Multi-Object State Estimation using Probabilistic Belief-Based Trackers

Connecting Low-Frequency Detection and High-Rate Prediction on Embedded Devices

Vitalii Mashkov[1] Qing Wang[1] Nitish Kumar[1]

[1] Delft University of Technology

1 Introduction

- High-accuracy detectors force edge hardware into a **low-frequency mode**, introducing latency.
- Standard trackers assume frequent updates and fail to bridge the **large spatial uncertainty gaps** during observation gaps.
- **Probabilistic Belief Tracker** resolves this by decoupling belief propagation from perception.

RQ: To what extent can a dual-rate, asynchronous belief-based architecture maintain object-state continuity and identity consistency under sparse detections on resource-limited hardware?

2 Background

- Standard Kalman filters [6] use a **unimodal representation** that fails when sparse updates violate the constant-velocity assumption.
- Without continuous observations, error accumulation causes **search gates to expand and overlap**, leading to identity switches (IDS W).
- Detector latency introduces **Out-of-Sequence Measurements (OOSM)** [1], creating a temporal mismatch.

Motivation & Requirements

Maintaining a continuous world-state on a **low-power edge budget** while satisfying Active Inference demands to minimize planning 'surprise' [3] requires:

a. **Temporal Belief Continuity:** Maintain a continuous, high-frequency state distribution rather than relying on frame-bound updates.
b. **Identity Stability:** Maintain consistent identities across observational gaps to avoid 'surprise' spikes from track fragmentation.
c. **Latency Compensation:** Retroactively realign delayed detections to prevent prediction errors from stale data.

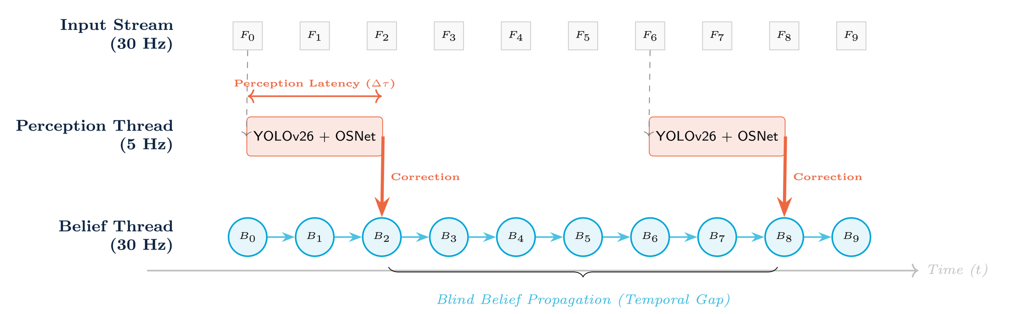

Figure 1: Asynchronous rate-decoupled pipeline.

3 Methodology: Belief Tracker

The tracker resolves the trade-off between detection lag and tracking stability using probabilistic dual-rate architecture.

Belief State Representation

- **Kinematic state X:** Tracks bounding box center, aspect ratio, height, and velocities.
- **Motion Hypotheses:** Branches GSF components into competing paths to capture spatial uncertainty during gaps.
- **Appearance state A:** Tracks deep visual embeddings via an Incremental GMM (IGMM) to preserve identity.

Belief Propagation (High-Rate)

- Belief propagation runs at camera frequency (**30 Hz**) for continuous state estimation.
- During gaps, the tracker branches GSF components into parallel motion hypotheses (Fig. 3).
- Over large gaps, numerical sub-stepping (**50 ms**) maintains filter stability (Fig. 3).

