An improved TLD target tracking algorithm based on Mean Shift

Aijun Zhou[1], Yu Liufang[2]

[1,2]Research Scholar,Information Engineering College,Nanjing Noamal University Taizhou College
Taizhou, 225300, China

Abstract **–Tracking-Learning-Detection (TLD) is an excellent long-term tracking method which has the advantages of high accuracy of tracking rate and self-detection mechanism. Noting that TLD algorithm is sensitive to illumination change and clutter results in drift even missing, and the corresponding tracker designed based on the pyramid Lucaks-Kanade optical flow method needs vast computation. To overcome these shortcomings, an improved target tracking scheme by integrating mean-shift and TLD algorithm is proposed.** ***When the confidence level of the TLD tracking box is high, the center position of target of the TLD output is used as the starting point of the Mean Shift tracking algorithm. When the confidence level is low, the center position of the target box in the previous frame is used as the iterative starting point for the Mean Shift. The results show that the improved algorithm achieves higher precision, especially for occlusion and target jitter. In order to solve the problem that there are more useless points in the feature points obtained by uniform sampling of TLD algorithm, a more robust Susan corner point is introduced into the TLD tracking module. This algorithm can track the object through the pyramid LK optical flow after selecting the corner. It not only preserves feature points with rich information during the tracking process, but also suppresses the tracking drift caused by more useless points. The results show that this method has high robustness and real - time compared with the original TLD algorithm.***

Keywords *– TLD algorithm, confidence level, target tracking, Mean Shift, Susan corner point*.

I. INTRODUCTION

Detecting and tracking of moving targets, as an important foundation of computer vision, have been widely applied in various fields, such as intelligent monitoring, unmanned aerial vehicles, mobile robots, man-machine interaction,navigation, guidance, etc. Although a number of methods on target tracking have been presented, some traditional problems still remain unaddressed, such as change of light, pose and scale change, fast moving of targets, occlusions between targets, appearing of similar targets. Jia, et al. [5] enhanced the Nearest Neighbor (NN) classifier via Local Binary Pattern (LBP) algorithm.The issue of target tracking by fusing kalman filter revealed that the proposed method can properly detect and accurately track a target in complex situations[6]. Given that the realization of Kalman filter must meet the follow three important conditions: (1) The system for modeling is a linear system; (2) The predicted or measured noise should be white noise; (3) The distribution of noise should be a Gauss distribution. However, due to the existence of occlusion, shadow, scale changes and other problems in the process of target tracking, the traditional target tracking algorithm can only deal with specific problems, and only in the short-term tracking to achieve good results. TLD[7,8] is an efficient long-term tracking algorithm which proposed by Kalal in the Ph.D. period. This algorithm combines the traditional tracking algorithm and detection algorithm to solve the problem that encounter partial occlusion and deformation in the process of tracking. When the tracker fails to track, the target can be detected again by the detector, allowing the tracker to recover from the error.

This paper is organized as follows. In section II, preliminary work including the TLD algorithm and Susan corner point algorithm will be introduced briefly. In section III, an improved algorithm for TLD target tracking based on Mean Shift is discussed in detail. In Section IV, experimental results are provided to validate the performance of the proposed approach. Finally, the conclusion will be presented.

II. METHODOLOGY

A. TLD algorithm

TLD is a target tracking algorithm that combines tracking and detection. And the algorithm refers to the online learning mechanism, making the tracking process in the case of occlusion and deformation more robust. TLD mainly includes tracking module, detection module and learning module.

- Tracking module

TLD tracking module use the algorithm for adding error detection on the basis of Median Tracker. This algorithm requires to use the LK optical flow method twice and adds the forward-backward error detection function. However, feature points will not just return to the original position after twice optical flow, there is a certain deviation. Assuming that the deviation of the twice optical flows is less than the threshold we set, the tracking is judged to be successful. The difference between the twice optical flows can be represented as

$$d = | p - p'' | \tag{1}$$

Where p'' is the point after twice optical flows.

The tracking module of TLD algorithm uses forward-backward error to calculate the tracking error on the basis of pyramid LK optical flow method, and filter out the sample points with large relative error. And finally through a reliable feature points to calculate the current target of the new location. The implementation process of this algorithm is shown in Fig.1.

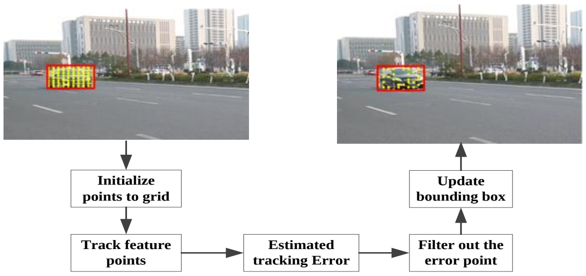


Fig. 1 Implementation process of the median flow tracking algorithm

- Detection module

The TLD detection module consists of a patch-variance classifier, an ensemble classifier, and a 1-NN classifier. It is through the sliding window to scan the entire image. When the target is lost during the tracking process, the tracker cannot recover from the error. In this case, the target is detected by the detector to find the target that is missing. Each detection window is only accepted by three classifier, it will become the target window.

