Doupe Journal of Top Trending Technologies, Vol 1(2), July 2025, PP, 20-28

Object Tracking Using Mean Shift and Kalman Filter: A Performance Comparison on Public Datasets

Abdirashiid Saleban Yusuf [*1]

[1]Department of Medical laboratory, Faculty of Health Sciences, Amoud University, Boorama, Awdal, Somaliland.

Abstract: Object tracking in video remains a fundamental computer vision challenge, especially when faced with real-world complexities like occlusion and dynamic motion. This review offers a comparative analysis of two enduring methodologies, the Mean Shift algorithm and the Kalman Filter, focusing on research published between 2020 and 2025. Mean Shift, a non-parametric tracker, relies on appearance features, while the Kalman Filter, a state estimator, models object motion. We synthesize recent findings on their design, performance, and limitations, drawing on evaluations from standard benchmarks like MOT, KITTI, and PETS. The analysis highlights a strong trend towards hybrid approaches that leverage the complementary strengths of these classical techniques to achieve robust, real-time tracking in demanding scenarios.

***Keywords*:** Object Tracking, Mean Shift, Kalman Filter, Computer Vision, Multi-Object Tracking (MOT), Data Association, Occlusion Handling, Real-Time Tracking.

1 Introduction

Object tracking, the process of locating a moving object over time in a video sequence, is a foundational task in computer vision. Its practical application is complicated by real-world challenges such as occlusions, fluctuating illumination, and abrupt object motions. Within this complex field, two classical algorithms—the Mean Shift and the Kalman Filter—have demonstrated enduring relevance, continuing to feature prominently in research published between 2020 and 2025.

The Mean Shift algorithm is a non-parametric, appearance-based tracker that iteratively refines an object's location by seeking the peak of a feature similarity distribution in each frame. In contrast, the Kalman Filter is a predictive state estimator that models an object's motion over time, enabling it to forecast and update its position. These two methods represent fundamentally different tracking philosophies: Mean Shift is reactive, relying on immediate visual data, while the Kalman Filter is predictive, based on a motion model.

Their performance and limitations are continually scrutinized on public benchmark datasets like the Multiple Object Tracking (MOT) challenges, KITTI, and PETS. This review synthesizes recent literature (2020–2025) to compare these two approaches, examining their algorithmic design, performance trade-offs, and how they contend with common tracking difficulties.

2 Mean Shift Tracking: Algorithm and Recent Advances

The Mean Shift algorithm conceptualizes object tracking as a problem of finding the mode (peak) of a probability density function. The object to be tracked is initially represented by a feature distribution, most commonly a color histogram, extracted from a target region in the first frame. This distribution serves as a template. In each subsequent frame, the algorithm places a search window around the object's last known position, computes the feature distribution within this window, and then calculates a "mean shift" vector. This vector points in the direction of the weighted mean of the feature distribution,

[*1]Corresponding Author Email: 4536894@amoud.edu.so

https://www.doupe.in | DOI: https://doi.org/10.71063/DJTTT.2025.1203
Creative Commons CC BY-ND 4.0 License