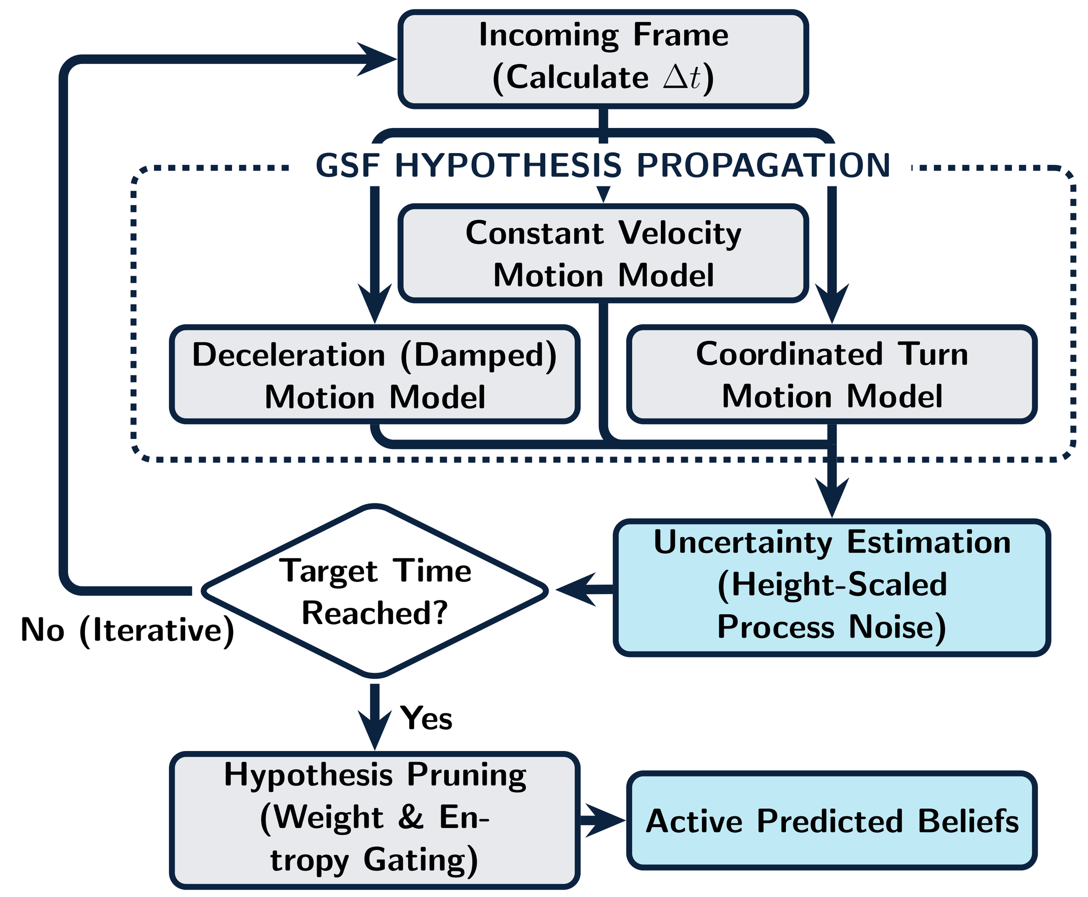

Figure 2: High-rate belief prediction pipeline.

Belief Update (Low-Rate)

- Detector latency creates a temporal mismatch on belief timeline (Fig. 1).
- Rollback mechanics **rewind the belief history** to apply updates at their original capture times.
- **Multi-Stage Association** matches detections sequentially using hybrid likelihoods (Fig. 1).
- **Lifecycle Management** controls track births and merges redundant hypotheses below 1.2 Mahalanobis distance.

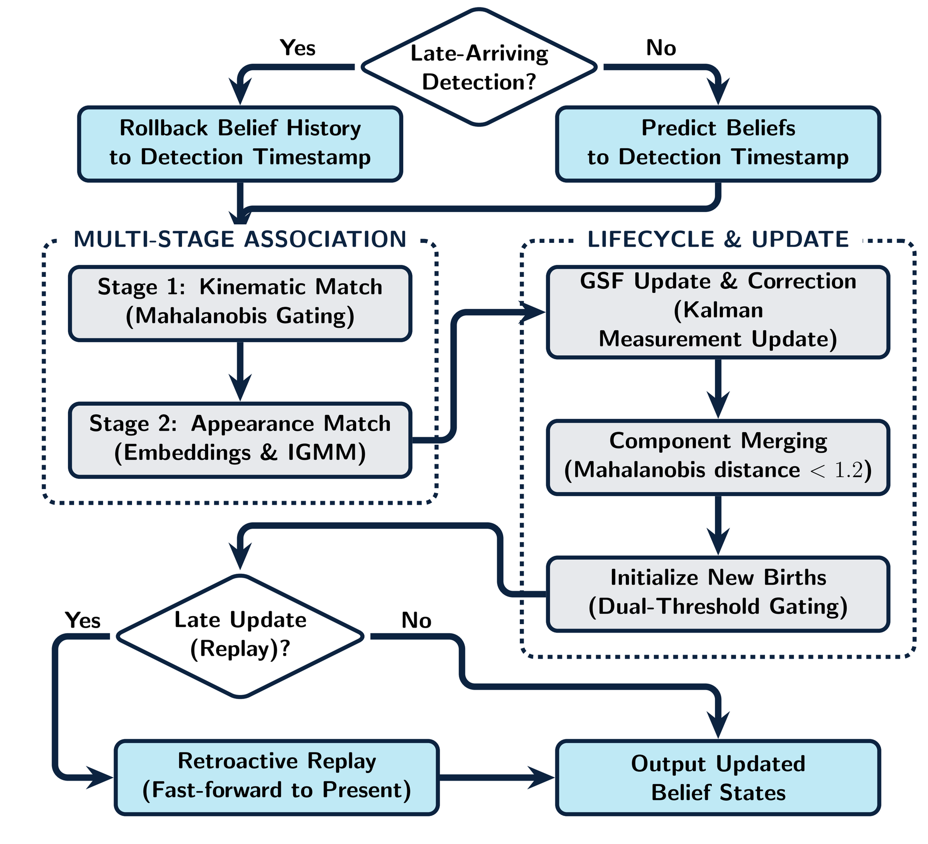

Figure 3: Low-rate belief update pipeline.

