Scale Adaptive Target Tracking Algorithm Based on Correlation Filtering

Fajun Lin¹, Haiping Wei¹*, Yuanchen Qi¹ and Qinqin Li¹

¹Liaoning Shihua University, Liaoning, China

*1847749934@qq.com

Abstract. Focusing on the issue that the kernel correlation filter tracking method can not accurately adapt to the target scale change, a scale estimation method is added under the kernel correlation filter tracking frame, and a scale adaptive target tracking algorithm is proposed. The algorithm firstly uses the kernel function to obtain the kernel correlation filter, and estimates the position of the target for the candidate sample detection; then uses the scale estimation method to detect the target scale to further determine the optimal scale of the target to be tracked in the current frame; Finally, according to the occlusion judgment mechanism, the parameters of the target performance model and the scale model are updated online. The experimental results show that compared with kernel correlation filtering algorithm, the proposed algorithm can accurately target tracking process to adapt to changes in the target scale, and the center position error reduced by an average of 4.39 pixels.

1. Introduction

Target tracking has a wide range of applications in video surveillance [1], human motion analysis [2], human-computer interaction [2], and unmanned [3]. However, target tracking is still a challenging problem due to scale changes and target occlusion. The target tracking algorithm can be divided into a generation method and a discriminant method [4]. In recent years, due to the development of machine learning, the discriminant-based tracking method has achieved good tracking results, and the discriminant method trains a classifier based on the target and background information, and extracts the target from the background area [5].

Bolme et al. [6] introduced correlation filtering into the target tracking field for the first time, and proposed a Minimum Output Sum of Squared Error (MOSSE) based on correlation filtering. The algorithm used gray features to train the appearance model of the target area, and used Discrete Fourier Transform (DFT) to convert the similarity calculation between the target and all candidate areas to the frequency domain, which greatly enhanced the target tracking speed. Henrique et al. [7] proposed a Circulant Structure of Tracking by Detection with Kernels (CSK), the algorithm shifts training samples circularly, which can greatly increase the number of training samples. On the basis of this, Henrique et al. [8] proposed a Kernelized Correlation Filter (KCF) based on CSK, extracted the Histogram of Oriented Gradients (HOG) features to replace the former gray features [9], which improved the robustness of the algorithm.

Although the above algorithms have achieved good results in target tracking, these tracking algorithms do not take into account the scale changes of the target. Aiming at this problem, this paper proposes a method of target scale estimation, which realizes scale-adaptive target tracking. In order to verify the effectiveness of the proposed algorithm in this paper, the Matlab is employed to numerical...
simulations [10], further we verify the correctness of the algorithm theoretical derivation.

2. Tracking framework
The KCF algorithm trains the RLS classifier, solve the RLS classifier to obtain a filter template, and then detect the candidate samples to achieve tracking of the target [8].

2.1. Training kernel correlated filter
The purpose of training the Kernel Correlation Filter is obtained an objective function \( f(x) = w^T x \) that the minimizes the squared error of all samples and their corresponding regression labels.

\[
\min_w \sum_i (f(x_i) - y_i)^2 + \lambda \|w\|^2
\]  

(1)

The \( \lambda \) is a regularized parameter that controls overfitting. The \( w \) is a weight coefficient matrix, which is used to describe the relationship between the sample and the classifier.

To reduce the amount of computation, the input samples \( x \) are mapped onto a high dimensional feature space \( \phi(x) \), and a kernel function is defined \( k(x,x') = \phi^T(x)\phi(x') \). Expressing the solution \( w \) as a linear combination of the samples [8], [11]:

\[
w = \sum_i \alpha_i \phi(x_i)
\]  

(2)

The \( \alpha_i \) is classifier coefficients, the variables under optimization are thus \( \alpha \), instead of \( w \). The solution to the kernelized version of Ridge Regression is given by [8]:

\[
\alpha = (K + \lambda I)^{-1} y
\]  

(3)

Where \( I \) is an identity matrix, and each element of \( y \) is a regression target \( y_i \). Since the equation (3) has a matrix inversion operation, the circulant matrix can be used to simplify the operation.

2.2. Simplification of the circulant matrix
Consider an \( n \times 1 \) vector representing a patch with the object of interest, denoted \( x = (x_1, x_2, \cdots, x_{n-1}, x_n) \) , We will refer to it as the base sample. The cyclic matrix \( X \) can be expressed as:

\[
X = F \text{dig} (\hat{x}) F^H
\]  

(4)

Where \( F \) is a constant matrix, and \( \hat{x} \) denotes the DFT of the \( x \). From now on, we all always use a hat \( \hat{\cdot} \) as shorthand for the DFT of a vector. Equation (3) can be transformed into the Fourier domain, so that the calculation amount of the training is greatly reduced.

2.3. Target position detection
An image block of size \( z \) is extracted at the same center position of the current frame and the last frame. The element of matrix \( K^z \) is defined as \( k^z = k(x, z) \), and the matrix \( K^z \) is used to represent the kernel matrix between the training samples \( x \) and the candidate image patches \( z \) [8].

