Spatio-Temporal Context Tracking Algorithm Based on Correlation Filtering

FANG Qi\textsuperscript{1a} ZHOU Hao\textsuperscript{2b} ZHANG Lu\textsuperscript{3c}

\textsuperscript{1}School of Information, Yunnan University
Information Institute
Yunnan University
+86 18987400211

\textsuperscript{2}School of Information, Yunnan University
Information Institute
Yunnan University
+86 13312525281

\textsuperscript{3}School of Information, Yunnan University
Information Institute
Yunnan University
+86 15023678304

\textsuperscript{a}493323675@qq.com \textsuperscript{b}zhouhao@ynu.edu.cn \textsuperscript{c}1069741887@qq.com

ABSTRACT: The Correlation Filter tracking algorithm is different from the traditional method based on target feature, which has high accuracy and fast tracking speed. In this paper, the Spatio-Temporal Context (STC) target tracking algorithm is combined with Correlation Filtering, the position of the target is predicted by using Correlation Filtering and the prediction result is corrected by using the STC target tracking algorithm. The target tracking algorithm combined with Correlation Filtering and STC can realize effective tracking when the target is occluded or there is an interference target. The experimental results show that the proposed algorithm can not only be applied to the tracking of visual targets in complex backgrounds such as illumination changes, target rotation and background region interference, but also have better robustness.

1. INTRODUCTION
With the rise of intelligence and automation, visual tracking has become an indispensable part of many industries, such as military guidance, security monitoring, motion analysis, activity identification, human-computer interaction, and many other fields. The field of computer vision is an important job. There are already many excellent visual tracking algorithms, such as Visual Tracking Decomposition (VTD)\textsuperscript{[1]} Tracking-Learning-Detection (TLD)\textsuperscript{[2]}, and Struck algorithm\textsuperscript{[3]}. These algorithms have a certain effect on accuracy. However, in the actual tracking process, on the one hand, the target may be deformed or rotated, and on the other hand, due to the influence of external environment such as illumination or obstruction, the above algorithm cannot accurately achieve target tracking.

Correlation filter tracking algorithms\textsuperscript{[4]} have proven to be more competitive than very complex methods because of the high frame rate and performance achieved with very few computational resources. The Spatio-Temporal Context (STC)\textsuperscript{[5]} algorithm is a simple, fast and robust algorithm that
uses the space-time scene model of the image for tracking. The spatiotemporal context target tracking algorithm is proposed under the Bayesian framework, so it is difficult to avoid the problem of tracking target drift.

In order to solve the problems of severe occlusion, severe deformation and high-speed motion in the tracking process, combined with the correlation filter, the STC target tracking algorithm has been improved, so that the target can be tracked effectively and stably when the target is completely occluded or moved at high speed.

2. CORRELATION FILTER

All The correlation filter was first applied to signal processing to describe the correlation between two signals. The correlation filter trains a linear classifier to distinguish whether the image block and its translation are targets. The correlation filter tracking algorithm [6] mainly utilizes the convolution of two image blocks to be equivalent to multiplying by element in the frequency domain. Therefore, by modeling the problem in the frequency domain, for multiple translations of the image, the output of the ideal linear classifier can be obtained by one calculation. Thereby the correlation filtering is very widely used in target tracking. In the target tracking process of the correlation filter tracking algorithm[6], the regression function is trained by the least squares method, and the formula is as follows:

\[
\min_w \|A_0w - y\|_2^2 + \lambda_2 \|w\|_2^2 \tag{1}
\]

Where \(A_0\) is a set of image blocks \(a_0\) obtained by cyclic shift in the training process, vector \(w\) represents a learning correlation filter, \(y\) is a regression target.

Let \(x^*\) denote its conjugate, and \(\hat{x}\) denote its Fourier transform \(F^Hx\), where \(F\) is the DFT matrix. Solve equation (1) using the following characteristics of the circulant matrix:

\[
X = F \text{diag}(\hat{x})F^H \quad \text{and} \quad X^T = F \text{diag}(\hat{x}^*)F^H \tag{2}
\]

After the formula (2) is diagonalized, the closed function can be solved by the objective function:

\[
\hat{w} = \frac{\hat{a}_0^* \hat{y}}{\hat{a}_0^* \hat{a}_0 + \lambda_k} \tag{3}
\]

The image block \(z\) represents the search window of the next frame, so that the trained filter \(w\) is convolved with \(z\) in the next frame, and the maximum corresponding position is the target position in the search window. The original detection formula is as follows:

\[
r_p(w, Z) = Z \leftrightarrow \hat{r}_p = Z \hat{\otimes} \hat{w} \tag{4}
\]

Where \(Z\) represents the circular matrix of the search window.

3. STC TRACKING ALGORITHM COMBINED WITH CORRELATION FILTERING

3.1 STC Target Tracking Algorithm

Based on the Bayesian framework, the STC algorithm establishes the spatio-temporal relationship between the tracking target and surrounding content, and models the statistical relationship between the target and the nearby region on the low-order features. The statistical correlation between the target and the surrounding low-order features is obtained. To estimate the confidence map of the target position \(x\) in a new frame, by calculating the confidence map, the position with the largest likelihood probability is the tracking target position of the new frame we obtained.

The calculation formula of the target position confidence map is as follows:

\[
c(x) = P(x|\omega) = \sum_{c(z) \in X^c} P(x|c(z) \omega)P(c(z)|\omega) \tag{5}
\]

Where \(P(x|c(z) \omega)\) represents the conditional probability of the spatial relationship between the modeling target and the surrounding context information, and \(c(z)\) represents the context feature at the \(z\)-point, including the pixel intensity and coordinate information of the image point; \(P(c(z)|\omega)\) represents the context prior probability of modeling each point \(x\) of the local context; \(\omega\) indicates that the target appears, and \(X^c\) is the current frame context feature set.

