Robust Object Tracking Based on Complementary Feature Fusion and Channel Reliability

Junchang Zhang, Bochuan Zhang and Jie Gao

ABSTRACT

In this letter, a robust object tracking method is proposed, which is divided into two parts: robust tracker based on complementary feature fusion with channel reliability, and reliable validator based on convolutional neural network. Compared with the existing tracking algorithms, our algorithm tracks the target more robustly and accurately, and suppresses template drift problems effectively. Experimental results on the standard benchmark show that our algorithm can compete with advanced algorithms and achieve a better level.

KEYWORDS
Object Tracking, Complementary Feature Fusion, Channel Reliability, Validator.

INTRODUCTION

As one of the important branch of computer vision, object tracking is applied to many fields such as video surveillance, artificial intelligence, security systems, and intelligent transportation. Scholars have proposed lots of object tracking methods, which are roughly divided into two classes: generative model and discriminative model. The former considers the candidate target that best matches the target model as the current actual target estimated, while the latter regards the object tracking as a classification problem.

In recent years, due to its simple calculation and high efficiency, tracking algorithms based on correlation filtering have attracted much attention. The main improvement methods include the introduction of kernels, the addition of scaled estimates (DSST[1]), the addition of color histogram and color name features, the addition of background information (CA-CF[2]), the reduction of boundary effects (SRDCF), and the addition of deep CNN features (Siamese RPN[3]), etc.

With the development of hardware technology, deep learning methods based on CNN are also applied in the field of object tracking, such as ECO[4]. In order to avoid online fine-tuning, a dual-input network based on similarity measure is introduced. the Siamese network is widely used for object tracking. For example, Hen Fan[5] et al combined correlation filter and Siamese network to improve the performance of the tracking system.
Appropriate feature fusion can generally improve the performance of the tracking system. For instance, Staple[6] combines the HOG feature with color histogram, which greatly improves the performance of the tracker. Our algorithm shows that fusion of complementary features (HOG, color histogram and edge feature) can lead to a significant increase in performance. In addition, the correlation filtering framework based on context-aware is widely used due to its robustness to partial occlusions. By adding background information from top to bottom and left to right as negative samples, the framework increases the diversity of samples. Once the target is occluded heavily or the target moves quickly, the tracker is likely to lose the target. To overcome this difficulty, in this paper, we add a validator branch based on the CNN to increase the robustness of the tracking system.

We propose a robust object tracking algorithm which has three contributions. (i) The algorithm is consisted of two parts: robust tracker with complementary feature and reliability validator which can detect and correct error for tracker. (ii) Combining multiple complementary features to represent the appearance of the target, which is gradient direction histogram features (HOG), color histogram, and edge feature of object, respectively. (iii) At the same time, we add channel reliability estimation based on the traditional HOG feature to further improve the expressive ability.

PROPOSED METHOD

ROBUST TRACKER

As the method proposed in the article[2], in every frame, we choose 4 context patches as negative samples on the top, bottom, left and right of the target, where each one is the same size as the target. As shown in Fig 1, the algorithm including two parts: tracker and validator. The tracker is responsible for fast and efficient tracking. The validator is responsible for verifying the accuracy of the tracking result, it returns a positive feedback when tracker passes the verification. Otherwise, the validator will re-detect the target.

The HOG feature is constructed by calculating and counting the gradient direction histogram of the local grid unit of the image, it maintains invariance to the optical deformation. In actual scenarios, there are usually many similar interferers around the tracked target. In this case, even though the maximum value accurately depicts the target position, multiple equivalent-form ratio results will appear.
In order to avoid the above situation, we further improve the HOG features by increasing the channel reliability adaptive weight to better improve the performance of the tracker. The specific method is to measure the response reliability of each feature channel based on the ratio between the next largest value and the maximum value in the response map. The smaller the ratio of the sub-large value to the maximum value in the response graph, the cleaner the response graph and the better the tracking effect. Then the reliability of the channel is higher. In order to avoid multiple equivalent forms, $\frac{\rho_{\text{max} 1}}{\rho_{\text{max} 2}}$ is limited to an upper 0.5. Therefore, the reliability of each individual feature channel is estimated as formula 1). Finally, the final filter response $y_i^h$ is obtained.

$$w_d^{(\text{det})} = 1 - \min\left(\frac{\rho_{\text{max} 2}}{\rho_{\text{max} 1}}, \frac{1}{2}\right)$$ (1)

The object tracking method based on color features has the advantages of rotation, translation and scale invariance. The DAT algorithm not only use the Bayesian classifier to establish model between the target and surrounding background information, but also establish the classification model between the target and the target similar area. These two models are linearly weighted and ultimately achieve the goal of suppressing target drift, which effectively overcomes similarity interference. Therefore, the DAT algorithm is used as a color model in this letter to obtain the final color response $y_i^c$.