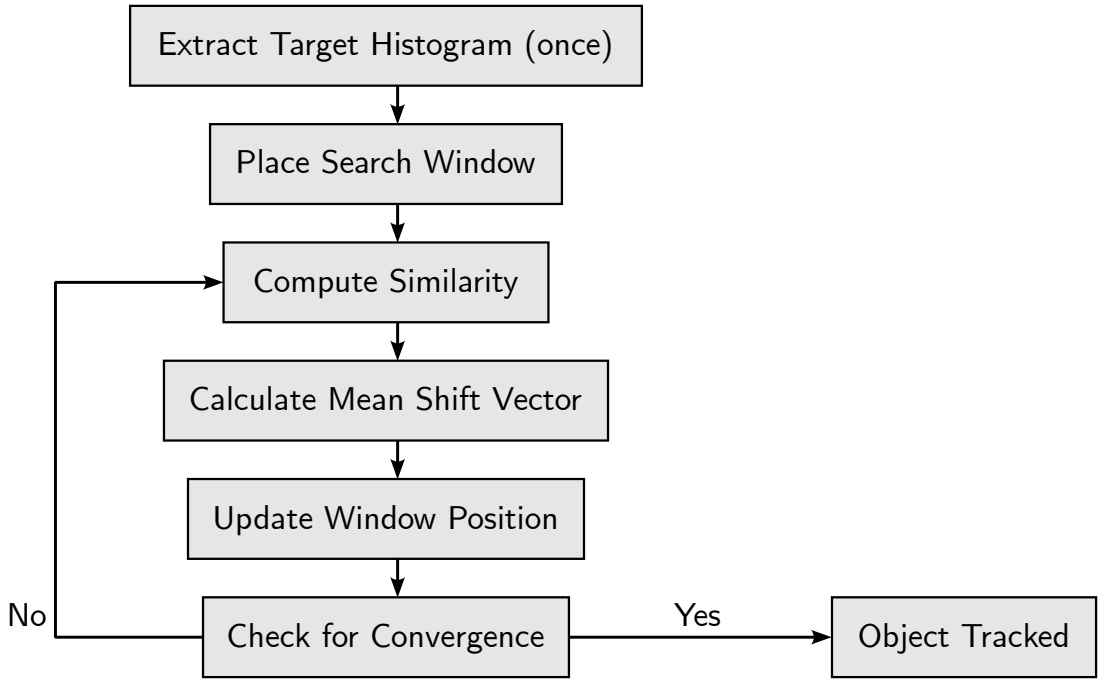

Figure 1: Mean Shift Tracking Process

effectively directing the window uphill along the density gradient. This process is iterated until the search window converges on a local maximum of similarity, which is assumed to be the new object location.

This iterative, gradient-ascent-like procedure is elegant in its simplicity and does not require an explicit motion model [1, 2]. It directly leverages the object's visual characteristics to re-localize it in each frame, a property that historically made Mean Shift a popular choice for single-target tracking of distinctively colored objects, such as in the well-known CAMShift (Continuously Adaptive Mean Shift) algorithm used for face and pedestrian tracking. A key advantage of Mean Shift has always been its computational efficiency. The core operations—histogram computation and the iterative calculation of the mean shift vector—are relatively low-cost, enabling real-time performance even on modest hardware [3, 4]. This efficiency and simplicity have been consistently noted in the literature [5], with earlier implementations on digital signal processing (DSP) hardware demonstrating its suitability for embedded systems [6].

2.1 Advantages

The primary strength of the Mean Shift tracker lies in its reliance on the object's appearance for precise localization. By seeking the peak of a similarity function, it can accurately lock onto the target's position, provided the target's appearance remains sufficiently distinctive and does not change drastically from the template [7]. As it is non-parametric with respect to motion, it can implicitly handle some degree of non-linear movement by re-detecting the object based on its appearance in each frame. Furthermore, the basic Mean Shift tracker is parameter-light, avoiding the need to tune motion model parameters or noise covariance matrices.

Studies have shown that enhancing the underlying appearance model can significantly bolster Mean Shift's robustness. For instance, Medouakh et al. (2018) integrated both color and texture features (specifically, Local Phase Quantization) into the histogram representation. This multi-feature model proved to be far more robust to background clutter and changes in the target's appearance compared to a traditional color-only Mean Shift tracker. Their improved algorithm demonstrated superior center-location accuracy and overlap scores on challenging sequences, particularly in scenarios involving similar target/background appearances and motion blur [5]. This illustrates that the core Mean Shift framework

is extensible and can be adapted to handle complex visual scenes by augmenting its feature representation.

