Exercise 5: Long-Term Tracking

Veljko Dudić

I. Introduction

The SiamFC short-term deep-CNN tracker was evaluated on the car9 sequence from a long-term tracking benchmark. The algorithm was enhanced with a re-identification module to resume tracking after occlusion or failure. Multiple experiments explored different threshold and number of samples settings to identify optimal parameters. Performance on the car9 sequence is reported in terms of precision, recall, and F-score.

II. Experiments

A. SiamFC: short-term tracker

The SiamFC short-term tracker was evaluated with the scripts' default parameters, and the same metrics specified in the instructions were applied.

Precision	Recall	F-score
0.64	0.27	0.38

Table I: SiamFC metrics on the sequence car9.

B. Long term tracking

In the second stage of the assignment, the tracker's update routine was revised and a re-detection procedure incorporated to handle short-term tracking failures. A failure is flagged whenever the correlation response drops below a predefined threshold. When this occurs, multiple patches are sampled uniformly across the full image; if the patch with the highest response exceeds the threshold, tracking switches back to short-term mode.

Figure 1 shows a frame where the short-term tracker loses the target, prompting activation of the re-detection module. Figure 2 presents the subsequent frame, in which the target is successfully re-identified and the system resumes short-term tracking.

Figure 1: When occlusion prevents the tracker from locating the car, the system uniformly samples multiple patches across the entire frame for evaluation.

Figure 2: After occlusion, the tracker fails to locate the car, so the system samples multiple patches uniformly over the entire frame. The red-highlighted patch yields a correlation response above the recovery threshold, allowing the tracker to resume short-term operation.

Long-term tracking performance hinges on two hyperparameters: the correlation-response threshold and the total number of sampled patches. On the car9 sequence, we evaluated a range of threshold and sample-count combinations. Setting the threshold too high leads to spurious failure detections—even when the target remains visible—while thresholds that are too low effectively disable re-detection, reverting to pure short-term tracking. Optimal values must be chosen empirically and will vary by application. As reported in Table II, the best overall precision–recall balance was obtained with a correlation threshold of 5.

Table II: Performance metrics for different correlation thresholds on the *car9* sequence.

Threshold	Precision	Recall	F-score
1	0.64	0.27	0.38
4	0.60	0.59	0.59
5	0.60	0.59	0.60
6	0.62	0.49	0.55
7	0.57	0.05	0.09
10	0.29	0.03	0.05

An additional study examined how the number of redetection patches affects tracker performance. As shown in Table III, precision, recall and F-score steadily improve with more samples, peaking at six patches (F-score 0.60). Beyond six (8, 10, 15, 30 samples), the F-score only falls marginally to 0.59, while the computational cost continues to climb. Using fewer than six patches yields noticeably lower accuracy (e.g. F-score 0.54 at one sample). These results suggest that six samples provide the optimal balance between re-detection reliability and execution speed, although applications demanding even higher confidence could still benefit—albeit minimally—from larger sample counts.

Table III: Performance metrics for varying sample counts on the *car9* sequence.

Threshold	Precision	Recall	F-score
1	0.61	0.48	0.54
3	0.59	0.58	0.59
5	0.60	0.59	0.59
6	0.60	0.59	0.60
8	0.61	0.57	0.59
10	0.60	0.58	0.59
15	0.60	0.59	0.59
30	0.60	0.59	0.59

III. Conclusion

This study assessed the performance of SiamFC—a short-term CNN-based tracker—on the car9 sequence from the provided long-term tracking dataset. To enable robust re-acquisition after occlusions or tracking failures, the baseline was extended with a redetection module. A series of experiments varying the correlation-response threshold and the number of redetection samples revealed that optimal performance is achieved with a threshold of five and six uniformly sampled patches. Future work could focus on parallelizing the sampling and correlation-response computations and investigating Gaussian-based patch sampling to further improve speed and accuracy.

References