Illumination-Adapted Long-Term Tracking Based on Multiple Correlation Framework

Xikai Han, Gang Yu, Han Hu, Jia Liu
School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China
gangyu@hit.edu.cn

Abstract. Object tracking is a hot field in image processing, among which the kernel correlation filtering algorithm is widely applied based on the advantages of precision and speed. However, due to information decay, it has poor performance in the long-term tracking. In this paper, we develop an illumination-adapted long-term tracking method based on correlation filtering tracker. Firstly, we introduce illumination adaptive normalization and information entropy weighted feature fusion to improve the prediction accuracy. Secondly, we estimate the object scale quickly based on the binary strategy to improve the speed of the algorithm. Then, we design re-detection based on online SVM detection for target loss. Finally, we design an appropriate updating mechanism to combine the whole tracking framework with various judgment thresholds. The experiment on the data set shows that all of our modules are effective and have great improvement as compared with other correlation filter type trackers.

1. Introduction
According to the mode of model generation, visual target tracking algorithms can be divided into two types: generative and discriminant. Due to outstanding tracking effect, discriminant tracking algorithm is the main direction of current research. As the first algorithm to apply correlation filtering method to target tracking. It adopted grayscale feature and reached the speed of 669FPS. Henriques [1] proposed Kernelized Correlation Filter (KCF) algorithm on the basis of cyclic matrix and kernel technique, which led to a boom in the development of correlation filtering algorithm. Scale adaptive kernelized correlation filter tracker introduced scale estimation on the basis of KCF, which could adapt to the tracking target's scale changes. Danelljan [2] improved the results from the aspects of training samples, filter coefficients and template update, and proposed Efficient Convolution Operators (ECO). Kalal [3] proposed the Tracking-Learning-Detection (TLD) algorithm, added the detection module to re-detection the target and improve the algorithm's long-term Tracking ability.

We mainly focus on three problems: the influence of illumination or deformation, the error accumulation of long-time target tracking and the problem of occlusion or disappearance. Based on the classical tracking model, we developed position filter and scale filter respectively, and used SVM to re-detect the missing target. Combined with relevant thresholds, we set the update mechanism to construct tracking framework. It can effectively deal with the problems of illumination, deformation, occlusion and target vanishing during the long time following.
2. Construction and Geometrical Dimensions of Specimens

2.1 Correlation Filter Tracker

Discriminant algorithm can be regarded as a binary classification problem, which mainly solves the problem of distinguishing objects and backgrounds, but its computational efficiency is not high. In the spatial domain, the kernel correlation filter training model can be expressed as follows:

\[
\text{loss} = \min_{w} \|Xw - y\|_2^2 + \lambda \|w\|_2^2
\]

where, \(X\) is all samples generated by cyclic displacement, \(y\) is training label, and \(w\) is ideal filter. In addition, the training model introduces the regularization parameter \(\lambda\) to avoid overfitting. After partial derivative with respect to \(w\) is simplified, the closed solution of \(w\) is obtained as follows:

\[
w = (X^T X + \lambda I)^{-1} X^T y
\]

In order to avoid the inverse process, transform the analytic solution in the Fourier domain, and the regression equation is as following:

\[
\tilde{w} = \frac{x^\odot y}{x^\odot x + \lambda}
\]

Multiple feature channels are reserved in the framework of nuclear correlation filtering, which provides the possibility for improvement of feature fusion and scale estimation.

2.2 Re-detection

When the target is moved out of sight, obscured or other problems occur, the traditional tracker will cause the target to be lost. Therefore, target re-recognition and fuzzy criteria have important significance. Therefore, many tracking algorithms combine correlation filter tracker and recognition technology to get good long-term tracking effect. Wang [4] combines yolov3 to re-identify tracking targets for long time tracking of airborne infrared targets. They re-detected the target with a random forest filter when the long-term reliable visual tracking with UAVs was lower than the threshold for the conservative similarity of the tracker output. Yu [5] used APCE and maximum response value as additional confidence criteria, which can improve the accuracy of pedestrian recognition.

3. The proposed method

The basic idea of our algorithm combines position filter, scale filter and re-identification tracker with update mechanism based on confidence evaluation mechanism. The algorithm mainly consists of four parts: position filter \(Ra\), scale filter \(Rb\), Re-identification \(Rc\) and judgment threshold.

