ME5413: Autonomous Mobile Robot

Homework 1: Perception

Name: WANG Yexiang   Matric Number: A0304719A

Task1: Single Object Tracking

In Task 1, I used three methods: ***template matching***, ***Object detection algorithm and association*** and ***template matching with Kalman filtering*** (improved method). First, I will explain my thought process for each method individually. Then, I will present the final result table and relative analysis. Finally, I will analyze the problems encountered in this assignment and propose possible improvements.

Method and Code Implementation

Template Matching Methods

First, the ***groundtruth.txt*** file is read, and the template is extracted from the first frame of the image. Then, using the template matching function provided in the ***cv2*** library, the template and the next frame are passed into ***cv2.matchTemplate***. The returned data is then processed to match the required format for the assignment. This process is repeated in a loop, and finally, all data is stored in ***trackresults_TM_seqX.txt***.

Notably, I set a search range in this function, allowing template matching within a restricted area to reduce interference and improve tracking accuracy. Additionally, I update the template after processing each frame to enhance search precision.

Object Detection Algorithm and Association

In the Object Detection Algorithm and Association method, I followed the assignment requirements by loading the model and passing each frame into it for processing. I then selected the detected target with the highest score and returned its coordinates. The returned data was processed to match the required format for the assignment. This process was repeated in a loop, and finally, all data was stored in ***trackresults_ODA_seqX.txt.***

Improved Method: Template Matching Methods with Kalman Filtering

Building on the template matching method, I incorporated Kalman filtering. During tracking, for each frame, Kalman filtering first predicts the target's position. This predicted position is then used as the center of the search region for template matching or to adjust the search area, thereby reducing the search range and improving efficiency and accuracy. The rest of the process remains the same as template matching. Finally, all data is stored in ***trackresults_improved_seqX.txt***.

Results and Relative Analysis

Results Tables

		Seq1	Seq2	Seq3	Seq4	Seq5	Average
1	**Template Matching**	0.35	0.51	0.79	0.63	0.18	0.492
2	**Detection with Association**	0.61	0.37	0.26	0.99	0.83	0.612
3	**Improved Method**	0.37	0.51	0.79	0.65	0.18	0.5

Table 1.1 Evaluation of SOT methods using ***Precision*** metrics

		Seq1	Seq2	Seq3	Seq4	Seq5	Average
1	**Template Matching**	0.35	0.48	0.12	0.32	0.74	0.402
2	**Detection with Association**	0.39	0.37	0.33	0.01	0.17	0.254
3	**Improved Method**	0.33	0.48	0.12	0.31	0.74	0.396

Table 1.2 Evaluation of SOT methods using ***Success*** metrics

Results Analysis

In the Detection with Association method, it can be observed that when there are fewer similar objects in the image, the precision performs significantly better. This is likely because some frames in multi-frame detection may contain incorrect target detections.

In comparison, for template matching, when there is no object overlap or occlusion, the detection results tend to be much better.

However, in some cases, the success rate of the Detection with Association method is very poor, likely due to persistent detection errors by the model, such as in Seq4 and Seq5. On the other hand, template matching may incorrectly track the wrong target during movement, leading to results that are partially correct and partially incorrect.

It is speculated that the introduction of Kalman filtering in the improved method did not enhance template matching because it did not address the issue of tracking loss or incorrect target tracking. As a result, the data in the third row is not significantly different from that in the first row.

Problems and Possible Improvements

1. In the template matching method, tracking loss or errors may occur due to object occlusion. Possible solutions include restricting an appropriate search range, applying suitable preprocessing to the images, and incorporating trajectory prediction to improve tracking accuracy.

Specifically, taking the images from seq1 as an example, teammates wear identical clothing and have similar appearances. This leads to tracking loss when they are running. Moreover, since I update the template continuously, an incorrect match will result in persistent errors in subsequent frames. As shown in the image below (**Figure1**), this is an example of a tracking error: the red box represents the template matching result, while the green box represents the ground truth.

Figure 1 Tracking Error in Case Seq1

2. When using the model for object detection, I encountered memory shortages and Jupyter Notebook kernel crashes. Possible improvements include: 1) Applying appropriate image preprocessing. 2) Incorporating trajectory prediction to perform object detection within a smaller search area.

3. The parameters of the Kalman filter, such as the covariance matrices for process noise and measurement noise, need to be appropriately set. These parameters should ideally be adjusted based on the specific scenario. However, for this task, I used the default values. Future optimizations can be achieved by fine-tuning these parameters to improve performance.

