experimental pipeline consists of five main modules: Degradation Engine, Detection Wrapper, Tracking Hub, GT Processor and Evaluation Engine. The Degradation Engine generates degraded video sequences according to selected parameters. The Detection Wrapper performs unified object detection using a fixed detector. The Tracking Hub executes the selected tracking algorithm. The GT Processor adapts ground-truth annotations to spatial or temporal changes caused by degradation. The Evaluation Engine calculates tracking metrics and robustness indicators. The functional roles, input data and output data of these modules are summarized in Table 1.

Table 1. **Functional modules of the experimental pipeline**

Module	Purpose	Input data	Output data
Degradation Engine	Modelling controlled degradations	Frames, degradation parameters	Degraded frames
Detection Wrapper	Unified object detection	Degraded frames	Bounding boxes with confidence
Tracking Hub	Execution of tracking algorithms	Detections, frames	Trajectories with IDs
GT Processor	Ground-truth adaptation	GT, degradation parameters	Synchronized GT
Evaluation Engine	Metric and robustness-indicator calculation	Trajectories, GT	MOTA, HOTA, IDF1, RS, BP

Source: **compiled by the authors**

The key methodological principle is factor isolation. In one experimental run, only one degradation parameter is changed, while all other conditions remain fixed. For example, when the influence of blur is investigated, resolution, compression, frame rate and occlusion parameters remain unchanged. This makes it possible to establish a direct relationship between a particular degradation and changes in tracking metrics.

Let M0 be the value of a selected metric on the reference video stream without degradation, and let M(theta) be the metric value under degradation with intensity theta. The relative quality decrease is defined as Delta M(theta) = ((M0 - M(theta)) / M0) * 100 %. This indicator provides a normalized measure of algorithm sensitivity to a specific distortion type and enables comparison of trackers with different baseline accuracies.

To ensure reproducibility, random degradations such as noise or random frame loss should be generated using a fixed pseudorandom-number generator seed. This guarantees that all trackers receive exactly the same degraded input frames, which eliminates uncontrolled variability between experimental runs. The configuration of each experiment should include dataset identifier, sequence name, tracker name, detector version, degradation type, intensity level, random seed, confidence threshold, non-maximum suppression threshold and evaluation settings.

Within the methodology, degradations are divided into three groups: spatial, temporal and observability-related. Spatial degradations change the quality of an individual frame. They include blur, noise, compression artifacts and reduced resolution. Such distortions primarily affect the detector and the visual features used by appearance-based trackers.

Blur is modelled as convolution of the frame with a Gaussian or motion kernel. It causes loss of high-frequency details, which worsens object localization and reduces the stability of appearance features. Noise simulates sensor operation under poor illumination and changes pixel-level statistics. Compression artifacts appear when the video stream is encoded at low bitrates and may destroy contours and texture patterns. Reduced resolution decreases the number of pixels that represent an object and is especially critical for small or distant targets.

Temporal degradations are particularly critical for trackers that rely on motion models. When FPS is reduced, the temporal interval between neighboring observations increases; therefore, the prediction error grows. Random or burst frame loss creates gaps during which the tracker must extrapolate object position without measurements. Observability-related degradations are mainly represented by occlusions, when an object is partially or completely hidden by other objects or scene elements. The degradation types and the corresponding intensity levels used in the proposed methodology are summarized in Table 2.

Table 2. **Example parameterization of degradation levels**

Degradation type	Parameter	L1	L2	L3
Blur	Kernel size, sigma	3x3; 1.0	7x7; 2.5	15x15; 4.0
Noise	Standard deviation sigma	10	25	50
Compression	JPEG quality / CRF analogue	75	40	15
Low resolution	Scale coefficient s	0.5	0.25	0.125
FPS reduction	Target frame rate	15 FPS	10 FPS	5 FPS
Random frame loss	Probability p	0.05-0.10	0.15-0.25	0.30-0.40
Occlusion	Area ratio and duration	alpha=0.3; 5 frames	alpha=0.6; 15 frames	alpha=1.0; 50 frames

Source: **compiled by the authors**

Evaluation correctness depends on the consistency between tracking results and ground-truth annotations. Spatial degradations such as noise, blur or compression do not change frame geometry; therefore, GT coordinates remain unchanged. However, when low-resolution degradation is modelled by downscaling, the bounding-box coordinates must be scaled by the same coefficient as the frame.

