Exercise 2: Mean-Shift Tracking

Ana Poklukar

I. Introduction

In this exercise, we implemented and analyzed the mean-shift tracking algorithm [1], evaluating its performance through systematic experiments. Our investigation compared different parameter configurations on both synthetic data and the VOT2014 benchmark [2], measuring tracking accuracy and computational efficiency. Furthermore, we identified characteristic limitations of the approach. The study provides practical insights for parameter selection and reveals opportunities for algorithmic improvements in mean-shift based tracking systems.

II. Experiments

A. Synthetic Images

Firstly, we implemented mean-shift tracking and evaluated it on synthetic images featuring diverse patterns (peaks, multiple modes, ridges). Testing various kernel sizes and termination tolerances, we analyzed convergence behavior from distributed starting points to characterize attraction basins.

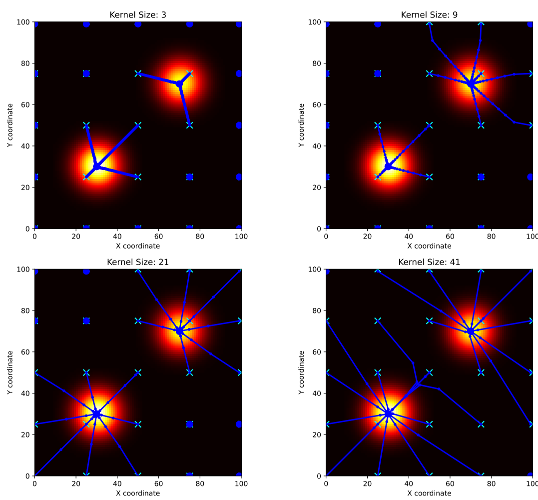

Figure 1. Convergence of the mean-shift algorithm with different kernel sizes.

Figure 1 demonstrates the kernel size tradeoff: smaller kernels converge precisely but only from nearby starting points, while larger kernels find global optima from farther away but require more computation. The broader kernel influence prevents local minima traps at the cost of slower convergence.

Table I
Average number of iterations for different kernel sizes.

Kernel Size	Avg. Iterations
3	66.5
9	21.6
21	9.40
41	7.20

Table I shows larger kernel sizes reduce convergence iterations, as their broader search area avoids local minima. However, this comes at the cost of reduced precision for small features, while smaller kernels offer finer localization but require more iterations.

Next, we tested how different tolerance values influenced the number of iterations required and the convergence behavior, while keeping the kernel size fixed at 15.

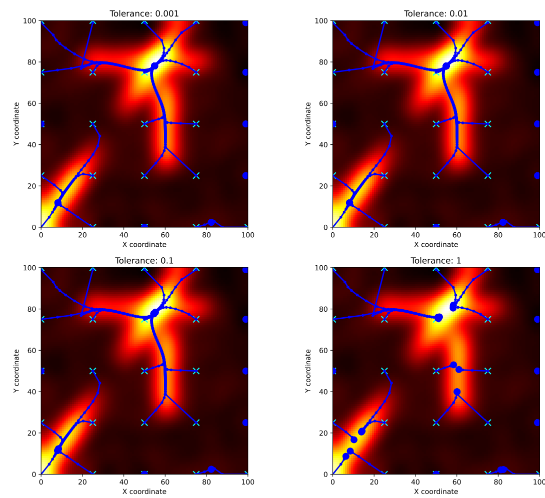

Figure 2. Convergence behavior of the mean-shift algorithm with different tolerance values.

Table II
Average number of iterations for different tolerance values.

Tolerance	Avg. Iterations
0.001	39.0
0.01	31.0
0.1	23.0
1	6.68

Figure 2 shows precise convergence with small tolerances, while larger values cause premature termination. Table II confirms this tradeoff - stricter tolerances require more iterations but achieve better optima.

In conclusion, convergence can be sped up by increasing the kernel size or tolerance. Larger kernels allow the algorithm to converge faster by considering a broader area and avoiding local minima, while larger tolerance values reduce the precision of convergence, leading to quicker termination. However, a trade-off between speed and accuracy must be considered when adjusting these parameters.

B. Visual Object Tracking Benchmark

After synthetic experiments, we tested the tracking algorithm on the Visual Object Tracking 2014 benchmark. All tests were conducted on an Apple M1 Pro chip.

Using optimal parameters (15px kernel, 16 bins, $\alpha = 0.01$, convergence threshold $\epsilon = 10e - 7$), our tracker achieved on average 470 FPS with 42 failures, mainly during occlusions and lighting changes (Fig. 3).

