Exercise 5: Long-Term Tracking

Gregor Zadnik

I. Introduction

In this exercise we were tasked with implementing a long-term tracker using the SiamFC short-term tracker. In this report I explain the experiments I did and present my results.

II. Experiments

A. Short-term tracker

To have a baseline result, I first ran the SiamFC short-term tracker on the car9 sequence. The results are presented in table I. The Precision value is pretty high, while the Recall and F-score are low due to the tracker getting lost when the target is occluded.

Precision	Recall	F-score
0.64	0.27	0.38

Table I
Results of the SiamFC short-term tracker.

B. Long-term tracker

I modified the SiamFC code to turn it into a long-term tracker and ran it again on the same sequence. The results in table II show us that precision was a bit lower while the recall and f-score measures drastically improved. This is due to the tracker recognizing when the target is no longer visible and finding it when it appears again, instead of wandering all over the image like in the previous case.

Precision	Recall	F-score
0.60	0.59	0.6

Table II
Results of the SiamFC long-term tracker.

C. Defining the confidence score

In order to perform long-term tracking we must know when the target is no longer in frame. We can do that by setting a threshold for the maximum correlation response (our confidence score). I tried setting the threshold (th) to start re-detecting to the following values: 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, and 5. The threshold to stop re-detecting was always $th + 2.5$, which empirically seemed to be a good value. Figure 1 shows that a low threshold ($th = 2$) is ideal for the car9 sequence because (in my case) that means a lower re-detection threshold which results in the tracker finding the target immediately after it is visible.

The threshold, however, must not be set too low because then the tracker thinks the target is always visible and acts the same as the short-term tracker. It also is not ideal for it to be set too high because at $th = 5$ the threshold to stop re-detecting is 7.5, which is too strict. This high value results in a lower recall and f1-score due to the tracker needing a lot of frames to find a correlation response higher than 7.5.

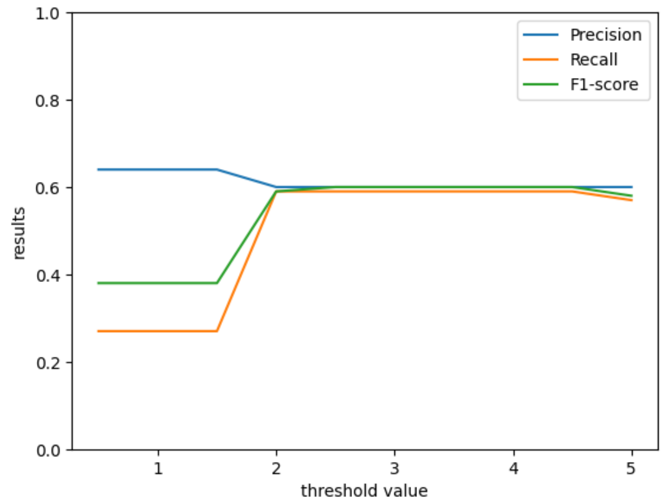

Figure 1. Tracker results changing with different confidence score thresholds.

D. Different number of sampled regions

After the target is no longer seen by the tracker, it starts sampling random regions of each frame to see where the target is most likely located. The number of regions (n) being taken into account can have a drastic impact on how fast the tracker finds the target again. Figure 2 shows how the number of frames needed to re-detect the target changes with the number of uniformly sampled regions.

The expected result would be that the number of frames would go down with n because a larger area of the image is checked per frame. This is mostly the trend in the curve with 2 exceptions at n=10 and n=60. The first could be explained with random sampling finding the correct region in the image quickly simply by chance. The second only has a 1 frame difference, which is essentially insignificant.

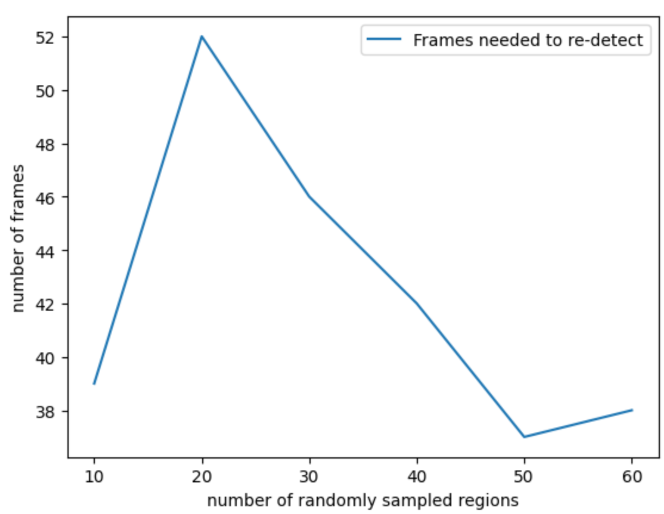

Figure 2. The number of frames needed to re-detect the target with respect to the number of randomly sampled regions.

E. Gaussian sampling

So far I only used uniform sampling to get random regions in the image. I also implemented and tried a gaussian sampling method, which samples around the last known position with a standard deviation. In the first approach the standard deviation is fixed to 3000 while in the second it starts at 3000 and gets multiplied by 1.1 every frame until the target is detected again. The results are presented in table III along with the baseline (uniform) sampling approach from section II-D. The Frames column is the number of frames needed for the tracker to re-detect the target.

Sampling	Precision	Recall	F-score	Frames
Uniform	0.60	0.59	0.6	52
Gaussian fixed	0.60	0.59	0.6	34
Gaussian growing	0.60	0.59	0.6	36

Table III
THE IMPACT OF SAMPLING APPROACHES ON TRACKING PERFORMANCE WITH NUMBER OF RANDOM SAMPLES SET TO 20.

We can see that the performance measures all stay the same while re-detection is considerably faster with gaussian sampling. Using the growing standard deviation makes the re-detection slightly slower, which could be explained by the target in the car9 sequence reappearing very close to the location where it disappeared and not far away.

III. Conclusion

I have implemented and tested the SiamFC long-term tracker. I tested some of the parameters and explained how they affect the performance of the tracker.