Temporal degradations require frame re-indexing. If some frames are removed due to FPS reduction or frame loss, the corresponding GT records must also be removed, and the remaining frame numbers must be renumbered sequentially. Without this procedure, the evaluation tool would compare predictions and annotations belonging to different time instants. This would distort FP, FN, IDSW and trajectory fragmentation values.

In addition, temporal degradations require adaptation of the max_age parameter, which determines how many frames a track may remain active without an update. If the baseline max_age corresponds to one second of video at 30 FPS, then after reducing the frame rate to 5 FPS the same real-time interval corresponds to six times fewer frames. Therefore, the parameter must be adjusted to preserve the same temporal interpretation of track lifetime.

The methodology is intended for comparative evaluation of trackers with different association principles. The main characteristics, strengths and expected vulnerabilities of the tracking algorithms considered for testing are summarized in Table 3.

Table 3. **Characteristics of tracking algorithms for testing**

Algorithm	**Approach class**	**Strengths**	**Expected vulnerabilities**
SORT	Kalman + Hungarian + IoU	High speed, simple integration	Occlusions, frame loss, ID switches
DeepSORT	SORT + Re-ID features	Better identity preservation	Compression, noise, low resolution
ByteTrack	High/low-score detection association	Robustness to confidence reduction	Risk of FP under strong artifacts
OC-SORT	Observation-centric motion model	Robustness to occlusions and missing frames	Dependence on detection quality
CenterTrack	Joint detection-tracking	Integrated motion association	GPU demands, dependence on training data

***Source:* compiled by the authors**

The MOTA metric aggregates the number of false negatives, false positives and identity switches relative to the total number of ground-truth objects [9]: MOTA = 1 - (FN + FP + IDSW) / GT. It provides a general view of tracking accuracy but does not fully characterize identity stability. Under severe degradation, MOTA can be dominated by false negatives and may not reveal whether a tracker preserves identity for the objects that remain detected.

The IDF1 metric estimates the ability of the algorithm to preserve correct object identity over time [10]. It is especially informative under occlusions, frame loss and degradation of visual features. HOTA is important for degradation analysis because it separates the influence of detection quality and association quality [11]. If a particular distortion mainly reduces DetA, the problem is related to object detection. If AssA decreases, the main problem lies in temporal association and identity preservation.

The first integral indicator is Robustness Score (RS), calculated as the normalized area under the degradation curve. For discrete degradation levels $\theta 0$, $\theta 1$, …, θL, where $\theta 0$ denotes the undistorted stream, the trapezoidal estimate is $RS = [\Sigma(i=1\ldots L)\ ((Mi-1/M0 + Mi/M0)/2)(\theta i-\theta i-1)] /(\theta L-\theta 0) \cdot 100\ \%$. Here M0 is the baseline metric and Mi is its value at level θi. The use of normalized metric values makes it possible to compare trackers that have different baseline accuracy. An RS value close to one hundred percent means that quality is preserved over the investigated range, whereas a low value indicates rapid degradation.

The second indicator is Breaking Point (BP), defined as the lowest degradation intensity at which the normalized quality falls below an admissible coefficient β: $M(\theta BP)/M0 < \beta$. In the experimental protocol $\beta = 0.50$ is used unless the application imposes another reliability threshold. When the threshold lies between two tested levels, BP is estimated by linear interpolation between the adjacent normalized metric values. If the threshold is not crossed, BP is reported as being above the maximum investigated degradation level. This definition makes the indicator reproducible and independent of the absolute scale of the selected metric.