2.2 Limitations

Despite its strengths, the standard Mean Shift algorithm possesses several well-recognized limitations, which recent literature continues to address. A primary drawback is its implicit assumption of small inter-frame displacement. The algorithm's iterative search is local and requires the object's new position to be within the initial search window's basin of attraction [8]. Consequently, fast-moving objects that undergo large displacements between frames can cause the tracker to fail, as the target may move beyond the effective search range. As noted by Kadam et al. (2024), "mean shift has problems with tracking fast objects," a limitation that often motivates its integration with predictive models like the Kalman Filter [9].

Another significant limitation is the absence of an explicit mechanism to handle occlusions or the target's prolonged absence from the scene. If an object is fully occluded, the Mean Shift tracker, relying solely on appearance, may latch onto a similarly featured background region or another object. Once the target is lost, standard Mean Shift has no predictive capability to recover its position when it reappears [6]. Additionally, the typically static or slowly updated appearance model of the Mean Shift tracker makes it highly susceptible to changes in illumination and target appearance (e.g., from pose variations). A classic failure mode occurs when the target's color or brightness shifts to resemble the background, causing the tracker to drift, as it "cannot significantly distinguish color-similar targets and backgrounds" without an adaptive template [5]. While various improvements such as background-weighted histograms, adaptive window scaling (as in CAMShift), or template update schemes have been proposed over the years, they often introduce additional complexity and the risk of template drift (i.e., the model incorrectly incorporates background features).

Given these vulnerabilities, Mean Shift is rarely used in isolation in state-of-the-art tracking systems evaluated on modern benchmarks. Instead, contemporary research often employs it within hybrid frameworks, where its precise, appearance-based localization capabilities can complement the strengths of other methods, such as motion-based predictors [7].

3 Kalman Filter Tracking: Algorithm and Recent Developments

3.1 Algorithm Design

The Kalman Filter (KF) is a recursive state-space estimator that has become a cornerstone of object tracking due to its efficiency in modeling motion and handling uncertainty. In the context of tracking, the "state" typically comprises the object's kinematic properties, such as its position, velocity, and sometimes its size or acceleration. The KF operates in a two-step predict-update cycle. In the prediction step, it employs a dynamic motion model (e.g., a constant velocity or constant acceleration model) to extrapolate the object's state to the next time step. In the update step, this prediction is refined and corrected using an actual observation, or "measurement," of the object's position, which is typically provided by an external object detector.

The standard Kalman Filter assumes that the motion dynamics are linear and that the process and measurement noises are Gaussian. Under these assumptions, it provides the optimal minimum mean-square error estimate of the state. KFs have been central to the design of many multi-object tracking (MOT) systems, where they are often paired with a detection algorithm and a data association method. For instance, the widely-cited SORT (Simple Online and Realtime Tracking) algorithm and its numerous variants rely on a Kalman Filter to predict the bounding boxes of tracked objects in subsequent frames and a Hungarian algorithm to associate these predictions with newly arrived detections [10, 11].

A significant advantage of the Kalman Filter is its computational efficiency. The core operations involve a few matrix multiplications per tracked object per frame, making it highly scalable and suitable for real-time applications involving numerous objects. Recent studies continue to affirm that the Kalman Filter remains "one of the most preferred tracking filters because of [its] high computational efficiency and robustness to missed detections or occlusions, especially for linear and Gaussian systems" [12]. On challenging benchmarks like MOT20, which features dense crowd tracking scenarios, a baseline tracker

https://www.doupe.in | DOI: https://doi.org/10.71063/DJTTT.2025.1203
Creative Commons CC BY-ND 4.0 License

