Long term tracking

Aljaž Konec

I. Introduction

Objects in a given video sequence are sometimes ocluded or even completely move out of the field of view. In such cases, it is desirable for a tracker to be able to recover the object once it reappears in the field of view. To achieve this, we implement object re-detection using thresholding and sampling of possible locations of the object in the image.

II. Comparison of Short-term and Long-term Tracking

As a baseline short term tracker we used the SiamFC tracker that we later adopted to a long term tracker. Table I shows the comparison of the two trackers on the entire provided dataset. The use of re-detection results in 12% less false negative predictions, leading to significanty increas in recall.

	Precision	**Recall**	**F-score**
SiamFC	0.596	0.318	0.415
Long-Term SiamFC	0.588	0.436	0.501

Table I: Comparison of short-term and long-term tracking.

Confidence Score Thresholding

In our implementation, we used simple thresholding to determine if the object was lost. To define the confidence score, we first calculate reliability score q_t for each frame t as:

$$q_t = MAX(R_t) * PSR(R_t) \tag{1}$$

Where R_t is the response map of the tracker and PSR is the peak-to-sidelobe ratio. The confidence score is then defined as:

$$c_t = \frac{\overline{q_t}}{q_t} \tag{2}$$

where $\overline{q_t}$ is the mean over all past frames. Through experimentation, we set the threshold value to 3. Decreasing the threshold value would result in the tracker being more sensitive to changes in the object's appearance, thus running the redetection more often. This would come at the cost of FPS.

III. Sampling of possible Locations

Sampling new possible locations is dependent on two factors: the number of samplings and the sampling distribution.

Number of Samplings

Table II shows the comparison of the number of samplings on the entire provided dataset. We tested the following numbers of sampling points: 10, 30, 50 and 100 using a uniform distribution of sampling points. Using random sampling resulted in better performance than the short-term tracker. The best overall performance was achieved with 30 samplings and as such we used this value as a baseline model. Precision for all tested values was similar, which is expected as the number of false positve matches should not increase if more non target locations are sampled. A higher recall value means that the target was re-detected faster, leading to lower numbers of false negatives.

Number of Samplings	**Precision**	**Recall**	**F-score**
10	0.596	0.386	0.469
30	0.588	0.436	0.501
50	0.579	0.372	0.453
100	0.585	0.384	0.463

Table II: Comparison of the number of samplings.

Sampling Distribution

Sampling of new target locations can be done in multiple ways. For all previous cases we used uniform sampling over the entire image. A more sophisticated approach uses a Gaussian distribution around the last known location of the object and uses a growing standard deviation. Table III shows the results of using a Gaussian distribution for sampling. Here we implemented Gaussian samplin centered around the previous location of the object and an increasing standard deviation. The standard deviation starts at one half the size of the search window size and is increased by 10% for each next frame that has to be re-detected. It can be observed that the uniform distribution performs better than the Gaussian distribution. Upon further expection of the sequences we conclude that this is due to the targets appearing in completely different locations in the image.

Sampling Distribution	**Precision**	**Recall**	**F-score**
Uniform	0.588	0.436	0.501
Gaussian	0.564	0.316	0.405

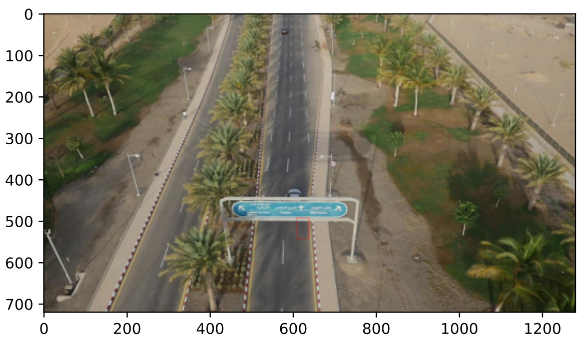
Table III: Comparison of sampling distributions.

IV. Visualizing the tracking results

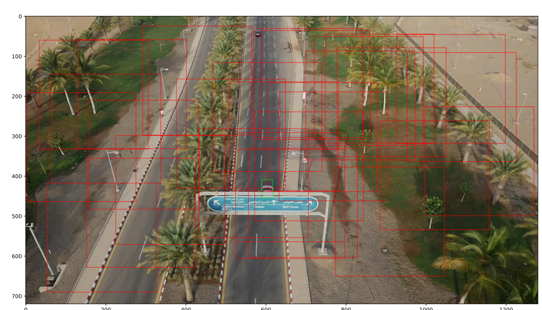
Figure 1 shows an example of the diffrence between the short-term and long-term tracker. The long-term tracker is able to recover the car after it is ocluded by the sign and continue tracking it.

V. Conclusion

Even a simple re-detection procedure can increase the tracking performance. Using adtional stratagies like video stabilization using Lucas-kanade and Kalman filtering could further improve the tracking performance. The drawback of this implementation is the number of samples that has to be drawn, leading to a drop in FPS performance.



(a) Short-term tracker looses target after occlusion.



(b) Long-term tracker and the new samples of possible locations in red.

Figure 1: Example of short-term and long-term tracking.