Table 4 illustrates the calculation procedure using normalized IDF1 values for one tracker under increasing blur: 1.00 for the original stream, 0.88 at the first level, 0.64 at the second level and 0.41 at the third level.

Table 4. **Illustrative calculation of Robustness Score and Breaking Point under increasing blur intensity**

Blur level	Normalized IDF1	Threshold status	Interpretation
0	1.00	Above 0.50	Baseline
1	0.88	Above 0.50	Weak quality loss
2	0.64	Above 0.50	Operational
3	0.41	Below 0.50	Breaking point crossed

Source: **compiled by the authors**

For equally spaced levels, trapezoidal integration gives RS = 74.2 %. The quality crosses the $\beta = 0.50$ boundary between the second and third levels; linear interpolation gives BP = 2.61 in the adopted level scale. The example demonstrates the calculation procedure only and is not used as an experimental result. For practical interpretation of the obtained robustness scores, the proposed RS ranges are classified as shown in Table 5.

Table 5. **Interpretation of Robustness Score values**

RS, %	Robustness category	Practical interpretation
90-100	High	The algorithm almost completely preserves effectiveness
60-89	Satisfactory	Quality decreases moderately and the system remains operational
30-59	Low	Significant fragmentation and error growth are observed
< 30	Critical	The algorithm loses practical applicability

Source: **compiled by the authors**

The experimental protocol consists of three levels. At the first level, a baseline run without degradation is performed to determine initial metric values for each algorithm. At the second level, one-factor experiments are conducted, in which one degradation type is applied at several intensity levels. At the third level, combined or stress-test scenarios can be performed to determine the behavior of trackers under realistic compound distortions.

The baseline run is performed on the original sequence before any degradation is introduced. Its MOTA, HOTA, IDF1 and processing-speed values serve as M0 and as the reference for all relative losses. To prevent unsupported generalization, absolute values must be reported directly from the archived experimental logs for each sequence, detector version and hardware configuration. The methodology therefore requires the publication package to include the configuration file and machine-readable result table; values from other papers or from a different detector must not be substituted because this would violate the factor-isolation principle.

Spatial degradation is generated with OpenCV operations: Gaussian or motion blur, additive Gaussian noise, JPEG compression and downscaling followed by restoration to the detector input size. Temporal degradation is generated by deterministic frame subsampling and pseudorandom frame removal with a fixed seed. Occlusions are modelled by masks whose area and duration are controlled. Video decoding and compression parameters are fixed using FFmpeg. Tracking results are converted to the MOT Challenge format, and MOTA, HOTA, IDF1, identity switches, false positives, false negatives and fragmentations are calculated with the TrackEval toolkit. Processing speed is measured after a warm-up interval and reported as the median number of processed frames per second.

Experimental implementation of the protocol. The validation series uses pedestrian sequences from MOT17 and MOT20. MOT17 provides scenes with different viewpoints, motion patterns and pedestrian densities, whereas MOT20 focuses on highly crowded scenes with long and frequent occlusions. The same sequence fragments and synchronized annotations are used for all trackers. A fixed YOLOX detector supplies identical bounding boxes, confidence values and class labels to SORT, DeepSORT, ByteTrack and OC-SORT; CenterTrack is evaluated separately because detection and tracking are integrated in one network. Detector confidence, non-maximum suppression and image-preprocessing settings remain unchanged within each comparison series.

For one video fragment, the number of basic one-factor runs is defined as $N = M * D * L + M$, where M is the number of algorithms, D is the number of degradation types, L is the number of intensity levels and the additional M corresponds to baseline runs. For example, when five algorithms, seven degradation types and three intensity levels are used, the protocol requires 110 runs for one video sequence.