The **Rollback-and-Replay (RR) mechanism** retroactively realigns delayed detections by rewinding belief history, avoiding the computational overhead of parallel track histories.

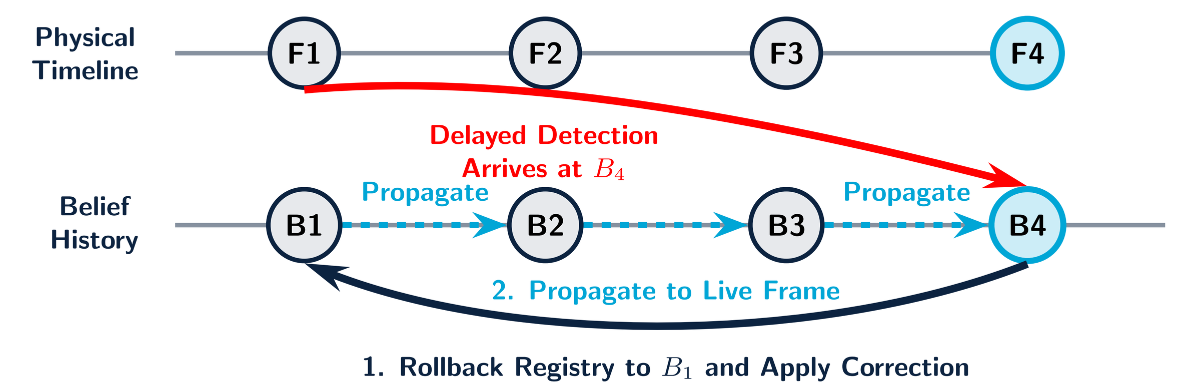

Figure 4: Rollback and replay timeline realignment.

4 Results

Sensitivity Analysis of Frequency and Latency

Sweeping tracking frequency on the MOT17 validation set [4] shows HOTA degrades under sparse detections, yet identity stability is preserved, maintaining 157 IDSW even at 1 FPS (Fig. 4). Under detector lag, identity switches decrease as delay increases (Fig. 4). The RR mechanism accumulates prediction uncertainty during replay and starts new trajectories instead of making incorrect associations.

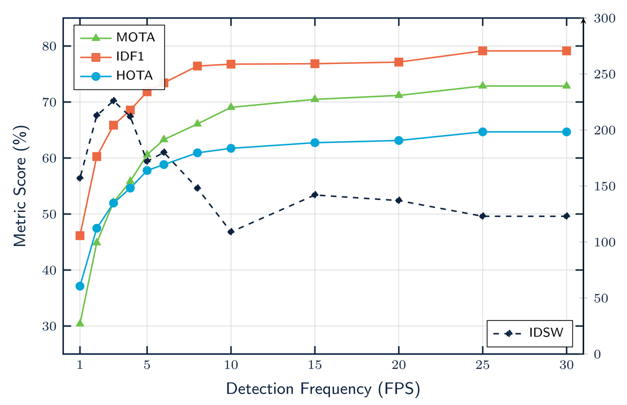

Figure 5: Metric scores and identity switches over detector frequency.

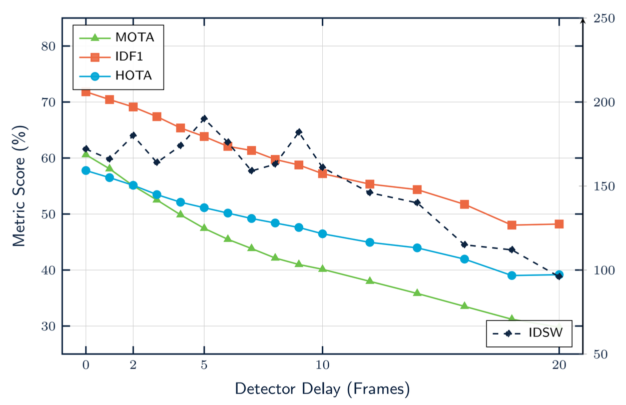

Figure 6: Metric scores and identity switches over detector latency.