From equation (2), we can compute the regression function for all candidate patches get all the response:

\[
\hat{f}(z) = \hat{k}^z \otimes \hat{\alpha}
\]  

(5)

Where \( \otimes \) denotes dot product. The target position of the current frame is estimated by the detected maximum value of the response. Then the classifier is updated as follows:

\[
x_t = (1 - \eta)x_{t-1} + \eta x
\]

\[
\alpha_t = (1 - \eta)\alpha_{t-1} + \eta \alpha
\]  

(6)

(7)

Where \( \eta \) is the learning factor, \( x_t \) and \( x_{t-1} \) represent the target global appearance model of the current frame and the last frame. The \( \alpha_t \) and \( \alpha_{t-1} \) represent the classifier update parameters of the current frame and the last frame.
3. Scale adaptive target tracking
In this paper, the target scale estimation strategy and the model update strategy with occlusion judgment.

3.1. Adaptive scale estimation
In practice, the target has changed in size. The characteristics of the collected samples are inaccurate; therefore, it is necessary to solve the problem of scale change.

Assuming that the target size of frame $t-1$ is $M \times N$ and the scale filter size of frame $t-1$ is $S \times 1$. In the target position of the current frame, a total of $2M+1$ images of different scales are extracted [12], [13] and the scale formula can be written as:

$$l_i = 1 \pm \mu \cdot m \quad m = 0, 1, 2, \ldots, M; i = 1, 2, \ldots, 2M+1 \quad (8)$$

Where $\mu \in (0, 1)$ is the parameter for judging the scaling scale? When $\mu > 0$, $l_i > 1$, indicates the scale of magnification; when $\mu < 0$, $l_i < 1$ indicates the scale of reduction; when $l_i = 1 (m = 0)$, the target scale remains unchanged. Image patches can be represented as $S_i = l_iM \times l_iN$. The Responses and updates of scale filters refer to Section 2.

3.2. Classifier update
In the target tracking, when the target is occluded, a method for occlusion judgment detection update is adopted.

In the assumption of forward-backward motion trajectory consistency [14], [15], the main factor that interferes with target tracking is that it is blocked by other objects. Therefore, it can be used as a basis for judging whether the target is occluded. According to this idea, this paper adopts a method of judging occlusion by using the Euclidean Distance between two trajectories.

Assuming that the $S = (m, m+1, \ldots, m+i)$ is a sequence of images, forward tracking $k$ step from $t$ to $t+k$ can obtain a moving track $L_{f} = (x_0, x_1, \ldots, x_{i+k})$, $f$ represents forward motion, $k$ represents step size, $x$ represents coordinate position of target center point. Then, a backward motion trajectory $L_{b} = (\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_{i+k})$ is obtained by tracking $k$ step backwards from $t+k$ to $t$. $b$ represents backward motion, $\tilde{x}_{i+k} = \tilde{x}_i$. Forward-backward error can be expressed by the Euclidean Distance $d$ between two trajectories [15].

$$d = (|x_0 - \tilde{x}_0|^2 + |x_{i+k} - \tilde{x}_{i+k}|^2)^{1/2} \quad (9)$$

Where $x_0, \tilde{x}_0$ are the coordinates of the target central point? Normally, $d$ is relatively small, indicating that the target is not occluded. Take the smaller threshold $\delta$. If $\delta > d$, judge the target to be occluded. The update factor in the classifier and scale filter of the current frame is reduced by half. If $\delta < d$, the tracking is continued according to the original update method.

3.3. Algorithm steps
In this paper, target scale estimation method and the model update strategy with occlusion judgment, the steps of the algorithm are summarized as follows:

Input: An entire video sequence; the target position and scale of the previous frame.
Output: The position and scale of the current frame target.

1. According to the position of the previous frame, the size of $z$ image patch is extracted, the image patch and training sample are detected quickly. The maximum response value is obtained by the equation (5) to determine the position of the current frame.
2. At the position of the target, a total number of $2M+1$ images of different scales are extracted, the images are quickly detected, and get the Best Scale.
3. When the target is occluded, to update classifiers and scale filters.

4. Experiments
In order to verify the effectiveness of the proposed algorithm, the experimental analysis was compared
with the original algorithm using the evaluation method in reference [10].

4.1. Qualitative Analysis
In order to verify the effectiveness of the algorithm, the test video sequences are compared. The two algorithms use the same video sequence. Here, four video sequences are selected.

Experiment 1: For the case of scale change, the CarScale video sequence was selected. See Figure 1(a), Figure 1(b), the tracking frame size of the KCF algorithm is always unchanged, this algorithm can adapt to the change of the target scale well, and can reduce the relevant parameters of model updating in time under occlusion, such as 166th frame.

Figure 1. Tracking results of CarScale (Frame 19, 139, 166 and 240).

Experiment 2: For the case of scale change and occlusion, the Girl video sequence was selected. There are obvious scale changes and occlusion in the target of the video sequence. As shown in Figure 2(a) and Figure 2(b), the scale of the tracking frame of the KCF algorithm does not change, While the size of the tracking frame of the algorithm in this paper keeps pace with the target scale.

Figure 2. Tracking results of Girl (Frame 8, 156, 360 and 436).