3.1. Position filter \(Ra\)

Under changeable illumination conditions, the acquired image has uneven illumination distribution, which interferes with the target tracking algorithm. In order to adaptively correct non-uniform illumination images, the illumination components of images are firstly extracted by reference to multiscale Gaussian function, and then the illumination threshold is calculated.

For images with uneven light distribution, we complete the correction processing based on two-dimensional gamma function to reduce the brightness of the over-illuminated position and improve the brightness of the over-illuminated position. The function expression is as follows:

\[
O(x, y) = 255 \left( \frac{F(x,y)}{255} \right)^\gamma, \quad \text{where} \quad \gamma = \left( \frac{1}{2} \right)^{\frac{1}{m}}
\]

where, \(O(x, y)\) donates as the output image after correction, \(\gamma\) represents brightness correction and \(m\) is the average brightness of each pixel of the image.

HOG feature can describe abundant edge gradient information and CN feature is related to the object contained in the image. The features are suitable for different scenes. Entropy weighting is adopted to fuse the features information to improve the robustness of our tracking algorithm.

Based on two-dimensional entropy operation, assume that the amplitude and direction of HOG feature gradient have M layer and N layer respectively, and combine them into feature array. The calculation method of entropy is as follows:
\[ H_h = -\sum P^{ij} \log p^{ij}, \quad p^{ij} = \frac{m^{ij}}{u} \]  

(5)

where \( m^{ij} \) represents the total number of pixels corresponding to the gradient feature number \((i, j)\), \( u \) donates as total number of image pixels, and \( H_h \) is the change rate of HOG feature gradient.

CN features describe feature information and introduce spatial weights \( f \) to express spatial characteristics of tonal distribution. Assuming that the tonal mean and components in the neighborhood have \( M \) layers and \( N \) layers respectively, they are combined into a two-dimensional array. Entropy calculation formula is as follows:

\[ H_c = -\sum P^{ij} \log p^{ij}, \quad p^{ij} = \frac{m^{ij}}{u} \]  

(6)

where \( m^{ij} \) represents the total number of pixels whose average color is \( i \) and the tonal component is \( j \), \( H_c \) is the rate of change of CN feature gradient.

The larger the gradient change rate, the more obvious the gradient change in the region, indicating that the gradient contains more discriminative information in this image. In this paper, the weights are determined according to their corresponding entropy, and the decision-level information fusion is carried out for the two characteristics. The normalized weights are calculated as follows:

\[ \alpha = \frac{2H_h}{(H_h + H_c)} \]  

(7)

3.2. Scale filter \( R_b \)

In the process of training the scale filter, the processing sample size obtained by frame \( T \) is set as \( M \times N \), and the target image is scaled to build the scale pyramid with the processing result of the position filter as the center. The processing formula is as follows.

\[ X_s = sM \times sN \]  

(8)

where \( s \) is the scale factor.

Based on the assumption that the target scale does not change much between successive frames, a scale pool is established for the scale estimation. The obtained scale pyramid image block is processed by bilinear interpolation to keep the size consistent with the initial processing sample. Then the HOG feature in the image is extracted, the least square classifier is trained, and the scale filter is obtained.

Dichotomy method is introduced to estimate the scale. When the accuracy of scale calculation is the same, using formula \( s_t \), can significantly reduce the computational complexity.

\[ s_t = \text{dich}(f(s), s_1, s_2) \]  

(9)

where \( f(s) \) is the maximum value of the current scale response, and \((s_1, s_2)\) is the boundary value of dichotomy. First, the initial dichotomy boundary value is set as the scale value boundary, and then compare with the corresponding maximum response value. If \( f(s_1) > f(s_2) \), set \( s_2 \) as \((s_1 + s_2)/2\), otherwise, set \( s_1 \) as \((s_1 + s_2)/2\), and carry out the calculation until the end of iteration to obtain the corresponding optimal scale factor. Assuming the number of scales needed to compare is \( n \), the complexity of the traditional exhaustive algorithm is \( O(n) \), and its computational complexity of the scale after the improved dichotomy is \( O(\log n) \), which accelerates the tracking speed.