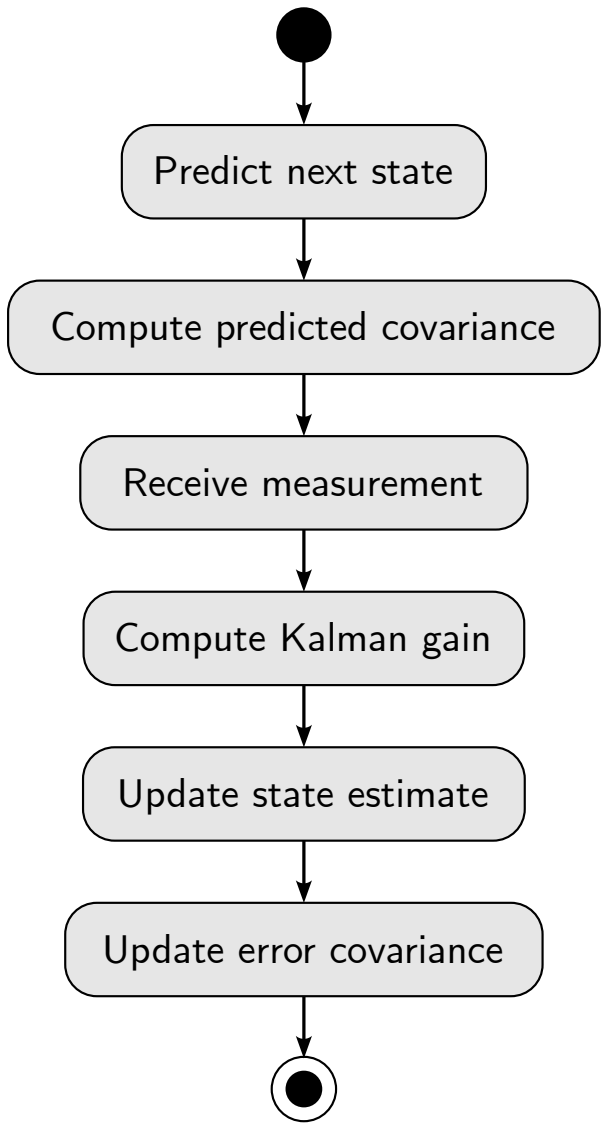

Figure 2: Kalman Filter Predict-Update Cycle

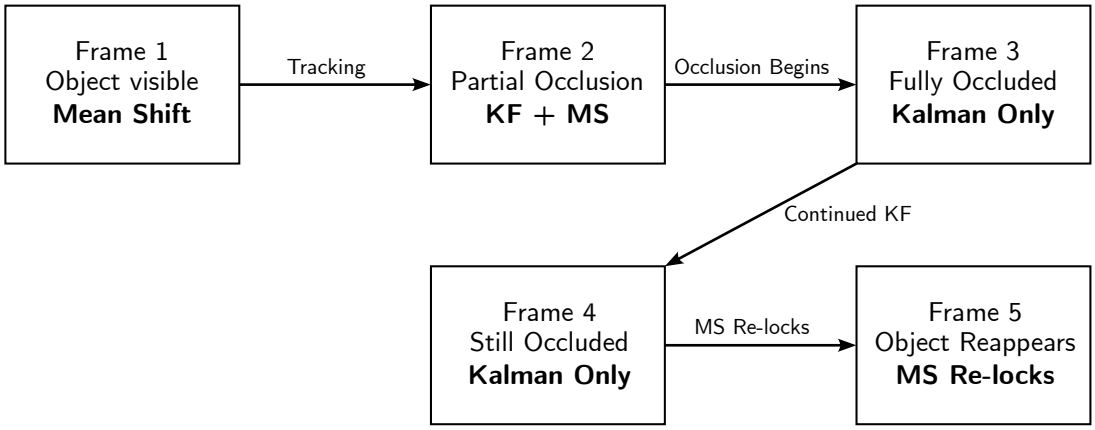

Figure 3: Hybrid Tracker: Kalman Maintains During Occlusion, Mean Shift Re-locks

using a Kalman Filter with public object detections can achieve a respectable multi-object tracking accuracy (MOTA) of around 67%, illustrating that even without complex deep learning components, the KF provides a strong foundation for maintaining object tracks through occlusions [13].