For experiments in pedestrian multi-object tracking, datasets of the MOT Challenge family are especially relevant. MOT17 contains scenes with different camera types, object densities and lighting conditions [12]. MOT20 focuses on crowded scenes and provides a challenging test for identity preservation under frequent occlusions [13]. The combination of these datasets enables evaluation of robustness under both typical and extreme surveillance conditions.

For occlusions, the key criterion is the ability to preserve the same ID after the object reappears. Therefore, it is advisable to analyze not only MOTA but primarily IDF1, IDSW and trajectory breaks. In crowded scenes, such as those represented in MOT20, the difference between algorithms may become more pronounced because degradation amplifies the ambiguity of data association.

An additional diagnostic step is the analysis of error components. If FP increases sharply under noise, the detector becomes unstable and the tracker receives many false observations. If FN increases under blur or low resolution, true objects are not detected and the tracker cannot update their states. If IDSW increases under FPS reduction or occlusion, the main failure mechanism is association ambiguity. This separation of causes is one of the main advantages of the proposed methodology. The diagnostic interpretation of the main metric changes under degradation is summarized in Table 6.

Table 6. **Diagnostic interpretation of metric changes under degradation**

Observed effect	**Most probable cause**	**Recommended analysis**
Increase in FN	Object detector misses true objects	Check detector confidence distribution and object scale
Increase in FP	Noise or artifacts create false detections	Analyze NMS and confidence threshold
Increase in IDSW	Association ambiguity or wrong re-identification	Analyze IDF1, HOTA AssA and occlusion intervals
Increase in fragmentations	Track lifecycle is too short or observations are missing	Check max_age and frame-loss patterns
FPS decrease	Computational overload or too many active tracks	Separate detection time and tracking time

***Source*: compiled by the authors**

The expected robustness profiles differ because the trackers use different association cues. SORT is computationally efficient but is most sensitive to frame loss, frame-rate reduction and long occlusions because its identity continuity is determined mainly by motion prediction and geometric overlap. DeepSORT usually preserves identity better during short occlusions, but blur, compression and low resolution reduce the discriminative value of appearance descriptors. ByteTrack is less sensitive to moderate confidence reduction because it also associates low-score detections, although severe noise may increase false trajectories. OC-SORT is expected to be comparatively stable under missing observations and non-linear motion, while its performance still declines when the detector stops producing usable observations. CenterTrack may benefit from joint motion estimation but requires a separate evaluation due to its integrated detector.

Differences between MOT17 and MOT20 should be interpreted through scene density rather than as a dataset-specific ranking alone. In MOT17, moderate degradation often first affects detection accuracy, so false negatives and MOTA dominate the change. In MOT20, dense crowds, persistent overlap and visually similar pedestrians amplify association ambiguity; therefore IDF1, HOTA association accuracy, identity switches and fragmentation become more sensitive than the aggregate accuracy metric. Consequently, the same degradation level can produce a substantially earlier breaking point on MOT20, especially for trackers without robust appearance or observation-centric association.

At the same time, the methodology has limitations. Synthetic degradation models cannot fully reproduce all physical properties of real cameras, weather phenomena or network transmission errors. For example, real compression artefacts depend on codec implementation, scene motion and bitrate-control strategy. Real rain, fog or snow may affect not only contrast but also object visibility in a spatially non-uniform way. Therefore, controlled synthetic experiments should be complemented by real degraded sequences when such data are available.

Before using the methodology for final algorithm comparison, the experimental bench should be validated on simple scenarios where the expected result is clear. The first validation scenario is the baseline run on clean video. The results should be close to known values reported for the selected algorithms and datasets. Significant deviation may indicate an implementation error, incorrect detector configuration or incompatible evaluation settings.

