Technical Report on Multi-Object Tracking using Kalman Filters and Data Association with Occlusion Handling

Introduction

This report shows how to use the KITTI Object Tracking Benchmark Suite to build a multi-object tracking (MOT) system for autonomous cars. The system incorporates a Kalman Filter for motion prediction, the Hungarian algorithm for data association, and RANSAC for reliable tracking, especially in difficult situations such as occlusions and noisy detections. The approach focuses on tracking moving objects across image frames, updating object states, and handling occlusions, which occur when objects temporarily disappear from view.

Methodology

1. Kalman Filter-Based Tracking

The Kalman Filter is a critical component of the tracking system. It anticipates each observed object's future position and modifies its state in response to fresh measurements (detections). The Kalman Filter model has eight state variables, including position and velocity, and four measurement variables, which are the bounding box coordinates (x1, y1, x2, and y2).

- State Transition Matrix (F): It represents the relationship between the position and velocity of objects. This matrix forecasts the object's future state based on its current position and velocity.
- Measurement Matrix (H): It relates the projected state to observed measurements, with a focus on positional variables.
- Process Noise (Q) and Measurement Noise (R): These matrices account for uncertainty in the model and measurement procedures, respectively. The process noise is set to be low for the position components, while the measurement noise is high for the velocity components.

2. Data Association Using the Hungarian Algorithm

The Hungarian algorithm is used to connect detections with existing trackers using Intersection over Union (IoU) scores. The IoU score between a tracker's anticipated position and a new detection is generated, and the Hungarian algorithm reduces the total cost of matching detections to trackers.

- IOU Calculation: Calculate IOU for each pair of detection and tracker bounding boxes to identify the best-matched pairs.
- Matching Threshold: A threshold of 0.35 is used to decide if a detection and a tracker should be paired. If the IoU is less than this threshold, the detection and tracker are deemed unmatched, and the tracker is either updated with no detection (indicating possible occlusion) or terminated if it has not been updated in a specified number of frames.

3. Robust Tracking with RANSAC

RANSAC (Random Sample Consensus) is used to eliminate outlier detections and provide accurate tracking in the presence of noisy data. RANSAC iteratively selects random subsets of detections, fits a model (linear in this example), and identifies inliers using a distance threshold. The best-fitting model is chosen, and only the inliers are used for subsequent processing..

- Normalization: Prior to using RANSAC, detections are normalized to account for data fluctuations and improve model fitting resilience.
- Thresholding: A threshold is used to filter out inliers and only evaluate detections that are near to the fitted model.

4. Handling Occlusions and Object Re-entry

Managing occlusions and object re-entry is critical in real-world tracking systems. Trackers in this approach keep track of the "time since last update" counter. If a tracker does not receive an update after a certain number of frames (indicating an occlusion), it is temporarily held with a lower confidence level. When the object reappears, the tracker can continue with updated measurements.

- Predicting Occluded States: During occlusion, the tracker uses the Kalman Filter to forecast the object's future state and is used for matching when the object reappears.
- Velocity Estimation and Smoothing: Velocity is estimated and smoothed using the difference between current and prior places. Additionally, a smoothing function is used to lessen the influence of noise and rapid changes in object motion.

5. KITTI Dataset and Evaluation Metrics

The system is tested using the KITTI Object Tracking Benchmark Suite, which contains labeled sequences of real-world driving events. Ground truth annotations include object bounding bounds for each frame, allowing for more exact assessment.

Evaluation Metrics:

- Precision and Recall: This assess object identification and tracking accuracy, comparing true positives (properly tracked objects) to false positives (incorrect detections) and false negatives (missing objects).
- F1-Score: It combines precision and recall into a single score for comparison.
- MOTA (Multi-Object Tracking Accuracy): It considers false positives, false negatives, and ID switches to assess total tracking accuracy.
- ID Switches: This indicator measures the number of times a tracked object is issued a different ID, indicating data association difficulties.

6. Results and Discussion

The tracking system was assessed using the KITTI dataset, specifically sequence '0002'. The evaluation metrics provide information about the tracking system's performance in comparison to the actual data.

Results:

- Precision: 0.8856
- Recall: 0.9995
- F1-Score: 0.9391
- MOTA: 0.6139
- ID Switches: 538.0000

These results show that the system tracks objects well under normal settings. However, issues develop during occlusions and object re-entry. Although RANSAC enhances robustness in the face of noise, there are still circumstances where ID switches occur due to complex occlusions or overlapping objects.

- Handling Occlusions: The system effectively handles short-term occlusions by estimating the object's position and then restarting tracking when it reappears. Long-term occlusions, or instances in which several objects overlap over extended periods of time, cause ID switches and lower accuracy.
- Comparison with KITTI Ground Truth: The system's performance when compared to the KITTI ground truth is competitive, particularly in terms of recall and F1-Score. The MOTA score, while slightly lower, reflects the difficulties of resolving occlusions and preserving object identities in complicated scenarios.

7. Conclusion and Future Work

The suggested MOT system successfully combines Kalman Filters, the Hungarian algorithm, and RANSAC to accomplish reliable tracking in real-world applications. The system performs well on the KITTI dataset, with strong precision and recall metrics.

Future Work:

- Improved Occlusion Handling: Deep learning-based re-identification could reduce ID switches and boost MOTA ratings.
- Adaptive RANSAC Thresholding: Changing the RANSAC threshold based on the scene context can improve resilience.
- Using 3D bounding boxes and point clouds from LiDAR data can improve object tracking and localization accuracy.