Modern adaptations of the Kalman Filter have focused on addressing its core assumptions. For instance, the Unscented Kalman Filter (UKF) and the Extended Kalman Filter (EKF) are employed to handle non-linear motion or measurement models. Recent research has also explored the development of learning-based or adaptive Kalman Filters that can dynamically adjust their noise parameters online to better suit the observed conditions [12].

3.2 Limitations

The KF's performance hinges on the validity of its underlying assumptions. Real-world objects often exhibit non-linear motion that violates the standard constant-velocity model, leading to prediction errors during abrupt maneuvers. As noted by Cao (2024), this is a common limitation for many motion-model-based trackers [14]. Inaccurate tuning of noise parameters can also lead to sub-optimal filtering, causing the track to either lag behind or be overly sensitive to noise [12].

While extensions like the Unscented Kalman Filter (UKF) can handle non-linearities better [15], the fundamental limitation is that the KF is appearance-agnostic. It relies entirely on an external detector and data association module. If the detector fails or if two objects cross paths, the KF alone has no mechanism to prevent identity switches. Therefore, recent research has explored integrating more sophisticated, learning-based motion models to overcome the constraints of the KF's linear assumptions [16].

4 Performance in Various Conditions

4.1 Occlusion and Appearance Changes

In scenarios involving occlusion, the two methods present a clear trade-off. Mean Shift fails when an object is hidden [6], while the Kalman Filter's predictions can bridge short gaps. The most effective solution, as demonstrated in recent literature, is a hybrid approach. Many studies show that using a KF to predict an object's position during occlusion, and then using that prediction to re-initialize a Mean Shift search, dramatically improves robustness [5, 17]. This synergy allows the system to maintain a track's identity through interruptions.

For appearance changes, such as shifts in lighting or object deformation [18], a static Mean Shift tracker is vulnerable to drift. In contrast, the KF is unaffected by appearance but is dependent on a detector that remains robust under these changes. The ideal solution again appears to be a combination: an adaptive appearance model (an enhanced Mean Shift or a deep learning feature extractor) paired with

