# Visual Tracking via Locality Sensitive Histograms

Yufeng Jiang

## 1. Multi-Region Tracking

The purpose of multi-region tracking algorithms is to capture the spatial information of a target object to account for appearance change. The target object is missing in single region tracking. Though using multiple regions to represent a target object is important, it is not useful to use a large number of them because the cost of local analysis and region-by-region matching is extremely high. Authors exploit the proposed locality sensitive histograms for multi-region tracking based on the efficientive illumination invariant features and region matching scores. One application of their method is to adapt the fragement-based method [1] to demonstrate the effect of using a large number of regions for robust object tracking.

### 1.1. Online Template Update

Visual tracking with a fixed template is not effective iver a long period of time as the appearance have changed sigificantly. To solve this problems, authors update the region histograms of the template. The most biggest advantage of using multiple regions is that allowing the template to adapt to the appearance change and alleviating the tracking drift problem during updating a fraction in each frame. Once the new target location is determined in [4], the local historgrams are updated as follows:

$$H^E_{p1}(\cdot) = H^E_{p2}(\cdot) \quad \text{if } F_1 \cdot M < d(S_1 \cdot S_2) < F_2 \cdot M, \tag{1}$$

where $M$ is the medium distance of all the region at the new position. $F_1$ and $F_2$ are the forgetting factors, which define the appearance model update rate. In general, the regions with high dissimilarity represent occlusion, and the region with low dissimilarity are perfectly matched. Thus, authors draw a conclusion that it is better to update only the regions with medium dissimilarity.

## 2. Experiments

In this part, auhtors evaluate the effectiveness and efficiency of the proposed tracking method. The working environment of this experiments which authors used in C with single core is a PC with an Intel i5 3.3GHz CPU and 8GB RAM. They have evaluated it using 20 video sequences which contain challlenging factors, including deastic illumination changes, pose and scale variation, heavy occlusion, background clutter and fast motion.

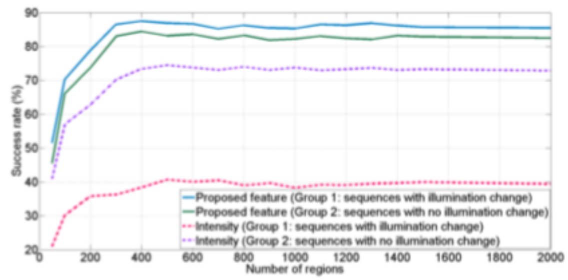


Figure 1. Success rate (%) of the proposed tracker with respect to different numbers of regions.

### 2.1. Quantitative Evaluation

There are two evaluation criteria used in their experiments: center location error and tracking success rate. Both of them are computed against manually labeled ground truth. Let $\frac{area(B_T \cap B_G)}{area(B_T \cup B_G)}$ denote the overlap ratio, where $B_T$ and $B_G$ are the bounding boxes of the tracker and the ground-truth. When the overlap ratio is larger than 0.5, the tracking result of the current frame is considered as a success.

Fig. 1 which firstly used in [4] shows the tracking performance of their method with different numbers of regions. The curves shows the average success rates of 20 video sequences. Tab. 1 researched from [4] shows the tracking performance and the speed of their method with the 12 other methods. After ananlysising the date, they note that the **TLD** tracker does not report tracking result when the drift problem occurs. So, they only report the center location errors for sequences that **TLD** does not lose the track of target objects. Fig. 2 cited from [4] shows some tracking results of different trackers. Authors only show the results by trackers with the top average success rate, *i.e.* **CT** [7], **MIL** [2], **TLD** [5], **Struck** [3], **DFT** [6] and **MTT** [8] to clear the presentation.

In this paper, authors propose a novel locality sensitive histogram method and a simple but effective tracking frame-