The second validation scenario is a controlled degradation with weak intensity. The quality metrics should decrease moderately or remain close to the baseline. If weak blur or weak noise causes a catastrophic drop in all metrics, the degradation implementation may be too aggressive or the preprocessing pipeline may be inconsistent. This check is important because parameter values can have different effects depending on image scale and detector input size. In addition to validation of the experimental bench, the results should be reported in a standardized form that allows degradation curves, error components, robustness indicators and computational characteristics to be compared. The recommended reporting structure is given in Table 7.

Table 7. **Recommended reporting structure for degradation-aware MOT evaluation**

Report component	Content	Purpose
Baseline metrics	MOTA, HOTA, IDF1, FPS without degradation	Define reference quality
Degradation curves	Metric values for L1-L3 levels	Visualize quality loss
Error components	FP, FN, IDSW, fragmentations	Identify failure mechanisms
Robustness indicators	RS and BP for each tracker	Compare reliability profiles
Computational indicators	Latency, FPS, memory use	Assess deployment feasibility
Configuration metadata	Detector, tracker, seed, hardware	Ensure reproducibility

Source: **compiled by the authors**

The third validation scenario is a temporal re-indexing test. A small sequence can be manually subsampled, and the correspondence between remaining frames and GT records can be verified. This prevents one of the most dangerous errors in MOT evaluation: comparing predictions and annotations from different time instants. The test should confirm that every prediction is compared with the correct ground-truth frame after frame removal.

The fourth validation scenario concerns coordinate consistency. When low-resolution transformation is used, the bounding-box coordinates in predictions and GT must belong to the same coordinate system. A visual overlay of predicted and ground-truth boxes for several frames is a simple but effective check. If boxes are shifted or scaled incorrectly, IoU-based metrics become meaningless.

The fifth validation scenario is reproducibility. The same experiment should be run twice with the same configuration and random seed. The results should be identical or differ only within numerical precision. If differences are observed, the source of nondeterminism should be identified. Possible causes include nondeterministic GPU operations, random data augmentation, unstable detector output or unordered processing of detections.

Conclusions. The paper proposes a methodology for evaluating the robustness of multi-object tracking algorithms to input video stream degradation. The methodology is based on a modular experimental pipeline, the principle of factor isolation, controlled modelling of spatial, temporal and occlusion-related degradations, adaptation of ground-truth annotations and multi-criteria metric evaluation.

It is shown that adequate analysis of tracking algorithms under real conditions cannot be limited to standard metrics on clean video. It is necessary to evaluate quality changes under progressive video-stream degradation, analyze error components and determine the operational boundary of an algorithm. MOTA, HOTA, IDF1, IDSW, FP, FN, fragmentation indicators and FPS should be considered jointly.

The proposed Robustness Score and Breaking Point indicators provide a compact quantitative description of tracker behavior under degradation. Robustness Score characterizes quality preservation over the whole range of distortion levels, while Breaking Point identifies the degradation threshold at which a tracker loses practical applicability. Together, these indicators make it possible to compare algorithms by reliability profiles rather than by a single baseline score.

The methodology can be used for comparative analysis of SORT, DeepSORT, ByteTrack, OC-SORT, CenterTrack and other algorithms, as well as for practical selection of solutions for intelligent video surveillance systems. Further research should focus on experimental validation of the methodology on MOT17 and MOT20 datasets, analysis of combined degradations, integration of real weather and network distortions, and development of adaptive mechanisms for selecting tracker parameters according to the current video-stream quality.

Conflict of interests: Authors declare that they do not have conflict of interests, in particular financial, personal, copyright or any of a different nature, which could have influenced the research, as well as the results published in this articles

Financing: Research was carried out without financial support

Accessibility data: All data available in digital or graphic form in the main text manuscript

Using artificial intelligence tools: During preparation manuscript ChatGPT Plus was used to write the article to improve the structure and English version of the content. All fragments processed by it were checked and edited by the author. AI was not used to generate text, formulas, and tables, formation scientific provisions, receipt results or formulation conclusions. Scientific the provisions, results and conclusions are my own, intellectual by the contribution of the authors

REFERENCES

1. Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W. & Kim, T.-K. "Multiple object tracking: A literature review". *Artificial Intelligence*. 2021; 293: 103448. DOI: https://doi.org/10.1016/j.artint.2020.103448.