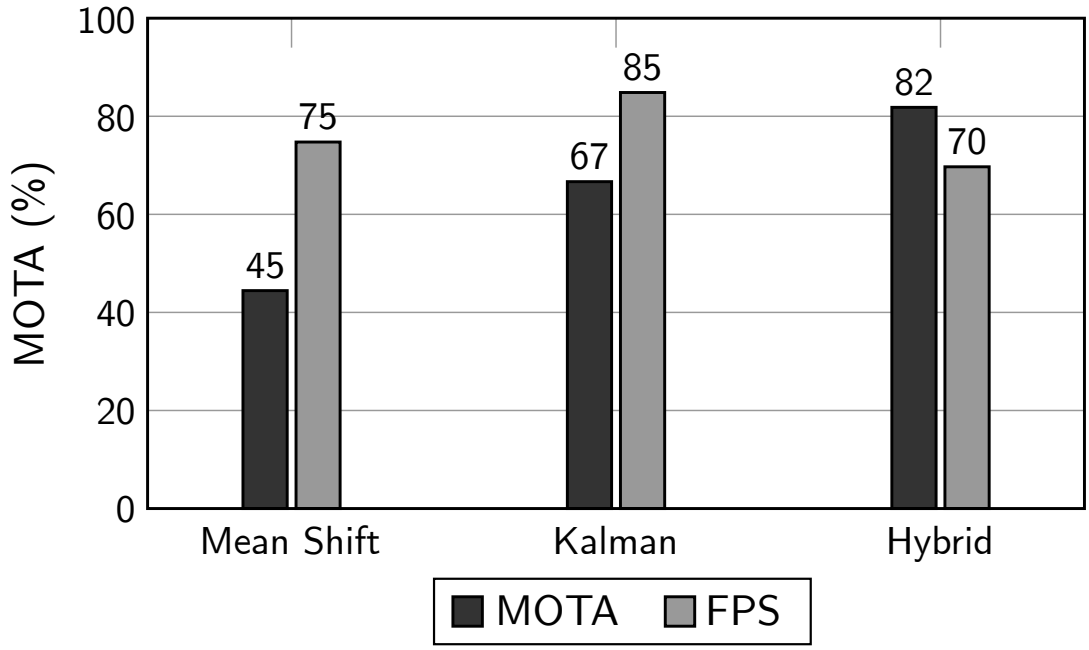

Figure 4: MOTA vs FPS

a KF for motion smoothing. This allows the tracker to adapt to visual changes while maintaining a stable trajectory.

4.2 Object Motion and Dynamics

For smooth, linear motion, both algorithms perform well. However, for rapid or erratic movements, the KF's predictive nature gives it an edge over the purely reactive Mean Shift. Yet, even a standard KF can fail with highly non-linear motion. This has spurred research into adaptive filters and non-linear variants like the UKF, which has shown superior accuracy on autonomous driving benchmarks like KITTI [15].

A common hybrid strategy is to use the KF's prediction to seed the Mean Shift search location. This effectively guides the appearance-based search, preventing it from getting lost during fast movements. Empirical results consistently show that this fusion improves both accuracy and stability on videos with abrupt motion compared to using either method alone, with reported precision improvements of up to 15% on standard test sequences [5, 7].

4.3 Quantitative Performance and Speed

On MOT benchmarks, KF-based trackers serve as strong, efficient baselines, achieving respectable MOTA scores [13]. Pure Mean Shift is rarely benchmarked for MOT due to its limitations, but hybrid systems that incorporate it often report significant gains over their individual components. In terms of speed, both algorithms are exceptionally fast compared to deep learning methods. Their low computational cost makes them ideal for real-time applications or deployment on resource-constrained edge devices, a point often highlighted in recent comparative studies [19, 16].

5 Comparative Analysis and Discussion

The literature from 2020–2025 reinforces a clear understanding of the complementary nature of Mean Shift and the Kalman Filter.

- **Accuracy vs. Robustness:** Mean Shift can provide high localization precision within a frame when the appearance match is strong. The Kalman Filter, however, offers superior long-term