Task2: Multi Object Prediction

In Task 2, the ***constant velocity model*** and ***constant acceleration model*** were used to predict the future trajectories of all target agents for 3, 5, and 8 seconds. The **ADE** (Average Displacement Error) and **FDE** (Final Displacement Error) were then calculated for each agent. Next, I will briefly explain the final results obtained, the problems that happened, and possible improvements.

Results and Relative Analysis

As shown in the figure below (**Figure2**), this is one of the results. The red line represents the ground truth, the green line represents the result obtained using the constant velocity model, and the blue line represents the result obtained using the constant acceleration model.

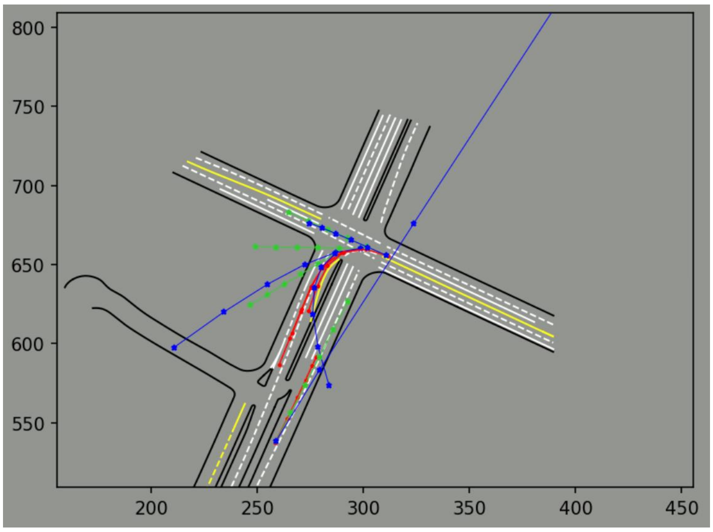

Figure 2 Prediction Results

It is evident that during turns, the green line, which represents the constant velocity model, deviates significantly from the ground truth. On the other hand, the blue line, representing the constant acceleration model, is relatively close to the ground truth at the beginning of the turn. However, as time progresses, the constant acceleration causes the velocity to increase excessively, leading to overestimation. This can also be observed in the figure, where some predicted trajectories extend beyond the image boundaries.

Correspondingly, as time increases, the predicted trajectories deviate more and more from the actual path, resulting in high **ADE** (Average Displacement Error) and **FDE** (Final Displacement Error). Overall, these two models are relatively simple—while they can achieve small errors for short-term predictions, their performance deteriorates significantly over longer time horizons.

Problems and Possible Improvements

1. There are cases where the groundtruth appears **shifted**, as shown in the figure below (**Figure3**). After analysis, it was found that this issue arises because some ground truth data is unavailable or unreliable. To address this, preprocessing should be performed to remove unusable data before using it.

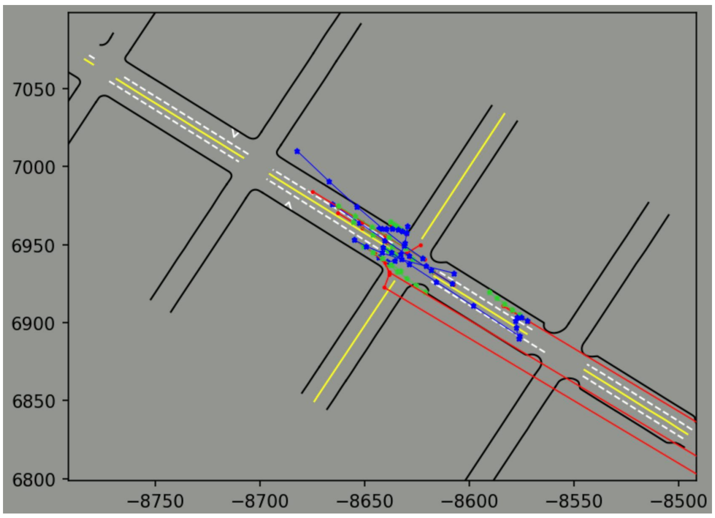

Figure 3 Ground Truth Trajectory Shifted

2. Both **ADE** and **FDE** are relatively large, so better models should be introduced, such as **LSTM** or **Kalman filtering**. Additionally, the prediction time should not be too long, and data should be updated in a timely manner.

Task 4: Evaluation

How much time did you spend on this homework for each task?

Since I was not familiar with the related content beforehand, I spent a lot of time on code modifications. The overall approach for the assignment is actually not difficult. The main challenge lies in optimization and improvement, which were not discussed in much detail during class and may require a longer time to complete. I spent nearly **12 hours** completing this assignment.