2. Park, Y., Dang, L. M., Lee, S., Han, D. & Moon, H. "Multiple object tracking in deep learning approaches: A survey". *Electronics*. 2021; 10 (19): 2406. DOI: https://doi.org/10.3390/electronics10192406.

3. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W. & Wang, X. "ByteTrack: Multi-object tracking by associating every detection box". *Computer Vision – ECCV*. 2022. p. 1–21. DOI: https://doi.org/10.1007/978-3-031-20047-2_1.

4. Cao, J., Pang, J., Weng, X., Khirodkar, R. & Kitani, K. "Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023. p. 9686–9696. DOI: https://doi.org/10.1109/CVPR52729.2023.00934.

5. Zhou, X., Koltun, V. & Krähenbühl, P. "Tracking objects as points". *Computer Vision – ECCV*. 2020. p. 474–490. DOI: https://doi.org/10.1007/978-3-030-58548-8_28.

6. Zhang, Y., Wang, C., Wang, X., Zeng, W. & Liu, W. "FairMOT: On the fairness of detection and re-identification in multiple object tracking". *International Journal of Computer Vision*. 2021; 129: 3069–3087. DOI: https://doi.org/10.1007/s11263-021-01513-4.

7. Meinhardt, T., Kirillov, A., Leal-Taixé, L. & Feichtenhofer, C. "TrackFormer: Multi-object tracking with transformers". 2021. DOI: https://doi.org/10.48550/arXiv.2101.02702.

8. Zeng, F., Dong, B., Zhang, Y., Wang, T., Zhang, X. & Wei, Y. “MOTR: End-to-End multiple-object tracking with transformer”. *Computer Vision – ECCV*. 2022. p. 659–675. DOI: https://doi.org/10.1007/978-3-031-19812-0_38.

9. Dendorfer, P., Osep, A., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S. & Leal-Taixé, L. “MOTChallenge: A benchmark for single-camera multiple target tracking”. *International Journal of Computer Vision*. 2021; 129: 845–881, https://www.scopus.com/pages/publications/85098212408. DOI: https://doi.org/10.1007/s11263-020-01393-0.

10. Dai, Y., Hu, Z., Zhang, S. & Liu, L. “A survey of detection-based video multi-object tracking”. *Displays*. 2022; 75: 102317. DOI: https://doi.org/10.1016/j.displa.2022.102317.

11. Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L. & Leibe, B. “HOTA: A higher order metric for evaluating multi-object tracking”. *International Journal of Computer Vision*. 2021; 129: 548–578, https://www.scopus.com/pages/publications/85092405889. DOI: https://doi.org/10.1007/s11263-020-01375-2.

12. Dendorfer, P., Rezatofighi, H., Milan, A., Shi, J., Cremers, D., Reid, I., Roth, S., Schindler, K. & Leal-Taixé, L. “MOT20: A benchmark for multi-object tracking in crowded scenes”. 2020. DOI: https://doi.org/10.48550/arXiv.2003.09003.

13. Sun, P., Cao, J., Jiang, Y., Yuan, Z., Bai, S., Kitani, K. & Luo, P. “DanceTrack: Multi-object tracking in uniform appearance and diverse motion”. 2021. DOI: https://doi.org/10.48550/arXiv.2111.14690.

14. Ge, Z., Liu, S., Wang, F., Li, Z. & Sun, J. “YOLOX: Exceeding YOLO Series in 2021”. 2021. DOI: https://doi.org/10.48550/arXiv.2107.08430.

15. Rakai, L., Song, H., Sun, S., Zhang, W. & Yang, Y. “Data association in multiple object tracking: A survey of recent techniques”. *Expert Systems with Applications*. 2022; 192: 116300. DOI: https://doi.org/10.1016/j.eswa.2021.116300.