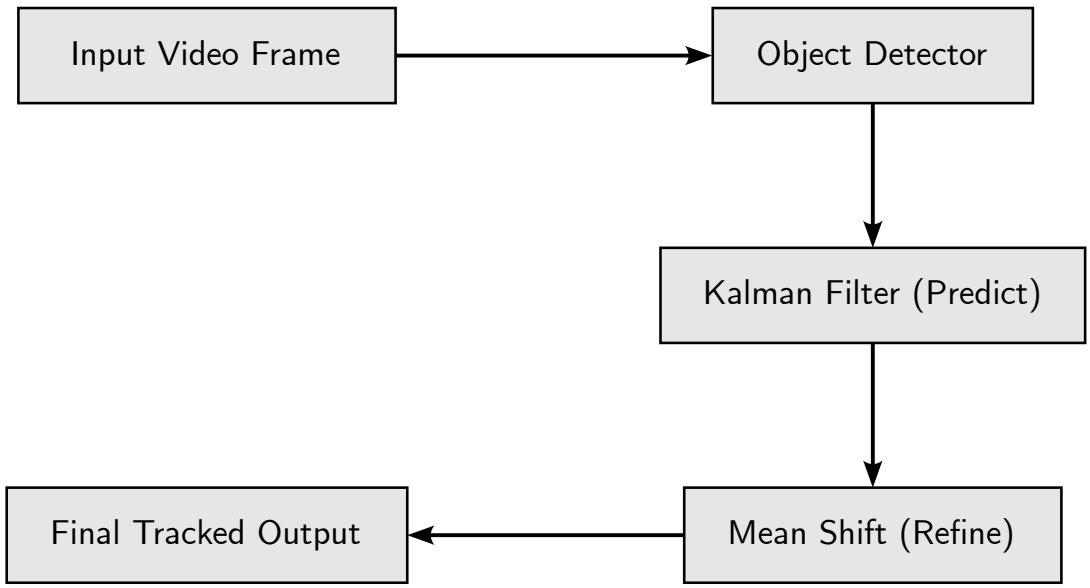

Figure 5: Architecture of a Hybrid Tracking System

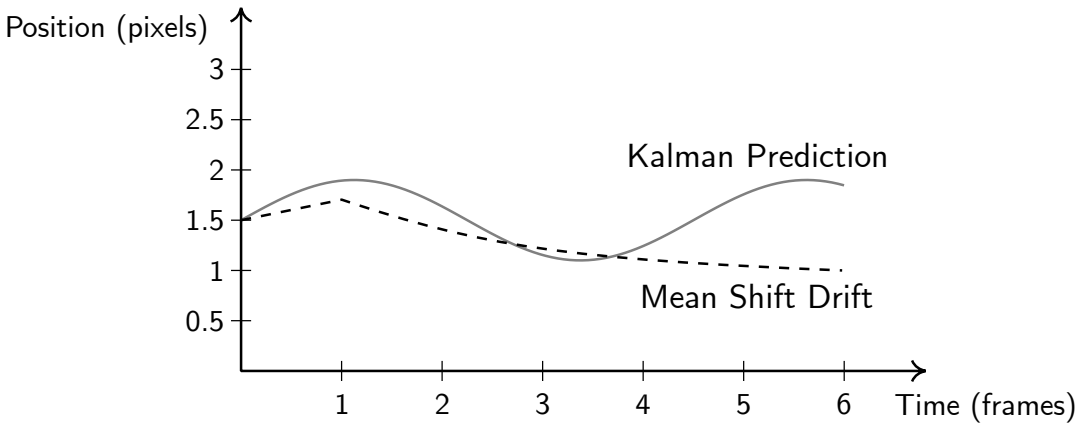

Figure 6: Comparison of Kalman Filter vs. Mean Shift on Curved Trajectory

robustness, preventing track loss during occlusions and smoothing trajectories. The consensus is that combining them—using the KF for robust trajectory maintenance and the Mean Shift for precise localization—yields the best overall performance [7].

- **Computational Efficiency:** A key shared advantage is speed. Both algorithms are computationally lightweight, making them suitable for real-time systems where deep learning models might be too slow. Their efficiency makes them valuable as components within more complex tracking frameworks or as standalone solutions for less demanding tasks.

The core trade-off is clear: Mean Shift's reactive, appearance-based search versus the Kalman Filter's predictive, motion-based estimation. Modern tracking challenges often require both, leading to the prevalence of hybrid designs.

6 Conclusion

The continued exploration of Mean Shift and the Kalman Filter in recent literature underscores their enduring value in the object tracking domain. While deep learning approaches now dominate high-performance benchmarks, these classical algorithms offer a compelling combination of efficiency, simplicity, and effectiveness that keeps them relevant. Mean Shift provides precise, appearance-driven localization, while the Kalman Filter offers a robust predictive framework for handling motion and occlusions.

A clear consensus from studies published between 2020 and 2025 is that the most effective use of these techniques often lies in their synergistic combination. By pairing the Kalman Filter's motion prediction with Mean Shift's appearance-based search, hybrid trackers can overcome the individual limitations of each method, resulting in systems that are robust to both occlusion and rapid motion. Quantitative evaluations on standard datasets consistently validate this approach, showing improved accuracy and robustness.

Ongoing innovations, such as adaptive Kalman filters and more sophisticated appearance models, continue to enhance the capabilities of these foundational algorithms. In an era of increasing complexity and computational demand, the efficiency and reliability of Mean Shift and the Kalman Filter ensure they will remain essential tools in the object tracking toolbox, whether as standalone solutions for resource-constrained applications or as integral components within more advanced, state-of-the-art tracking systems.

