Exercise 5: Long-Term Tracking

Matic Isovski, mi6568@student.uni-lj.si

I. Introduction

In this paper, we present comparison of results obtained using short-term and long-term SiamFC tracker. We also present the impact of parameters on re-detection capability of long-term tracker. An example of re-detection is also presented.

II. Experiments

In the first part, we ran short-term tracker SiamFC on car9 sequence and got these results:

Precision	Recall	F-score
0.6386	0.2712	0.3807

Tracker was run on just one sequence, because we do not own any GPU that we could use for performance boost.

Later, we extended tracker into a long-term tracker. Precision slightly dropped, but Recall and F-score improved significantly. Here are results for sequence car9 using 5 of randomly sampled regions during re-detection:

Precision	Recall	F-score
0.5977	0.5917	0.5947

For detecting if tracker got lost, we used adaptive thresholding:

$$t = score_{global} \times \delta$$

where score is defined as:

$$score_{global} = (1 - \alpha) \times score_{global} + \alpha \times score_{current}$$

We found that the best results were obtained using $\delta = 0.6$ and $\alpha = 0.1$. If maximum response was below threshold, tracker started to re-detect. Once maximum response exceeded threshold, re-detection was no longer needed.

We also tested different number of randomly sampled regions during re-detection, and how they impact re-detection capability. Results are shown in a table:

Samples	Precision	Recall	F-score	Frames
5	0.5991	0.5930	0.5960	13
10	0.6018	0.5957	0.5987	10
15	0.5977	0.5917	0.5947	9
20	0.5977	0.5916	0.5946	9
25	0.6015	0.5954	0.5984	9
50	0.6015	0.5955	0.5985	8

We can clearly see, that more samples bring better re-detection capability (fewer frames needed to re-detect), but only to one point. From 15 to 50 samples, there is no significant change.

In Figure 1 we can see an example when the tracker lost the target due to an overlap, followed by successful re-detection.

Figure 1. Re-detection example.

III. Conclusion

We can see how a minor change can improve tracker's reliability. Of course, result may differ if all the sequences would be used for testing.