3.3. Re-identification detector \( R_c \)

The missing target re-identification is realized by training the SVM detector updated online. The first frame of image target is collected cyclically as the initial positive sample, and then the interval update method is used. In the unoccluded state, the high confidence tracking target is extracted every \( \Delta t \) frame as the training positive sample. And at the same time, the cyclic shift method is used to obtain sufficient negative sample, extract the HOG features of positive and negative samples and update the SVM detector online.

The SVM detector calculates the score of each bounding box when re-identifies the row target, and takes the candidate target with the highest confidence score to the tracker. The SVM detector outputs the \( t_{x}, t_{y} \) of the predicted bounding box, and the offset between there-identification area and the tracking image is \((c_x, c_y)\), then the re-detection target position \((b_x, b_y)\) assigned to the tracker can be
obtained by the formula as the following.

\[
\begin{align*}
    b_x &= \sigma(t_x) + c_x \\
    b_y &= \sigma(t_y) + c_y
\end{align*}
\]  

(10)

3.4. Update Mechanism

(1) The illumination threshold \( Tr \): For images with abnormal illumination conditions, adaptive correction is carried out to avoid excessive processing of normal images and ensure tracking effect. The normal light intensity pixel ratio (IPR) of the image is taken as the illumination threshold to determine whether the current frame image needs to be subjected to illumination normalization. The IPR calculation formula is as follows:

\[
    \text{IPR} = \frac{I_{\text{high}} - I_{\text{low}}}{XY}
\]

(11)

where \( X \) and \( Y \) are the horizontal and vertical boundaries of the image. \( I_{\text{low}} \) and \( I_{\text{high}} \) are the number of pixels whose light intensity is lower than \( h_{\text{min}} \) and also lower than \( h_{\text{max}} \) in the image. Based on this, the illumination distribution level of the image is evaluated according to IPR, and the illumination threshold \( Tr \) is set. When \( \text{IPR} < Tr \), activate light adaptive normalization processing.

(2) Trace the threshold \( Tf \): Target deformation or occlusion is a common problem in target tracking. If the tracker’s strategy is continuously updated, it is easy to misjudge the target in the case of similar targets or occlusion without considering the accuracy of the tracking results of the previous frame. This paper introduces average peak correlation energy (APCE), whose value reflects the fluctuation degree of correlation response. APCE is defined as follows:

\[
    F_{\text{APCE}} = \frac{|F_{\text{max}} - F_{\text{min}}|^2}{\text{mean}(\sum_{w,h}(F_{w,h} - F_{\text{min}})^2)}
\]

(12)

where \( F_{\text{APCE}} \) is the average peak correlation energy, \( F_{\text{max}} \) is the maximum response value, \( F_{\text{min}} \) is the minimum response value, and \( F_{w,h} \) is the response value of pixel points \((w, h)\). In this paper, the ratio \( Tf \) of APCE to the historical average is set as the adaptive tracking threshold. When the threshold is too small, the updating of the filter tracker is stopped, so as to maintain the robustness of the template.

(3) Detect the thresholds \( T1 \) and \( T2 \): Aiming at the detection threshold problem, we adopts two kinds of confidence evaluation indexes APCE and \( F_{\text{max}} \). When \( \text{APCE} < T_{1}, F_{\text{max}} < T_{2} \), the SVM re-identifier is activated.

Fig1 shows the effect of the update mechanism on lemming sequence. When there is no background noise, the target with high confidence can keep the template updated. When partial occlusion happens, APCE is less than \( Tf \), so the filter model stop update. When there is heavy occlusion or whole occlusion, \( \text{APCE} < T_{1}, F_{\text{max}} < T_{2} \), the detection result is not credible, and the re-identifier is initialized to re-identify the target. The improved target tracking framework, the running process is shown in Algorithm 1:

![Figure 1: Schematic diagram of update mechanism](image-url)
**ALGORITHM 1: Tracking Algorithm**

**Input**: Initial target boundary box $X_0 = (x_0, y_0, s_0)$, illumination threshold $T_0$, tracking threshold $T_r$, detection threshold $T_T$, $T_S$.