- Learning module

The TLD learning module uses the semi-supervised learning method in machine learning, uses the P-N learning method to estimate each frame in the detector, uses the P expert to find the wrong negative sample, and the N expert finds the wrong positive sample. The results are fed back to the training set to update the classifier.

B. Susan corner point algorithm

Susan corner point algorithm is through a circular template to traverse the image, select the center of the circular template as the core point. The template pixel gray value and the core point gray value are comparison, expressed as

$$C(x_0, y_0) = \begin{cases} 1, | I(x, y) - I(x_0, y_0)| \le t \\ 0, | I(x, y) - I(x_0, y_0)| > t \end{cases} \quad (2)$$

The regions of the template that meet the criteria for the pixels form USAN. In general, a 37-pixel template is selected.
The size of the USAN area reflects the strength of the local features of the image. Its area is smaller, the core point is the greater possibility become corner. Calculating the area of the USAN area at that pixel. The function can be defined as

$$n(x_0, y_0) = \sum_{i=1}^{i=N} C(x_i, y_i) \quad (3)$$

Where $n(x_0, y_0)$ indicates the area of the USAN region.

By comparing the USAN value and the geometric threshold *g* of each pixel to obtain the initial response function of the current pixel, the calculation method can be expressed as

$$R(x_0, y_0) = \begin{cases} g - n(x_0, y_0), n(x_0, y_0) < g \\ 0 \quad ,\text{otherwise} \end{cases} \quad (4)$$

The geometric threshold *g* in (4) can be dynamically adjusted according to the actual needs. In corner detection, *g* = 0.5*Amax*, *Amax* represents the maximum area of USAN, depending on the template size. The pixel that satisfies the R response is listed as the candidate corner, and then the non-maximum value of the candidate corner to determine the corner point.

III. AN IMPROVED ALGORITHM FOR TLD TARGET TRACKING BASED ON MEAN SHIFT

A. Mean Shift algorithm principle

Mean Shift is a fast pattern matching algorithm based on probability density estimation. The color feature histogram is used to describe the target, and the similarity measure is made in each frame image, and finally realizing the pattern matching and target tracking. Assuming that give *n* sample points $x_i, i = 1, 2, ..., n$ in a d-dimensional space R^d . The basic form of the Mean Shift vector at x is defined as

$$\mathrm{M}_h = \frac{1}{k} \sum_{x_i = S_h} (x_i - x) \quad (5)$$

Where *k* is the number of all sample points falling into the S_h . S_h is a high-dimensional sphere area with a radius *h*, a set of *y* points that satisfy the following relationship.

$$S_h = \{ y \mid (y - x)^T (y - x) \le h^2 \} \quad (6)$$

Considering the distance between the sample and the offset point, its contribution to Mean Shift is different. So the kernel function is introduced and the different weights are given according to the different positions of each sample point. The Mean Shift vector is expressed as

$$M(x) == \frac{\sum_{i=1}^{n} G_H(x_i - x) w(x_i) x_i}{\sum_{i=1}^{n} G_H(x_i - x) w(x_i)} - x = m_h(x) - x \quad (7)$$

Where $G(x)$ is a kernel function, w_i is the weight of the sample point.

B. TLD tracking algorithm based on Mean Shift

First, the Mean Shift algorithm is used for modelling and parameter selection. The probability density estimate of the target template can be expressed as

$$\hat{q}_u = C\sum_{i=1}^{N} K(\| x_i^* \|^2)\delta[b(x_i)-u] \tag{8}$$

Where u is the colour index corresponding to the histogram. x_i^* indicates the target location normalized to the target center. C is normalized constant.

The search center coordinates y of the current frame are determined based on the tracking result of the previous frame. Each pixel in the region centered on y is represented by $\{x_i, i=1,2,...,n\}$, then the probability density estimate of the candidate template can be expressed as

$$p_u(y) = C\sum_{i=1}^{n} K(\| \frac{y-x_i}{h} \|^2)\delta[b(x_i)-u] \tag{9}$$

The Bhattacharyya (*BH*) coefficient is calculated by the (10), and $\hat{\rho}(y)$ represents the similarity between the target template and the candidate template.

$$\hat{\rho}(y) = [\hat{p}(y), \hat{q}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u \hat{q}_u} \tag{10}$$

In the process of iterative search for candidate targets, the commonly used offset is less than the threshold, the actual use of the number of iterations to control the search, the general choice of the number of iterations is 20 times. The bandwidth with the largest *BH* coefficient is taken as the optimal bandwidth. To avoid over-regulation, the bandwidth in the current frame is corrected. χ is an adjustment parameter, this paper selected 0.1 by testing.

$$h_{new} = \chi h + (1-\chi)h_{opt} \tag{11}$$

The steps of the TLD tracking algorithm based on Mean Shift are as follows.

