An Improved TLD Target Tracking Algorithm Combined with Kernel Correlation Filtering

Haoyu Duan1,*, Yumin Zhang1,*, Wei Sheng1,*, Xin Chen1,*, and Jianxin Ren2,*

1School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100083, China
2Beijing Aerospace Guanghua Electronic Technology Co. Ltd, Beijing 100854, China

*Corresponding author e-mail: zhyminus@163.com, *dhy0719@163.com,
bshengwei@buaa.edu.cn, *bournex@163.com, *buaarjx612@buaa.edu.cn

Abstract. Target continuous tracking problem is an important research topic in the field of computer vision. To conquer difficulties of target loss under complex situation, size changes and illumination changes, an improved tracking module is designed via combining tracking module in TLD with KCF (Kernelized Correlation Filters). The comparison between TLD and the proposed algorithm shows that the performance of the latter is improved in tracking accuracy. Under complex situation of target occlusion, scale change, illumination change, etc., algorithm presented can output the tracking results stably, which is more robust and more suitable for long-term target tracking.

1. Introduction
Target tracking has always been an important topic in the field of computer vision. To improve the robustness, in 2012, the TLD algorithm proposed by Kalal [1] et al. combines the tracking module with the detection module. When the tracking module has an error, the detection module is used to correct the result, and the new feature of the target is input into the learning module for learning, so that the target model is obtained. In view of the limitations of the TLD median stream tracker, this paper proposes a target tracking algorithm that combines the TLD tracking module with KCF. The KCF algorithm is used to supervise the TLD tracking in real time, and the similarity is calculated by the tracking result. The judgment of the detection module switching is performed, and the target tracking frame is adjusted in combination with the two tracking results. The results show that our algorithm can output the tracking results stably and accurately when encountering complex conditions such as target occlusion and illumination changes, which is robust and suitable for long-term single-target tracking.

2. Tracking algorithm of TLD and KCF
The TLD tracking algorithm uses the median stream algorithm as a base tracker. By comparing the similarity of results between the KCF algorithm and the median stream algorithm, the KCF algorithm is combined into the TLD algorithm to perform long-term comparison and supervision on the original TLD, which can help to be more robust by overcoming some shorts of the MF algorithm. At the same time, the high speed of KCF does not affect the overall speed of the algorithm [2], while avoiding the tracking drift caused by error accumulation when using KCF alone.
2.1. Tracking Box Similarity between KCF and MF Method

The similarity calculation between the tracking results of KCF and MF algorithm can be described as the similarity comparison of the target tracking frames calculated by the two algorithms on the same frame image [3]. In this paper, we use overlap rate to describe the similarity: the tracking boxes $A_{KCF}$ and $A_{MF}$ of the current frame are respectively calculated by the KCF and the MF tracking algorithm. Then the area ratio of the intersection of the two and the union is recorded as the similarity $M$ of the current frame:

$$M = \frac{(A_{KCF} \cap A_{MF})}{(A_{KCF} \cup A_{MF})}$$  \hspace{1cm} (1)

It can be intuitively seen that the larger the $M$ value, the closer the KCF and TLD prediction results are, and the higher the reliability of the tracking result is. When judging the magnitude of the similarity $M$, if $M$ is greater than the set threshold, it is considered that the tracking frame overlap is sufficiently high to continue tracking; if $M$ is less than the threshold, the overlap rate is so low that the detector begins to work and the learner no longer learns the untrusted result of this time [4].

In the same frame, the area of the KCF tracking box is represented by $S_{KCF} = L_1 \times W_1$; the TLD median flow algorithm $S_{MF} = L_2 \times W_2$, the area of the overlapping parts and the union of the two tracking frames represented as $S_{overlap}$ and $S_{merge}$, then the similarity $M$ can be expressed as:

$$M = \frac{A_{KCF} \cap A_{MF}}{A_{KCF} \cup A_{MF}} = \frac{S_{overlap}}{S_{merge}} = \frac{L_{overlap} \times W_{overlap}}{S_{KCF} + S_{MF} - S_{overlap}}$$  \hspace{1cm} (2)

2.2. Feature Point Screening and Tracking Box Adjustment

2.2.1. Feature Point Screening. According to the above ideas, we define that the tracking points in the common area of the tracking boxes of the two algorithms are trusted tracking points that conform to both algorithms, and the feature points in the tracking boxes but not in the public area are secondary trusted tracking points. These trusted points are selected as feature tracking points and used as the tracking basis for the next tracking.

As shown in Fig. 1, the 181st frame of the Car Scale in the OBT-2013 data set, the red dot is the feature point. (a) Shows the feature points selected by TLD tracking algorithm, and (b) for the feature points selected by the algorithm combined with KCF and TLD. It can be seen that there are some feature points mismatching in the former, and the ones generated by the latter are almost all distributed on the target with higher accuracy.

