Exercise 3: Correlation filter tracking

Matej Miočić, 63180206, mm9520@student.uni-lj.si

I. Introduction

In this exercise we look into the correlation filter tracking. We implemented a simplified version as well as the actual MOSSE tracker. For each version we compare performance as the number of failures, average overlap and average FPS for their parameter choice. We explore improvements such as using a larger template or simple scale estimation. For the actual MOSSE we also explore improvements such as patch normalization, filter initialization and PSR measure. We compare the obtained results on the VOT14 dataset.

II. Experiments

A. MOSSE optimal performance

To obtain tracking performaces we integrated our tracker with toolkit-lite. For the optimal parameter choice we implemented a grid search. We ran the tracker with all combinations of parameters choice and selected the best one for our tracker. We searched for the following values:

1) **enlarge_factor** $\in \{1, 1.1, 1.2, 1.5, 1.7\}$,
2) $\sigma \in \{0.5, 1, 2, 3\}$,
3) $\alpha \in \{0.05, 0.1, 0.125, 0.15, 0.2\}$,
4) **PSR_threshold** $\in \{5, 7, 10\}$,
5) **size_of_training_set** $\in \{0, 8, 16\}$.

We obtained different but similar results for the simplified and actual MOSSE tracker. Since the overlap measure partially depends on the number of failures (with each fail, the bounding box resets creating a better overlap), we show optimal parameters for smallest number of fails (see Table I).

Table I: Optimal parameter choice for each version of MOSSE tracker, where **e** is enlarge factor, σ is smoothing of Gaussian peak, α is update parameter, **t** is PSR threshold and **s** is size of training set. Simplified version is **simp** and **real** is the actual tracker. Results obtained from all sequences from VOT14.

	e	σ	α	**t**	**s**	failures	overlap	FPS
simp	1.2	2	0.15	\	\	70	0.46	677
real	1.2	3	0.125	5	8	45	0.50	321

We can see that optimal actual mosse tracker has an improvement of 25 failures over the whole dataset, at the cost of FPS which is lowered by more than half.

B. Parameter influence

In this section we will go through each parameter with plots that show number of failures and average FPS with that parameter change. To further show that real MOSSE is superior we use simplified MOSSE optimal parameters (**enlarge_factor** $= 1.2$, $\sigma = 2$, $\alpha = 0.15$). For the parameters that only actual MOSSE uses, we keep the gained optimal (**PSR_threshold** $= 5$, **size_of_training_set** $= 8$).

1) Enlarge factor: Taking a larger patch than the annotated bounding box allows us to look at a larger surrounding area around the initial target. This allows us to detect objects that move fast and change their position drastically in between two sequential frames. With increasing the patch size we also observe lower average FPS.

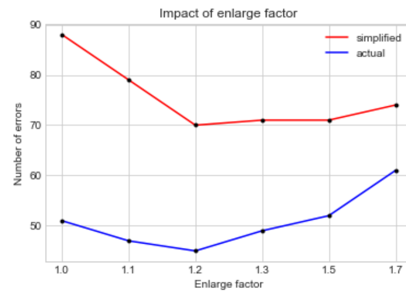

Figure 1: Number of failures based on **enlarge factor** for both simplified and actual version of MOSSE tracker.

In Figure 1 we observe major improvements when using a larger patch for both versions. We also observe rise in number of failures when taking a patch larger by a factor 1.2. We deduce that the target becomes too small and we take too much background information to be able to track the target successfully.

2) Sigma: With smoothing the Gaussian peak we tell our algorithm how precise we want to centralize our target object. With a small σ, we tell it that the corelation outputs should have a peak with a small radius. With increasing sigma, radius of peak gets bigger (see Figure 2).

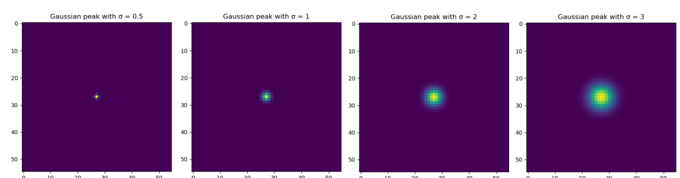

Figure 2: Gaussian peak for σ parameter choice 0.5, 1, 2 and 3.

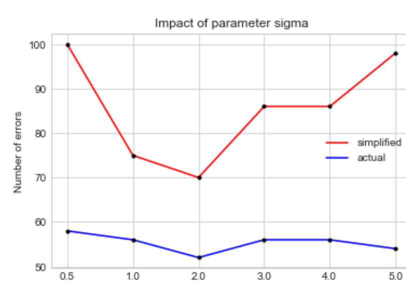

Figure 3: Number of failures based on **sigma** parameter for both simplified and actual version of MOSSE tracker.

In Figure 3 we observe that actual MOSSE does not change much based on σ, while simplified version is really dependant on using the right parameter.