**Output**: Prediction bounding box $X_t = (x_t, y_t, s_t)$

**Repeat**

1. Extract the illumination component with multi-scale Gaussian function and calculate IPR
2. If $IPR < T_r$, adaptive brightness correction
3. According to $(x_{t-1}, y_{t-1})$ determine prediction area, trimme samples and extract features
4. Make displacement prediction based on regression model $R_a$, and obtain the predicted position $(x_t, y_t)$
5. Establish the scale pool on $(x_t, y_t)$, estimate scale $s_t$ by regression model $R_b$, calculate APCE and $F_{\text{max}}$
6. If $APCE < T_1$ and $F_{\text{max}} < T_2$
   - Start SVM to re-detect position $(x_t', y_t')$, and get the optimal scale $s_t'$ by regression model $R_b$
7. Else if $APCE > T_r$
   - Update the regression model $R_a$ and $R_b$, interval frame $\Delta T$ update the re-detection $R_c$

**Until** last frame

4. Experimental results

In order to verify the effect of the improved algorithm, ablation experiments and qualitative and quantitative analysis are conducted on the benchmark dataset.

4.1. Evaluation criteria

The evaluation (OPE) and spatial robustness (SRE) evaluation methods are used to compare with other algorithms. Precision represents the proportion of images whose predicted location is within a given ground truth threshold distance setting to 20 pixels, and the success rate represents the degree of overlap between the predicted target area and the actual area setting the overlap threshold to 70%.

4.2. Ablation experiment

In order to verify the effectiveness of the improvement measures of the tracking framework, we conducted ablation experiments based on the SRE evaluation criteria. Based on the kernel-related filtering framework, respectively evaluate the illumination adaptive and feature fusion tracker (KCF_IB), model confidence update tracker (KCF_CU), loss re-detection tracker (KCF_RD) and tracker (Ours) that integrates all improved frameworks performance.

| illumination variation | occlusion | background clutter | deformation |
|------------------------|-----------|--------------------|-------------|
| precision | Success | precision | Success | precision | Success | precision | Success | precision | Success |
| Ours 0.787 0.733 | 0.644 0.563 | 0.783 0.738 | 0.670 0.608 | 0.859 0.817 |
| KCF_IB 0.723 0.510 | 0.618 0.485 | 0.681 0.458 | 0.650 0.579 | 0.823 0.662 |
| KCF_RD 0.700 0.516 | 0.494 0.408 | 0.711 0.500 | 0.560 0.515 | 0.822 0.713 |
| KCF_CU 0.677 0.490 | 0.565 0.477 | 0.684 0.473 | 0.665 0.598 | 0.687 0.576 |
| KCF 0.660 0.468 | 0.535 0.438 | 0.664 0.446 | 0.625 0.559 | 0.700 0.581 |

In Table 1, the marked red is top performance data while second marked blue. Our algorithm achieve the best performance among these algorithm and each module has a suitable scenario.

4.3. Overall Performance

Considering common illumination changes, similarly deformed or occlusion, we qualitatively analyzed our algorithm with other kernel correlation filter algorithm by selecting 15 video sequences.
Fig. 2 shows that our algorithm is better than the original KCF and other 7 commonly tracking algorithms. In order to better illustrate the advantages of the improved algorithm, this paper conducts a quantitative analysis on the OTB100 benchmark data set, and the results are shown in Fig. 3.

According to quantitative analysis, the improved tracking algorithm can achieve best result than others in long-term tracking. Compared with the original kernel correlation filtering algorithm, precision is improved by 5.1% and accuracy is improved by 13.8%.

5. Conclusion
In this paper, we proposed a new tracking framework to solve the long-term tracking problem. We train and fusion position filter Ra, scale filter Rb and SVM detection. Based on the threshold value, we design an update mechanism to improve tracking performance. Finally, ablation experiments contrast experiments proved that our algorithm is more effective in terms of accuracy and robustness.

References
[1] Henriques, J.F., et al., High-Speed Tracking with Kernelized Correlation Filters. IEEE Trans Pattern Anal Mach Intell, 2015. 37(3): p. 583-96.
[2] Danelljan, M., et al., ECO: Efficient Convolution Operators for Tracking, in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
[3] Kalal, Z., K. Mikolajczyk and J. Matas, Tracking-Learning-Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence
[4] Wang, X., et al., An Optimal Long-Term Aerial Infrared Object Tracking Algorithm With Re-Detection. IEEE Access, 2019. 7: p. 114320-114333.
[5] Yu, M., et al., Distractor-Aware Long-Term Correlation Tracking Based on Information Entropy Weighted Feature. IEEE Access, 2020. 8: p. 29417-29429.