Experiment 3: For the case of scale change, the Walking video sequence was selected. As shown in Figure 3(a) and Figure 3(b), comparing with this paper, the KCF algorithm has no scale change.

Figure 3. Tracking results of Walking (Frame 12, 86, 208 and 356).

Experiment 4: For the case of scale change and occlusion, the Walking2 video sequence was selected. See Figure 4(a) and Figure 4(b), the KCF algorithm tracking is not accurate, this algorithm is more effective tracking.

Figure 4. Tracking results of Walking2 (Frame 6, 178, 204 and 287).
4.2. Quantitative analysis
In order to verify the effectiveness of the proposed algorithm, adopt Central precision Error (CLE) and Distance Precision (DP) as evaluation criteria. \[ CLE = \left[ (x - x_c)^2 + (y - y_c)^2 \right]^{1/2}, \]
it refers to the distance between the center coordinates \((x, y)\) of the tracking target and the true center coordinates \((x_c, y_c)\).

| Algorithm | Video         | CarScale | Girl | Singer1 | David | Walking | Walking2 | Car4 | Dog1 | Doll | Skating1 |
|-----------|---------------|----------|------|---------|-------|---------|----------|------|------|------|----------|
| KCF       |               | 16.23    | 32.52| 12.73   | 9.48  | 7.31    | 20.24    | 9.77 | 3.95 | 10.07| 7.72     |
| Ours      |               | 14.32    | 12.14| 4.64    | 9.14  | 5.98    | 16.83    | 8.58 | 3.26 | 3.85 | 7.36     |

Table 1 shows the experimental results of 10 sets of center position errors. From Table 1, compared with KCF algorithm, CLE reduces by 4.39 pixels on average, which shows the effectiveness of the proposed algorithm.

DP is the ratio of the number of test video \(p\) frames to the total number of \(q\) frames, where \(p\) frames are the number of video frames whose CLE is less than a certain threshold (20 pixels in the experiment), obtaining \(DP = p/q\).

Figure 5 is the DP curve of a partial test video, where the abscissa represents the threshold of the center position error, the ordinate represents the precision, i.e. the ratio of \(p\) to \(q\). The higher the precision, the better the effect.

5. Conclusion
In this paper, a correlation filter algorithm for scale adaptive tracking is proposed. By adding a target scale estimation strategy, the problem that KCF algorithm has no target scale change is solved. In addition, the algorithm also adds a model update strategy with occlusion judgment, which reduces the tracking error by reducing the model updating coefficient under occlusion, so as to improve the robustness of the algorithm. The experimental results on several groups of test videos show that the proposed algorithm greatly reduces the CLE and improves the DP. Therefore, compared with the traditional KCF algorithm, the proposed algorithm has good tracking effect in scale transformation and target occlusion.
References

[1] Smeulders A W M and Chu D M. Visual Tracking: An Experimental Survey. J. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(7): 1442-1468.

[2] Zhang K and Zhang L. Fast visual tracking via deep spatio-temporal context learning. C. //Proc of the European Conference on Computer Vision, 2014.

[3] WANG N, SHI J and Yeung D Y. Understanding and diagnosing visual tracking systems. C. //Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE Press, 2015: 3101-3109.

[4] Liu Q, Zhao X and Hou Z. Survey of single-target visual tracking methods based on online learning. J. IET Computer Vision, 2014, 8(5): 419-428.

[5] Wei Q and Lao S Y. Review of visual target tracking based on correlation filter. J. Computer Science, 2015, 43(11): 1-5

[6] Bolme D S, Beveridge J R and Draper B A. Visual object tracking using adaptive correlation filters. C. Los Alamition: Computer Vision and Pattern Recognition, 2010

[7] Henriques J F, Caseiro R and Martins P. Exploiting the ncirculant structure of tracking-by-detection with kernels. C. New York: European Conference on Computer Vision, 2012

[8] Henriques J F, Caseiro R and Martins P. High-speed tracking with kernelized correlation filters. J. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596

[9] Dalal N and Triggs B. Histograms of oriented gradients for human detection. C. Santiago: Computer Vision and Pattern Recognition, 2005

[10] Wu Y, Lim J and Yang M H. Online Object Tracking: A Benchmark. C. // Computer Vision and Pattern Recognition. IEEE, 2013: 2411-2418.

[11] Schölkopf B and Smola A. Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT Press, 2002.

[12] Qian T H, Luo Z Q and Li G J. Scale adaptive improvement of kernel correlation filter tracking algorithm. J. Journal of Computer Applications, 2017, 37(3): 811-816.

[13] Danelljan M, Häger G and Khan F S. Accurate scale estimation for robust visual tracking. C. // British Machine Vision Conference. 2014: 65.1-65.11.

[14] Lv G J, Mao X and Hu Y J. One Method to Detect Tracking Occlusions Based on Evaluation Forward and Backward Errors. J. Infrared Technology, 2016.

[15] Zhao S S, Liu Xipeng and Shen J L. The Affection of Thickness Error on THz Refractive Index. J. Journal of Capital Normal University: Natural Science Edition, 2009, 30(2): 13-15