16. Sun, Z., Chen, J., Chao, L., Ruan, W. & Mukherjee, M. “A survey of multiple pedestrian tracking based on tracking-by-detection framework”. *IEEE Transactions on Circuits and Systems for Video Technology*. 2021; 31 (5): 1819–1833. DOI: https://doi.org/10.1109/TCSVT.2020.3009717.

17. Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T. & Meng, H. “StrongSORT: Make DeepSORT great again”. 2022. DOI: https://doi.org/10.48550/arXiv.2202.13514.

18. Aharon, N., Orfaig, R. & Bobrovsky, B.-Z. “BoT-SORT: robust associations multi-pedestrian tracking”. 2022. DOI: https://doi.org/10.48550/arXiv.2206.14651.

19. Ciaparrone, G., Sánchez, F. L., Tabik, S., Troiano, L., Tagliaferri, R. & Herrera, F. “Deep learning in video multi-object tracking: A survey”. *Neurocomputing.* 2020; 381: 61–88. DOI: https://doi.org/10.1016/j.neucom.2019.11.023.

20. Wu, S., Hadachi, A., Lu, C. & Vivet, D. “MOTT: A new model for multi-object tracking based on green learning paradigm”. *AI Open*. 2023; 4: 145–153. DOI: https://doi.org/10.1016/j.aiopen.2023.09.002.

21. Pang, J., Qiu, L., Li, X., Chen, H., Li, Q., Darrell, T. & Yu, F. “Quasi-dense similarity learning for multiple object tracking”. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2021. p. 164–173. DOI: https://doi.org/10.1109/CVPR46437.2021.00023.

22. Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. “YOLOv4: Optimal speed and accuracy of object detection”. 2020. DOI: https://doi.org/10.48550/arXiv.2004.10934.

23. Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., Tomizuka, M., Li, L., Yuan, Z., Wang, C. & Luo, P. “Sparse R-CNN: End-to-End object detection with learnable proposals”. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2021. p. 14454–14463. DOI: https://doi.org/10.1109/CVPR46437.2021.01422.

24. Cai, J., Xu, M., Li, W., Xiong, Y., Xia, W., Tu, Z. & Soatto, S. “MeMOT: Multi-object tracking with memory”. 2022. DOI: https://doi.org/10.48550/arXiv.2203.16761.

25. Zhang, Y., Wang, T. & Zhang, X. “MOTRv2: Bootstrapping End-to-End multi-object tracking by pretrained object detectors”. 2022. DOI: https://doi.org/10.48550/arXiv.2211.09791.

Received 16.07.2026
Received after revision 26.08.2026
Accepted 02.09.2026

DOI: https://doi.org/10.15276/ict.03.2026.16
УДК 004.932

Методика оцінювання стійкості алгоритмів багатооб'єктного трекінгу до деградації вхідного відеопотоку

Стрельцов Олег Васильович [1)]
ORCID: https://orcid.org/0000-0002-4691-5703; streltsov.o.v@op.edu.ua. Scopus Author ID: 57210373473
Поліщук Максим Володимирович [1)]
ORCID: https://orcid.org/0009-0004-4425-9564; 4107624@op.edu.ua
Гриньов Максим Андрійович [1)]
ORCID: https://orcid.org/0009-0007-7189-4485; grinevvv2002@stud.op.edu.ua
[1)] Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