3) Alpha: Updating our template H, allows the algorithm to adapt to our target that might change over time. Alpha parameter controls the update speed.

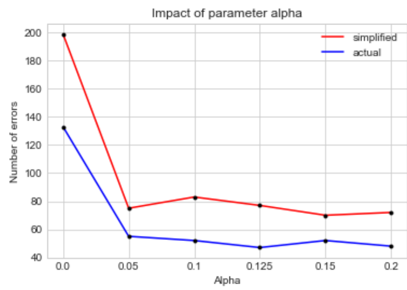

Figure 4: Number of failures based on **alpha** parameter for both simplified and actual version of MOSSE tracker.

In Figure 4 we see that not updating the template at all results in a lot of failures for both versions. This is because our dataset contains targets that are changing throughout the video. Updating the template allows the algorithm to adapt to these changes.

C. *Actual MOSSE parameter influence*

1) PSR threshold: "To compute the PSR the correlation output is split into the peak which is the maximum value and the sidelobe which is the rest of the pixels excluding an 11×11 window around the peak. The PSR is then defined as $\frac{g_{max}-\mu_{sl}}{\sigma_{sl}}$ where g_{max} is the peak values and μ_{sl} and σ_{sl} are the mean and standard deviation of the sidelobe."

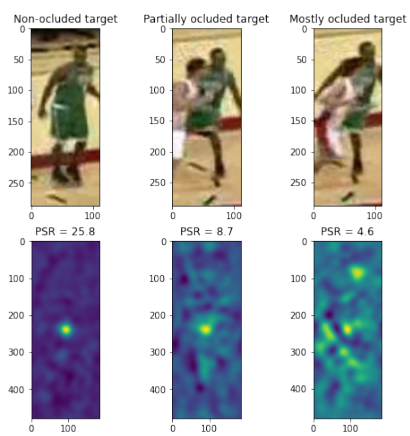

Figure 5: Examples of non-ocluded, partially ocluded and mostly ocluded targets and their correlation outputs. Above the correlation outputs we show their calculated PSR.

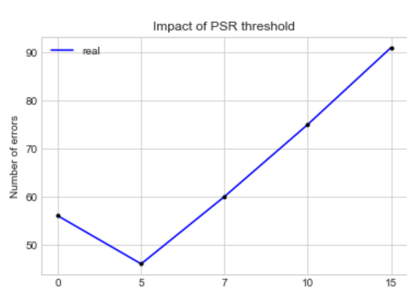

Figure 6: Number of failures based on **PSR_threshold** parameter for the actual version of MOSSE tracker.

In Figure 5 we show that ocluded targets have lower PSR. When a PSR is lower than **PSR_threshold**, we do not update the template. In Figure 6 we show how **PSR_threshold** influences number of failures.

2) Size of training set: We can try to tell our tracker in advance how our target can change, so we perform random affine transformations on the first frame and update our template. We used random $\pm$ 5 translation, $\pm$ 5 rotation and $\pm$ 0.1 scalling of our initial patch.

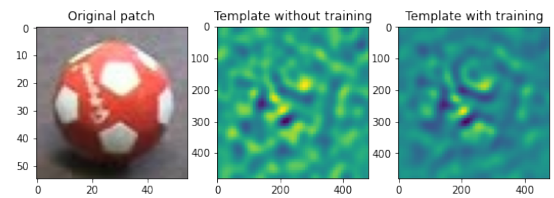

Figure 7: Change in template without and with training with 8 randomly affine transformed patches.

In Figure 7 we show how training with 8 randomly affine transformed patches changes the template. We can see it starts to resemble the shape of the ball more clearly.

D. *Tracking speed*

Table II: Average FPS to compute initial and tracking frame for both simplified and actual MOSSE tracker for first 10 sequences in VOT14.

	Simplified		Actual	
Sequence	Init FPS	Track FPS	Init FPS	Track FPS
Ball	429	499	167	319
Basketball	111	91	45	143
Bicycle	425	1099	87	132
Bolt	445	251	45	175
Car	996	375	22	172
David	333	165	58	330
Diving	153	153	61	290
Drunk	143	55	30	165
Fernando	117	93	28	80
Fish1	424	323	84	285
Average	358	310	63	209

In Table II we can observe that on average actual MOSSE performs slower than the simplified version. We also observe low values for FPS for initial frame. This is due to updating template with 8 randomly affine transformed patches.

III. Conclusion

We have explored the MOSSE tracker and compared simplifed and actual versions. We have compared parameters and found the parameters for optimal number of failures for each version. We have shown improvement ideas such as enlarging patch, updating template with randomly affine transformed initial patch, updating template only when PSR is higher than a threshold and how do these improvements affect the tracker. We have shown that the actual MOSSE outperforms the simplified version at the cost of computation speed. In the end we have shown how FPS changes for initial and tracking frame.