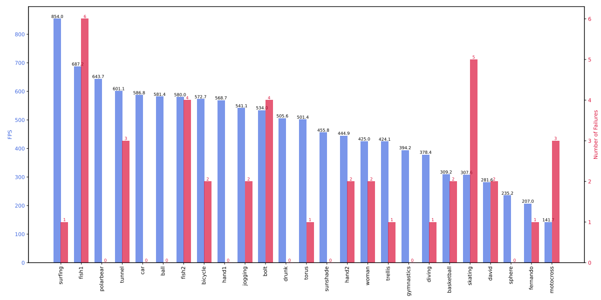

Figure 3. Tracking performance on VOT2014 benchmark sequences.

The tracker fails primarily during occlusions, lighting changes, and rapid motion (Fig. 4). Fast movements are particularly challenging as the fixed kernel cannot handle large displacements, while motion blur corrupts color histograms. This often causes convergence to similarly-colored background regions, highlighting mean-shift's fundamental limitations with dynamic targets.

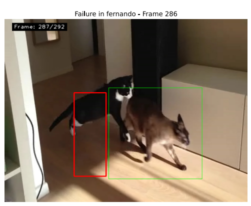

Figure 4. Tracking failure in the *fernando* sequence caused by rapid target movement.

To identify optimal tracking parameters, we conducted systematic experiments across the entire benchmark dataset, evaluating performance through both tracking failures and computational efficiency (FPS).

Kernel size tests (seen in table III) showed larger kernels reduce speed but maintain similar failure rates. The computational cost increases with kernel area, while failures stem primarily from occlusions and appearance changes rather than kernel selection. This suggests kernel size mainly affects processing speed without significantly impacting tracking robustness.

Table III
PERFORMANCE COMPARISON ACROSS KERNEL SIZES (FIXED: BINS=16, α=0.010, ϵ=1.0E-07) SHOWING TOTAL FAILURES AND AVERAGE FPS ON THE VOT2014 BENCHMARK.

Kernel Size	Total Failures	Avg. FPS
3	41	499
15	42	384
41	41	380

Our evaluation of histogram bins in table IV revealed 16 bins offers the best balance - sufficient color discrimination without excessive noise sensitivity. While 8 bins delivered highest FPS, it suffered more failures. The 32-bin configuration proved least effective due to oversensitivity to appearance changes. This shows mean-shift tracking benefits from moderate histogram resolution.

Our analysis of learning rates (α) (table V) reveals a key trade-off: smaller values maintain model stability with fewer failures, while larger values increase speed but risk overfitting to noise. The intermediate $\alpha = 0.01$ provides optimal balance, offering sufficient adaptation without compromising tracking reliability. This demonstrates mean-shift's inherent stability-speed tradeoff in dynamic environments.

Table IV
PERFORMANCE COMPARISON ACROSS DIFFERENT HISTOGRAM BIN COUNTS (FIXED: KERNEL=15, α=0.010, ϵ=1.0E-07) SHOWING TOTAL FAILURES AND AVERAGE FPS ON THE VOT2014 BENCHMARK.

Number of bins	Total Failures	Avg. FPS
8	49	536
16	42	464
32	59	283

Table V
PERFORMANCE COMPARISON ACROSS DIFFERENT LEARNING RATES (α) (FIXED: KERNEL=15, BINS=16, ϵ=1.0E-07) SHOWING TOTAL FAILURES AND AVERAGE FPS ON THE VOT2014 BENCHMARK.

α value	Total Failures	Avg. FPS
0.001	44	440
0.01	42	461
0.1	84	526

The convergence threshold ϵ significantly impacts tracking performance, as seen in table VI. Smaller values maintain accuracy through precise localization, while $\epsilon = 1$ fails catastrophically by terminating iterations prematurely. Despite similar FPS across values, proper convergence precision proves critical for reliable tracking.

Table VI
PERFORMANCE COMPARISON ACROSS DIFFERENT CONVERGENCE THRESHOLDS (ϵ) (FIXED: KERNEL=15, BINS=16, α=0.010) SHOWING TOTAL FAILURES AND AVERAGE FPS ON THE VOT2014 BENCHMARK.

ϵ value	Total Failures	Avg. FPS
$1e-8$	42	470
$1e-4$	42	453
1	341	432

III. CONCLUSION

Our analysis reveals mean-shift tracking offers efficient performance but faces limitations with rapid motion and occlusions. Optimal results require careful parameter tuning, particularly for kernel size and adaptation rate. While effective for stable scenarios, the method could benefit from combining with motion estimation techniques for challenging cases.

REFERENCES

[1] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[2] M. Kristan *et al.*, "The visual object tracking (VOT) challenge 2014," https://www.votchallenge.net/vot2014/, 2014.