| Sequence | Ours | LIT | SPT | CT | Frag | MIL | Struck | VTD | TLD | BHT | LGT | DFT | MTT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Basketball | 87 | 75 | 84 | 32 | 78 | 27 | 2 | 96 | 1 | 20 | 44 | 3 | 3 |
| Biker | 69 | 23 | 44 | 34 | 14 | 40 | 49 | 45 | 38 | 46 | 7 | 46 | 44 |
| Bird | 98 | 44 | 74 | 53 | 48 | 58 | 48 | 13 | 12 | 71 | 5 | 91 | 13 |
| Board | 93 | 3 | 47 | 73 | 82 | 76 | 71 | 81 | 16 | 38 | 5 | 23 | 63 |
| Bolt | 81 | 18 | 8 | 66 | 15 | 73 | 76 | 26 | 3 | 6 | 2 | 8 | 68 |
| Box | 84 | 4 | 8 | 33 | 42 | 18 | 90 | 34 | 60 | 8 | 9 | 37 | 25 |
| Car | 92 | 43 | 73 | 43 | 40 | 38 | 59 | 44 | 58 | 10 | 11 | 43 | 49 |
| Coupon | 100 | 24 | 98 | 58 | 67 | 77 | 100 | 38 | 98 | 58 | 12 | 100 | 100 |
| Crowds | 76 | 59 | 7 | 9 | 2 | 4 | 82 | 8 | 16 | 4 | 3 | 85 | 9 |
| David indoor | 93 | 41 | 64 | 46 | 35 | 24 | 67 | 32 | 90 | 7 | 24 | 45 | 92 |
| Dragon baby | 67 | 16 | 28 | 30 | 35 | 38 | 43 | 33 | 15 | 28 | 4 | 23 | 24 |
| Man | 100 | 98 | 41 | 60 | 21 | 21 | 100 | 31 | 98 | 18 | 8 | 21 | 100 |
| Motor rolling | 83 | 5 | 28 | 11 | 24 | 9 | 11 | 6 | 14 | 30 | 1 | 10 | 5 |
| Occluded face 2 | 100 | 60 | 22 | 100 | 80 | 94 | 79 | 77 | 76 | 42 | 8 | 49 | 82 |
| Shaking | 73 | 12 | 3 | 81 | 8 | 45 | 9 | 77 | 1 | 2 | 23 | 95 | 2 |
| Surfer | 75 | 1 | 3 | 3 | 2 | 2 | 67 | 2 | 86 | 2 | 1 | 3 | 3 |
| Sylvester | 89 | 48 | 34 | 74 | 66 | 70 | 87 | 72 | 76 | 78 | 8 | 54 | 96 |
| Tiger2 | 66 | 10 | 3 | 65 | 5 | 77 | 65 | 17 | 26 | 5 | 2 | 21 | 27 |
| Trellis | 91 | 67 | 72 | 35 | 18 | 34 | 70 | 54 | 31 | 18 | 2 | 45 | 34 |
| Woman | 83 | 8 | 80 | 6 | 71 | 6 | 87 | 5 | 30 | 34 | 4 | 80 | 8 |
| **Average** | 85.0 | 33.0 | 41.1 | 45.6 | 37.7 | 41.6 | 63.1 | 39.6 | 42.3 | 26.3 | 9.2 | 44.2 | 42.4 |
| **Average FPS** | 25.6 | 10.2 | 0.2 | 34.8 | 3.5 | 11.3 | 13.5 | 0.5 | 9.5 | 3.3 | 2.8 | 5.8 | 1.0 |

Table 1. The success rates (%) and the average frames per second (FPS) of the 20 sequences. The best and the second best performing methods are shown in red color and blue color, respectively. The total number of frames is 10,918.

Figure 2. Screenshots of the visual tracking results. The figures are placed in the same order as Tab. 1. Red, green, yellow, azure, purple, olive and blue rectangles correspond to results from **DFT**, **CT**, **MTT**, **MIL**, **Struck**, **TLD** and **ours**.

work. They can conclude from experimental results that the proposed multi-region tracking algorithm has a favorable performance compared to numerous state-of-the-art algorithms.

## References


[1] A. Adam, E. Rivlin, and I. Shimshoni. Robust fragments-based tracking using the integral histogram. In *CVPR*, 2006. 1

[2] B. Babenko, M.-H. Yang, and S. Belongie. Robust object tracking with online multiple instance learning. *IEEE TPAMI*, 33(8):1619–1632, 2011. 1

[3] S. Hare, A. Saffari, and P. Torr. Struck: Structured output tracking with kernels. In *ICCV*, 2011. 1

[4] S. He, Q. Yang, R. W. H. Lau, and J. Wang. Visual tracking via locality sensitive histograms. In *CVPR*, 2013. 1

[5] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-learning-detection. *IEEE TPAMI*, 34(7):1409–1422, 2012. 1

[6] L. Sevilla-Lara and E. Learned-Miller. Distribution fields for tracking. In *CVPR*, 2012. 1

[7] K. Zhang, L. Zhang, and M.-H. Yang. Real-time compressive tracking. In *ECCV*, 2012. 1

[8] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja. Robust visual tracking via multi-task sparse learning. In *CVPR*, 2012. 1