The object detection algorithm based on the edge information can effectively utilize the edge information and propose the appropriate recommended bounding boxes, which can reduce the scope of search and speed up the tracking. The specific method is: all the bounding boxes in the set of sample bounding boxes are scored by the edge information of the image; the score size indicates the probability of containing the target in the bounding box, and finally we can obtain the edge response map of the object $y_i^e$.

In this letter, the tracking process is divided into translational tracking and scaled tracking. The translational tracking phase not only combines three independent correlation filtering models; colour model regression models and object edge detection models, but also integrates channel reliability adaptive weights for each feature channel of HOG features; then we can obtain the final position response:

$$y_i = \partial y_i^h + \beta y_i^c + \left(1 - \hat{\partial} - \beta\right) y_i^e$$ (2)

Among them, the HOG feature is sensitive to deformation, suitable for illumination; the color histogram is not affected by image rotation and translation but sensitive to illumination; the edge detection model can obtain the edge information of the image and reduce the search range. These three models can complement each other and take advantage of each other to improve the performance of the tracker. As shown in Figure 2, after position estimate, the target scale is estimated using a one-dimensional correlation filter.
Figure 2. Block diagram of multi-complement feature fusion.

The DSST framework divides tracking into two problems: translation estimation and scale estimation by using a scale filter estimate the size of the target. The DSST has 33 scales, which are more precise, accurate and flexible. In the scale estimation stage, the HOG feature is used to construct the scale filter. In the new frame, the 2D position correlation filter is used to determine the new candidate position of the target, and then the 1D-scale correlation filter is used to take the current central position as the central point to obtain candidate patches of different scales, thereby finding the most matching scale.

RELIABLE VALIDATOR

In this paper, we use a two-input Siamese network as a validator to learn a matching function through offline training. The processing flow of the algorithm verification system as following: every $T$ frame, the tracker sends the tracking result to the validator, and the matching result is calculated according to the matching function to determine the accuracy of the tracking result. If the verification passes, the tracking continues; otherwise, the target will be redetection from the location of the error frame. Then start tracking again with the new location. As shown in Figure 3.

Figure 3. The flow of algorithm verification system.

EXPERIMENTAL RESULTS

In order to evaluate the performance and efficiency of the algorithm, this paper evaluates the algorithm on the Object Tracking Benchmarks 2013 (OTB-13)
benchmark datasets through MATLAB under the configuration of Core i7 4GHz CPU and 8G RAM. The test data set contains attributes such as occlusion, illumination changes, scale changes, motion blur, fast motion, out-of-plane rotation, and deformation.

We compare our tracker with some state-of-the-art tracking trackers including the CSK, KCF, Staple, STAPLE_CA, ECO. KCF is the baseline method of Staple. The Staple algorithm improves the robustness of the tracking algorithm by combining two complementary features. The STAPLE_CA algorithm effectively solves the occlusion and boundary effects by adding background information. The ECO algorithm uses neural network and deep features. In this letter, our algorithm integrates all of these advantages and adds validation modules to get a very powerful CF-based tracker. As shown in Figure 4, the comparison of the success rate and accuracy of the tracking results of different trackers on the OTB2013.

![Success and precision plots on OTB2013.](image)

**Figure 4. Success and precision plots on OTB2013.**

![Precision plots of some challenges on OTB2013.](image)

**Figure 5. Precision plots of some challenges on OTB2013.**

**CONCLUSION**

In this paper, robust object tracking method based on complementary feature fusion and channel reliability is proposed. The algorithm use similarity learning with Siamese discriminator as a validator. Experimental evaluation on the OTB2013 benchmark has shown that our tracker not only performs better than previous CF-
based trackers, but also exhibits significantly better performance than the ECO algorithm. As shown in Figure 5, our method works well for occlusion, deformation which closes to 0.9. For background clutter and rotation, our method also has great improvement.

Corresponding author: Junchang Zhang

REFERENCES

1. Danelljan M, Hager G, Khan F S, et al. Discriminative Scale Space Tracking [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, PP (99):1-1.
2. Matthias Mueller, Neil Smith, Bernard Ghanem. Context-Aware Correlation Filter Tracking. In CVPR, 2017.
3. Bo Li, Junjie Yan, Wei Wu, Zheng Zhu, Xiaolin Hu. High Performance Visual Tracking with Siamese Region Proposal Network. In CVPR, 2018.
4. Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, Michael Felsber. ECO: Efficient Convolution Operators for Tracking. In CVPR, 2017.
5. Heng Fan and Haibin Ling. Parallel Tracking and Verifying: A Framework for Real-Time and High Accuracy Visual Tracking. In CVPR, 2017.
6. Bertinetto L, Valmadre J, Golodetz S. Staple: Complementary Learners for Real-Time Tracking. Computer Science, 2016, 38, pp. 311-323.