Step 1: Read the video frame, use the mouse to manually mark the target, and then initialize the TLD and Mean Shift tracker.

Step 2: Start the TLD and Mean Shift trackers, respectively. Through the TLD tracker to predict the location of the target to determine its confidence. If the confidence level is high, the central position of the output of TLD as a Mean Shift tracking algorithm iteration starting point. If the confidence level is low, the center position of the target box in the previous frame of Mean Shift as the starting point for the tracking algorithm.

Step 3: Set the maximum number of iterations is 20, and in each iterative search, the search center position is shifted toward the center of the similarity.

Step 4: Adjust the radius of the kernel function of Mean Shift dynamically to make it adaptive scale changes. The final candidate target scale is obtained by using the radius of the kernel function corresponding to the candidate target with the largest similarity.

Step 5: Set the target box of the Mean Shift output as the final tracking prediction box.

C. *Algorithm improvement*

Many of the 100 sample points that are uniform sampling in the tracking module of the original TLD algorithm will become useless sample points, which will not only cause the drift of the tracker, but will also cause a lot of computation. Corner is a feature point with strong characteristics, including the gradient of the image changes or local grayscale changes. In this paper, we consider that the feature points generated by uniform sampling cannot accurately describe the tracking target. The Susan algorithm and the TLD algorithm are combined to replace the original uniform sampling points with the corners obtained by the Susan algorithm. Choosing a stronger Susan corner point as a tracking point to improve the robustness of tracking. The schematic diagram of the tracking module after the improved algorithm is shown in Fig.2.

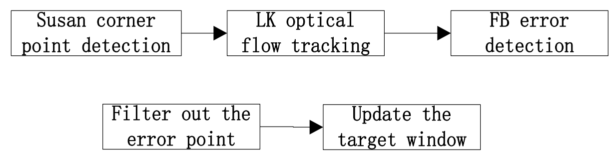


Fig. 2 Schematic diagram of the improved tracking module

IV. EXPERIMENTAL RESULTS AND ANALYSIS

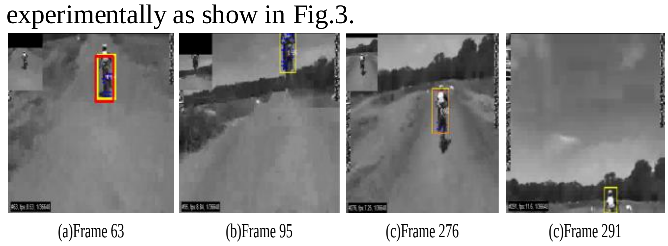
A. *Evaluation criterion*

For the stability of tracking, this paper uses the tracking success rate as a measure. The tracking success rate is expressed as the number of frames tracked to the target divided by the total number of frames in the target. For real-time of tracking, this paper uses the average value of the tracking frame rate as a measure. The tracking frame rate is expressed as the total number of frames divided by the total time.

B. *Results and discussion*

(1) Tracking stability

We experiment with the online published Motocross, David, Car and the video output of our camera to test the stability of our algorithm. The video sequence of the motocross is tested experimentally as show in Fig.3.



(a)Frame 63 (b)Frame 95 (c)Frame 276 (c)Frame 291

Fig. 3 Tracking results of Motocross video sequence

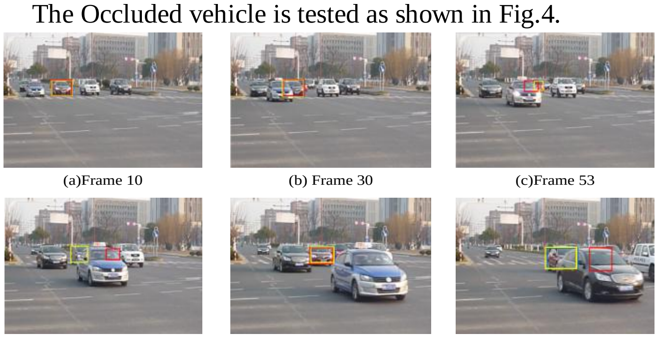
From Fig.3, we can see that the image in the frame 95 and 291, respectively, have a large scale changes and occlusion, the tracking target of the original TLD algorithm get lost. Our algorithm has strong anti-blocking ability, and the improved algorithm based on Susan corner point is effective to solve the scale change, so it can accurately track the target. Table I is a repeated experiment for TLD algorithm and our algorithm, and finally taking the statistical results of the average.

