Estimation of detection probability in multitarget filters using object advanced image processing techniques

Author: Ing. Michal Seibert | **Supervisor:** doc. Ing. Kamil Dedecius, Ph.D.
Czech Technical University in Prague, Faculty of Information Technology

FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Motivation

This work focuses on the problem of multi-target tracking with noisy measurements in a cluttered environment. Algorithms for tracking multiple targets with the uncertainty in the number of targets and their survival are often sensitive to the correct setting of parameters relative to the environment. One of the key parameters is the detection probability, which is often constant. In this work, a method to estimate this probability for each target at any time step is demonstrated through tracking objects in video footages. Advanced deep-learning algorithms for image processing are utilized to provide targets' measurements. Outputs of these methods are then exploited to calculate the detection probability. To analyse the capability of the proposed method, the Gaussian mixture probability hypothesis density (GM-PHD) filter is used. This filter falls among fundamental random finite sets statistics-based multitarget algorithms.

In real-world scenarios, the sensor performance is susceptible to various environmental conditions. Adverse weather, occlusions, or just the nature of the current scenario can lead to variations in detection probabilities. A mismatch between assumed and actual detection probabilities can result in a suboptimal tracking performance, leading to missed detections, false alarms, or inaccurate target state estimates.

Multi-target tracking establishment

Multi-target tracking (MTT), a fundamental aspect of surveillance and monitoring systems, has undergone significant advancements in recent years, transforming it into a critical field with diverse applications across various domains. The scope of MTT extends beyond mere tracking, encompassing tasks such as object detection, identification, and trajectory prediction. The primary goal is to maintain a comprehensive situational awareness, providing invaluable information for decision-making processes in various applications, ranging from defense and surveillance to autonomous systems and robotics.

Target tracking plays a pivotal role in surveillance systems, such as radars, where the primary objective is to estimate the trajectory, position, and other characteristics of moving objects within a surveillance area. Derived from the foundational principles of the Kalman filter, target tracking algorithms are tailored to the specific requirements and challenges inherent in radar applications. Radar-based target tracking encounters various complexities, including survivability and the presence of multiple targets in the same area. Survivability concerns the algorithm's ability to maintain accurate estimates despite target maneuvers, occlusions, or potential target loss.

These principles can be applied beyond radar systems. In this study, we focus on enhancement of target/object tracking on video footages.

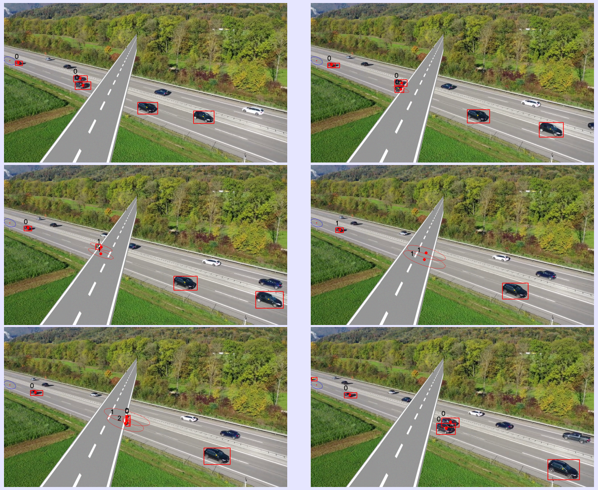
Object detection and segmentation models

Due to the existence of obstacles that cause the sensors to misdetect objects in certain situations, we focus on tracking objects in video data using target tracking algorithms with the dynamic detection probability. The image object detector serves as a sensor for obtaining measurements. As an object detector, we chose YOLO model [1] and Grounding Dino [2]. To obtain color characteristics of an object to compare them with the surrounding background of an object, we also need an object segmentation model. As a segmentation model, the Segment Anything [3] is utilized.

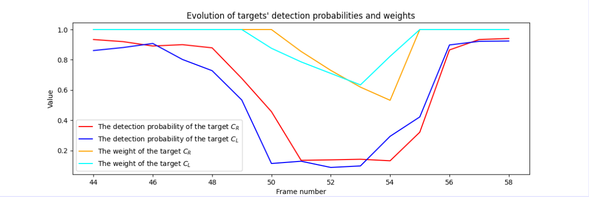
Experiments

Due to the lack of datasets including scenarios with targets that hide behind another object from the camera view, several experiments had to be made. Many of which are captured by ourselves, others are traffic videos with an added road that servers as a blockage of the camera's proper view to capture the objects under it.

The following figures represent a scenario with two cars on a road approaching a crossing road that unables the camera to capture the cars for a while. With our modified GM-PHD filter we are able to estimate the cars' positions together with covariance ellipses that show the uncertainty of the positions. The numbers above the targets mean the following states: 0 – detected, 1 – hidden, 2 – lost.



The following chart shows the evolution of the detection probability. Due to the lowered detection probability, the targets' weights do not decrease rapidly and thus the targets are able to survive.



Methodology

To track the targets in a video footage, the following steps occur:

- Detect the chosen objects in a scene by an object detection model.
- Segment the objects to obtain their characteristics.
- Compare gained color characteristics with color characteristics of a scene of a predicted object. This predicted position comes from the prediction step of the GM-PHD filter [4].

The GM-PHD filter assumes a target state x described by an intensity function $\nu(x)$. At time k, the prediction of the prior intensity $\nu_{k-1}(x)$ is given by

$$\nu_{k|k-1}(x) = \int p_{S,k}(\xi)\phi_{k|k-1}(x|\xi)\nu_{k-1}(\xi)d\xi + \nu_{\gamma,k}(x), \tag{1}$$

where $p_{S,k}(\cdot)$ is the probability of target survival, $\phi_{k|k-1}(\cdot|\cdot)$ is the target state transition density, and $\nu_{\gamma,k}(\cdot)$ denotes the prior PHD of the targets birth at time k. The predicted intensity $\nu_{k|k-1}$ is then updated by the measurement set Z_k given by sensors at time k according to the Bayesian update

$$\begin{aligned}\nu_k(x) &= [1 - p_{D,k}(x)]\nu_{k|k-1}(x) \\ &+ \sum_{z \in Z_k} \frac{p_{D,k}(x)g_k(z|x)\nu_{k|k-1}(x)}{\kappa_k(z) + \int p_{D,k}(\xi)g_k(z|\xi)\nu_{k|k-1}(\xi)d\xi},\end{aligned} \tag{2}$$

where $g_k(\cdot|\cdot)$ is the likelihood function, $p_{D,k}(\cdot)$ is the probability of detection, and $\kappa_k(\cdot)$ is the clutter density.

Conclusion

This work introduces an innovative approach aimed at estimating one of the key parameters crucial to tracking systems – the detection probability. This probability is profoundly influential in the survivability of tracked objects, making its accurate estimation paramount. Leveraging advanced deep-learning image processing techniques, we propose a method to estimate the detection probability for each target at every time step within video footage. Furthermore, we address the possibility of the target misdetection, a common occurrence when obstacles obstruct the camera's view.

References

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.

[2] S. Liu, Z. Zeng, T. Ren, *et al.*, *Grounding dino: Marrying dino with grounded pre-training for open-set object detection*, 2024. arXiv: 2303.05499 [cs.CV].

[3] A. Kirillov, E. Mintun, N. Ravi, *et al.*, *Segment anything*, 2023. arXiv: 2304.02643 [cs.CV].

[4] B.-N. Vo and W.-K. Ma, "The gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, 2006. DOI: 10.1109/TSP.2006.881190.