У статті запропоновано розширену методику оцінювання стійкості алгоритмів багатооб'єктного трекінгу до деградації вхідного відеопотоку в системах інтелектуального відеоспостереження. Актуальність дослідження зумовлена тим, що в реальних умовах відеодані зазнають просторових, фотометричних, часових та оклюзійних спотворень, зокрема розмиття, шуму, артефактів компресії, зниження роздільної здатності, втрати кадрів, зменшення частоти та оклюзій. Методика базується на модульному експериментальному конвеєрі, який охоплює підготовку відеоданих, моделювання контрольованих деградацій, детектування об'єктів, запуск алгоритмів трекінгу, адаптацію еталонної розмітки та розрахунок метрик якості. Для коректного порівняння використано принцип ізоляції факторів. Ефективність оцінюється за точністю багатооб'єктного трекінгу, точністю трекінгу вищого порядку, мірою ідентифікації, кількістю перемикань ідентичності, хибних спрацювань, пропущених об'єктів, фрагментацій траєкторій та швидкістю оброблення кадрів. Для аналізу робастності введено показники оцінки стійкості та граничної точки руйнування, що дають змогу визначити абсолютну якість трекінгу й темп її падіння зі зростанням інтенсивності деградації. Методика придатна для порівняльного аналізу алгоритмів простого онлайн-трекінгу в реальному часі, його поглибленої та спостережно-центричної модифікацій, ByteTrack і CenterTrack, а також для вибору рішень для систем відеоспостереження за обмеженої якості відеопотоку. Практична цінність методики полягає у її використанні під час проєктування, налаштування та валідації систем відеоаналітики, що працюють за несталої якості вхідного відеопотоку. Запропонований підхід може бути основою розробки адаптивних систем трекінгу, які автоматично змінюють параметри детектора або трекера відповідно до рівня деградації відеопотоку. Методика орієнтована на відтворюваність експериментів, тому передбачає фіксацію параметрів програмного середовища, версій бібліотек, конфігурації детектора, порогів асоціації та випадкового зерна для стохастичних перетворень. Підкреслено необхідність окремого аналізу складових помилки. Структура протоколу дає змогу відокремити ці механізми та сформувати практичні висновки щодо вибору алгоритмів для умов експлуатації. Зокрема, для систем із низькою пропускною здатністю важливішою є стійкість до компресії, для систем із нестабільною мережею – стійкість до втрати кадрів, а для щільних сцен – здатність зберігати ідентичність об'єкта під час оклюзій.

Ключові слова: багатооб'єктний трекінг; деградація відеопотоку; інтелектуальне відеоспостереження

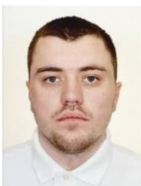
ABOUT THE AUTHORS

Oleh V. Streltsov - Lecturer / Researcher in Computer Engineering and Intelligent Systems. Odesa Polytechnic National University. 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-4691-5703; e-mail: streltsov.o.v@op.edu.ua
Research field: computer vision; intelligent video surveillance; multi-object tracking; embedded AI systems

Стрельцов Олег Васильович - канд. техн. наук, доцент кафедри Комп'ютерних систем Національний університет «Одеська політехніка», просп. Шевченка, 1. Одеса, 65044, Україна
Галузь досліджень: комп'ютерний зір; інтелектуальне відеоспостереження; відстеження кількох об'єктів; вбудовані системи штучного інтелекту

Maksym V. Polishchuk - Master of Computer Engineering. Odesa Polytechnic National University, 1,.Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0009-0004-4425-9564; 4107624@op.edu.ua
Research field: artificial intelligence in the control of mobile robotic systems

Поліщук Максим Володимирович - магістр Комп'ютерної інженерії. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна
Галузь досліджень: штучний інтелект в управлінні мобільними роботизованими системами



Maksym A. Hrynyov - Post Graduate Student of the Computer Systems Department. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0009-0007-7189-4485; grinevvv2002@stud.op.edu.ua
Research field: Improving models and methods for analyzing sensory information in autonomous mobile systems.

Гриньов Максим Андрійович - аспірант кафедри Комп'ютерних систем, Національний університет «Одеська політехніка», просп. Шевченка, 1. Одеса, 65044, Україна
Галузь досліджень: удосконалення моделей та методів аналізу сенсорної інформації в автономних мобільних системах