Visual Tracking and Continuous Identity at 1 FPS

Propagating motion beliefs maintains target identity across significant spatial gaps, as seen in Fig. 1.

Figure 7: Tracking comparison on MOT17-04: 1 FPS (left) vs. 30 FPS (right).

Decoupled Architecture vs. Synchronous Baselines

Under the empty-detection protocol at 5 FPS, synchronous trackers like SORT [2] and StrongSORT [5] drop to **0.00% HOTA** due to track fragmentation. Decoupling belief propagation from perception maintains track continuity at **57.77% HOTA**.

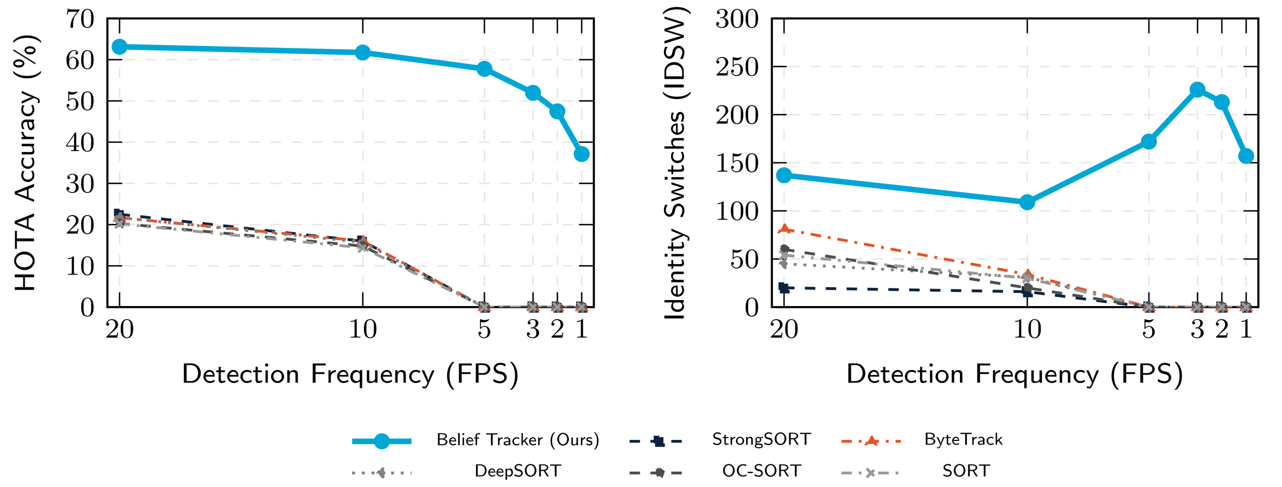

Figure 8: Tracking accuracy and identity switch comparisons against baselines.

Edge Profiling: NVIDIA Jetson Nano (10W MAXN, 3-Object Scenario)

Asynchronous Rates: 30 Hz Belief Thread (CPU) • 5 FPS Detection (GPU)
Resource Utilization: 72.3% CPU Load • 2.13 GB Memory Stabilization
Processing Latency: 17.0 ms Prediction • 39.0 ms Update

5 Conclusion

Decoupling belief propagation from perception resolves the trade-off between detection lag and tracking stability. The system achieves low-latency edge tracking by combining:

- Decoupled, high-rate motion belief propagation
- Hypothesis branching for spatial uncertainty
- Retroactive realignment for delayed updates

Continuous beliefs preserve target identity under sparse, delayed updates where standard trackers fail, establishing an efficient world model for edge systems.

6 References

[1] Y. Bar-Shalom. "Update with out-of-sequence measurements in tracking: exact solution". In: *IEEE Transactions on Aerospace and Electronic Systems* 38.3 (2002), pp. 769–777.
[2] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft. "Simple online and realtime tracking". In: *2016 IEEE International Conference on Image Processing (ICIP)* (2016), pp. 3464–3468.
[3] L. Da Costa, T. Parr, N. Sajid, S. Veselic, V. Neacsu, and K. Friston. "Active inference on discrete state-spaces: A conceptual and mathematical tutorial". In: *Journal of Mathematical Psychology* 99 (2020), p. 102447.
[4] P. Dendorfer, H. Rezatofighi, A. Milan, A. Bakic, F. Galasso, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler. "MOTChallenge: A benchmark for multi-target tracking". In: *arXiv preprint arXiv:2010.07547* (2020).
[5] Y. Du, Y. Song, B. Yang, and Y. Zhao. "Strongsort: Make deepsort great again". In: *IEEE Transactions on Multimedia* (2023).
[6] R. E. Kalman. "A new approach to linear filtering and prediction problems". In: *Journal of Basic Engineering* 82.1 (1960), pp. 35–45.

vmashkov@tudelft.nl