Table I. Tracking results comparison table

Video sequence	The original TLD tracked frames	Our TLD tracked frames	TLD tracking success rate	Our TLD tracking success rate
Motocross	1092	1357	0.57	0.71
David	724	748	0.95	0.98
Car	851	902	0.92	0.97

From Table I, we can see that the improved algorithm can effectively deal with the problem of occlusion and scale change. Therefore, the tracking success rate of our algorithm is further improved compared with the original TLD algorithm.

The Occluded vehicle is tested as shown in Fig.4.



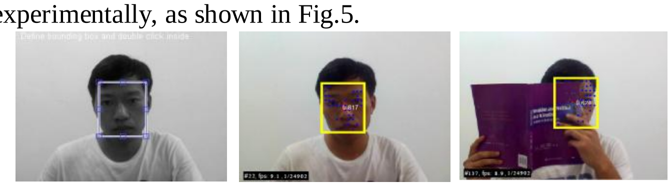
(a)Frame 10 (b) Frame 30 (c)Frame 53

(d)Frame 71 (e)Frame 92 (f)Frame 134

Fig. 4 Tracking results of occluded vehicles

As shown in Fig.4, the red box marks the original TLD algorithm, yellow box marks our algorithm. In the frame 10, the yellow box and the red box are initialized to select the red car. After the occlusion between frames 30 and 53, the original TLD algorithm appears tracking errors. After target reappears in frame 71, the original TLD algorithm failed to regain the target. And the improved algorithm is able to track the target in the face of occlusion. Experiments show that our algorithm has strong anti-blocking ability.

The video sequence of the camera output is tested experimentally, as shown in Fig.5.



(a) Manual calibration tracking target (b) Target tracking (c) target occlusion

Fig. 5 The occlusion test of our algorithm on the camera

Fig.5 (a) is to track the target by manual marking, Fig.5 (c) shows the effect of tracking when the tracking process encounters severe occlusion. Experiments show that our algorithm has strong anti-blocking ability.

(2) Real-time characteristic of tracking

In this paper, we use the average of the tracking frame rate as a measure of the tracking real-time comparison, through the video on the Motocross, David, Car video sequence and the camera output video sequence as experimental samples. The average frame rate of each video sequence under different algorithms is shown in Table II.

Table II. Comparison of the frame rate

Video sequence	Original TLD/fps	TLD based on Mean Shift /fps	Our TLD /fps	Our TLD growth rate/%
Motocross	9.21	9.17	11.13	20.8
David	8.45	8.26	10.21	16.1
Car	10.02	9.91	11.87	18.4
Camera	7.42	7.25	9.07	22.2

From Table II, we can see that: The frame rate of the original TLD algorithm and the TLD algorithm based on Mean Shift is not much different. In this paper, due to the introduction of the Susan algorithm in tracking module greatly reducing the amount of computation, so real-time significantly improved. The real-time have improved about 20%.

V. CONCLUSION

The contributions of this work can be concluded as below. (1) A TLD target tracking algorithm based on Mean Shift is proposed. (2) In the TLD tracking module introduced the Susan corner algorithm, reducing the amount of computation. (3) The algorithm has proved to have strong anti-blocking ability. And real-time about 20% increase.

REFERENCES

[1] B Babenko, M H Yang, S Belongie, Visual tracking with online multiple instance learning, IEEE Trans Pattern Anal Mach Intell,Vo1.33,No.8, 983-990, 2009.

[2] D S Bolme, J R Beveridge, B A Draper, et al, Visual object tracking using adaptive correlation filters, IEEE Conference on Computer Vision & Pattern Recognition,Vol.119, No. 5, 2544-2550, 2010.

[3] J Sanmer, C Leisfner, A Saffari, et al, PROST: Parallel robust online simple tracking, IEEE Conference on Computer Vision & Pattern Recognition, Vo1.23, No.3, 723-730, 2010.

[4] S Hare, A Saffari, P H S Torr Struok: Structured outputtracking with kernels, International Conference on Computer Vision, Vo1.23, No.5. 263-270, 2011.J

[5] Z Kalal, K Mikolajczyk, J Matas, Tracking-Learning-Detection,IEEE Transactions on Pattern llnalysis & Machine Intelligence, Vo1.34,No.7,1409-1422, 2012.

[6] B D Lucas, T Kanade, An iterative image registration technique with an application to stereo vision, International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers Inc, 285-289,1981

[7] C Jia, Z Wang, X Wu, et al, A Tracking-Learning-Detection (TLD) method with local binary pattein improved, IEEE InternationalConference on Robotics and Biomimetics，1625-1630, 2015

[8].. C Sun, S Zhu, J Liu. Fusing. Kalman filter with TLD algorithm fox target tracking, Chinese Control Conference, 3736-3741, 2015.