2.2.2. Tracking Box Adjustment Strategy. We have noticed that if the above scheme is used to filter the trusted tracking points and calculate the tracking box size, the box will gradually become smaller than the actual target size due to the continuous reduction of the feature points. To solve the problem of the tracking frame narrowing caused by the continuous reduction of trusted tracking points, in this paper, we use the secondary trusted tracking points mentioned above, centering on the rectangular box calculated by the trusted tracking points, and expanding the box with half of the trusted ones (N/2 or N/2+1) as the standard, and use it as the tracking box of the improved algorithm.
Figure 1. Tracking points of TLD algorithm and filtered trusted tracking points of ours.

Figure 2. Results before and after tracking box adjustment.

As shown in Fig. 2, the blue dotted line frame is the exact target location calibrated by Ground truth, the improved algorithm tracking box is closest to the exact target.

3. Experiments and analysis
In order to verify the effectiveness of the improved algorithm (hereinafter referred to as ours), we conducted a comparative experiment using the standard data Qualitative analysis set OTB2013. The six tracking algorithms used in the experiment, in addition to the improved algorithm in this paper, are the more popular tracking algorithms, namely: TLD, KCF, DCF, MOSSE filter and DSST. All comparison algorithms use default parameter values when executed. Experiments are implemented in Visual Studio 2015 and Matlab 2014a, and the experimental environment is: Intel(R) Core-i7 processor, clocked at 2.6GHz, 8GB memory, 64-bit Windows 10 operating system.

We will analyze the superiority of the improved algorithm from qualitative and quantitative aspects.

3.1. Qualitative analysis
In order to analyze the algorithm more qualitatively, we select two representative video sequences with one or several challenges including target fast motion, light change, size change and occlusion, as shown in Table 1.

| Test sequences | Total frames | Fast motion | Light change | Size change | Similar background | Occlusion |
|----------------|--------------|-------------|--------------|-------------|--------------------|----------|
| Car4           | 659          | no          | yes          | yes         | yes                | no       |
| Dog1           | 1350         | yes         | no           | yes         | no                 | no       |

Table 1. Description of experimental test sequences.
Figure 3. Qualitative comparison of different tracking algorithms.

Figure 3 is partial screenshots of the experimental results of video sequences. It can be seen that the motion of the target in the video is challenging for the tracking algorithm. In the video sequences Car4, there is obvious light changes. Except for DSST and ours, other algorithms eventually lose the target. In the two tested video sequences, the target sizes varied to varying degrees. Taking the 80th frame to the 1022th of Dog1 as an example, it can be seen from Fig. 3(b) that while the target scale changing from small to large, our algorithm adapts well to the size change of the target, and shows higher tracking accuracy since it has improved the tracking box adjustment scheme.

3.2. Qualitative analysis

When quantitatively evaluating the robustness of the tracking algorithm, two evaluation methods, precision plot and success rate plot, are used [5-6]. The former is defined as the ratio of the number of video frames whose center position error of all tracking frames is less than the given threshold to the total number of frames, and is used to evaluate the overall performance of the tracking algorithm. The latter success plot describes the ratio of the number of video frames with a degree of overlap greater than the given threshold to the total number of frames.

The plots of tracking accuracy and success rate of the six algorithms are shown in Fig. 4 and Fig. 5. As can be seen, the proposed algorithm has higher tracking accuracy and stability under different video sequences than other classical tracking algorithms.
4. Conclusion

The traditional TLD algorithm is prone to the defects of error accumulation leading to target loss in complex situations such as loss of target, rapid movement and light changes. In this paper, the tracking module is designed based on comparison of KCF and MF similarity. The improved TLD tracking algorithm supervises the median flow tracking through KCF, and uses the overlap degree of the two tracking boxes to judge the tracking similarity of the two algorithms, and so that judging the output credibility of the TLD tracking algorithm. The experimental results show that the proposed algorithm has higher tracking success rate than the original TLD algorithm. However, the information fusion adopted by the improved algorithm increases the computational cost to a certain extent. Therefore, the next research is to further optimize the algorithm while ensuring the tracking accuracy, reduce the complexity of the algorithm, and improve the detection efficiency.
Acknowledgments
Authors say many thanks to reviewers for their hard working and kindly comments. This work is supported by national science foundation of China with granted No.61374131.

References
[1] Kalal Z, Mikolajczyk K and Matas J, Tracking-learning-detection, IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34 (7): 1409 - 1422.
[2] W. J. Dai, T. Q. Chang and K. F. Su, Improved TLD Target Algorithm Based On Feature Fusion, Proc of Workshop on Advanced Research & Technology in Industry Application, 2016: 398 - 403.
[3] J. Q. Xu and Y. Lu, Robust visual tracking via weighted spatio-temporal context learning, Acta Automatica Sinica. 2015, 41 (11): 1901 - 1912.
[4] Q. Sun, S. X. Zhang and Z. X. Zhang, A weighted-overlap based metric for single visual object tracking evaluation. Acta Automatica Sinica, 2017, 45 (3): 753 - 761.
[5] M. Everingham, L. Gool, C. K. I. Williams, et al. The Pascal visual object classes (VOC) challenge, International journal of computer vision, 2010, 88 (2): 303 - 338.
[6] Y. X. Wu, N. Jia, J. P. Sun, Real-time multi-scale tracking based on compressive sensing, The visual computer, 2015, 31 (4): 471 - 484.