References

[1] Hung Duong and Arash Fahim. "Gradient ascent method for fully nonlinear parabolic differential equations with convex nonlinearity". In: *arXiv preprint arXiv:2406.06787* (2024).

[2] Brian John Julian. "Mutual information-based gradient-ascent control for distributed robotics". PhD thesis. Massachusetts Institute of Technology, 2013.

[3] Stefan Craciun et al. "A real-time, power-efficient architecture for mean-shift image segmentation". In: *Journal of Real-Time Image Processing* 14 (2018), pp. 379–394.

[4] Ido Leichter, Michael Lindenbaum, and Ehud Rivlin. "Mean shift tracking with multiple reference color histograms". In: *Computer Vision and Image Understanding* 114.3 (2010), pp. 400–408.

[5] Irene Anindaputri Iswanto, Tan William Choa, and Bin Li. "Object tracking based on meanshift and particle-kalman filter algorithm with multi features". In: *Procedia Computer Science* 157 (2019), pp. 521–529.

[6] Rahim Panahi, Iman Gholampour, and Mansour Jamzad. "Real time occlusion handling using Kalman Filter and mean-shift". In: *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*. IEEE. 2013, pp. 320–323.

[7] Priya Jain and Meenakshi Arora. "Optimized Object Tracking in Videos using Mean Shift and Kalman Filter". In: *International Journal of Research Publication and Reviews* 5.7 (2024), pp. 1518–1524.

[8] Fatih Porikli and Oncel Tuzel. "Object tracking in low-frame-rate video". In: *Image and Video Communications and Processing 2005*. Vol. 5685. SPIE. 2005, pp. 72–79.

[9] Pushkar Kadam, Gu Fang, and Ju Jia Zou. "Object tracking using computer vision: a review". In: *Computers* 13.6 (2024), p. 136.

[10] Bikal Lamichhane. "DirectionSORT: Multi-Object Tracking using Kalman Filters and the Hungarian Method with Directional Occlusion Handling". MA thesis. Villanova University, 2022.

[11] Megan Chung. *MULTI-OBJECT TRACKING USING KALMAN FILTER AND HUNGARIAN ALGORITHM*. Technical Report or Student Project. 2024.

[12] Jiahong Li et al. "Adaptive Kalman Filter for Real-Time Visual Object Tracking Based on Autocovariance Least Square Estimation". In: *Applied Sciences* 14.3 (2024), p. 1045.

[13] MOTChallenge. *MOT20 Benchmark Leaderboard*. https://motchallenge.net/results/MOT20/. Accessed: July 1, 2025. 2025.

[14] Jinkun Cao. "Improving Kalman Filter-based Multi-Object Tracking in Occlusion and Non-linear Motion". MA thesis. Pittsburgh, PA: The Robotics Institute, Carnegie Mellon University, Mar. 2024.

[15] Shiqi Liu et al. "Convolutional Unscented Kalman Filter for Multi-Object Tracking with Outliers". In: *IEEE Transactions on Intelligent Vehicles* (2024).

[16] Hsiang-Wei Huang et al. "Exploring learning-based motion models in multi-object tracking". In: *arXiv e-prints* (2024), arXiv–2403.

[17] Khizer Mehmood et al. "Context-aware and occlusion handling mechanism for online visual object tracking". In: *Electronics* 10.1 (2020), p. 43.

[18] Takahiro Kawabe et al. "Deformation lamps: A projection technique to make static objects perceptually dynamic". In: *ACM Transactions on Applied Perception (TAP)* 13.2 (2016), pp. 1–17.

[19] Bibek Das et al. "MMLT: Efficient object tracking through machine learning-based meta-learning". In: *Results in Engineering* 26 (2025), p. 104768.