The conditional probability function is defined as:
\[ P(x|c(z), o) = h^{sc}(x - z) \] (6)

Where \( h^{sc}(x - z) \) is the non-radial symmetric function of the relative distance and direction of the target \( x \) from the local context position \( z \).

### 3.2 Optimization of STC Tracking Algorithm

STC tracking algorithm exhibits the advantage of fast tracking for common problems such as illumination changes and target rotation in target tracking. When the target is severely occluded or rotated, the pixel intensity and relative position of the target area will change greatly. The prior probability model calculated by Bayesian estimation is inaccurate, and the obtained target estimation position and the actual target position will be greatly deviated, which will cause problems such as tracking drift. Therefore, based on the STC algorithm, a correlation filtering algorithm is introduced to predict the motion trend of the target, and to correct the tracking result of the STC tracking algorithm.

#### 3.2.1 Initialization

In the initial frame, the selected target area is manually located, and the space-time context area confidence map \( c(x) \) and the prior probability model \( P(c(z)|o) \) of the tracking target are calculated using equation (7):

\[ c(x) = be^{-\left|\frac{x-x^*}{\sigma}\right|^\beta} \] (7)

Where \( b \) is the normalized constant, \( \alpha \) is the scale parameter, and \( \beta \) is the shape parameter.

\[ P(c(z)|o) = I(z)\omega_o(z - x^*) \] (8)

Where \( I(z) \) is the gray value of point \( z \), \( \omega_o \) is a weighting function, and the closer the point \( z \) is to \( x \), the larger the weight. At the same time, the geometric center position \( x(0), y(0) \) of the target tracking frame is obtained, and this initial center position is taken as the initial state of the context-aware correlation filtering algorithm.

#### 3.2.2 Tracking Target

In the \( t \)-th frame, the STC tracking algorithm is first used to obtain the maximum value of the target confidence map to determine the initial position of the target, and then the target position output by the STC tracking algorithm is used as the search window \( z(n) \) of the \( t \)-th frame of the context-aware correlation filter. The maximum response value of the current frame.

Using the context-aware correlation filter to correct the target position of the STC output, a penalty term \( \sum_{i=1}^{k}||A_iw||^2_2 \) is added to the solution of the context-aware correlation filter, so that the filter template \( w \) to be trained and the background \( A_i \) can be as small as possible in response. The updated target function is as follows:

\[ \min_{w}||A_0w - y||^2_2 + \lambda_1||w||^2_2 + \lambda_2\sum_{i=1}^{k}||A_iw||^2_2 \] (9)

The original objective function in equation (9) can form a new data matrix \( B \in \mathbb{R}^{(k+1)nxn} \) and a new target location \( \tilde{y} \in \mathbb{R}^{(k+1)n} \) by the context image block around the target image block.

\[ B = \begin{bmatrix} A_0 \\ \sqrt{\lambda_2A_1} \\ \vdots \\ \sqrt{\lambda_2A_k} \end{bmatrix} \text{ and } \tilde{y} = \begin{bmatrix} y \\ 0 \\ \vdots \\ 0 \end{bmatrix} \] (10)

Since the original objective function in equation (9) is convex, by setting the gradient to zero to minimize it, it produces:

\[ w = (B^TB + \lambda_1I)^{-1}B^T\tilde{y} \] (11)

After diagonalization according to formula (2), the closed solution in the Fourier domain is obtained as:

\[ \hat{w} = \frac{\hat{a}_o\hat{y}}{\hat{a}_o\hat{a}_0 + \lambda_1 + \lambda_2\sum_{i=1}^{k}\hat{a}_i\hat{a}_i} \] (12)
An update filter template \( \hat{\omega} \) of the current frame target position is obtained, and the target window determined by the updated filter template \( \hat{\omega} \) is taken as the final position of the target. At the same time, the final position is passed to the STC target tracking algorithm, and the spatiotemporal context models \( h^{sc}(x) \) and \( H_{t+1}^{stc} \) are updated according to the following formula.

\[
H_{t+1}^{stc} = (1 - \rho)H_{t}^{stc} + \rho h_{t}^{sc}
\]

(13)

Where \( \rho \) is a learning parameter.

The tracking model uses the confidence map of the \( t + 1 \) frame to calculate the maximum likelihood probability to obtain the new target coordinates. The calculation formula is as follows:

\[
x_{t+1}^* = \arg \max_{x \in \Omega(x_t)} c_{t+1}(x)
\]

(14)

The algorithm flow of this paper is shown in Figure 1.

![Figure 1 The algorithm flow](image)

At the same time, in order to speed up the algorithm, the target context area of the STC tracking algorithm is reduced from the original target area by 2 times to 1.7 times.

4. EXPERIMENT AND RESULT ANALYSIS

In order to verify the real-time and robustness of the proposed algorithm, the algorithm of this paper is experimentally evaluated on the open test data set. The experimental platform is: Inteli7 4GHz processor, 32GB memory, software environment is windows7 operating system and Matlab2015b. In this paper, the algorithm of this paper is tested on the tracker_benchmark_v1.0 platform. The selected public test data set is OTB-100[7], which basically covers various challenging factors in an uncontrolled environment.

In addition to the STC algorithm, this paper also selected seven mainstream tracking algorithms for testing, namely Tracking-Learning-Detection (TLD)[2], Adaptive Structural Local Sparse Appearance Model (ASLA)[7], Visual Tracking Decomposition (VTD)[1], Context Tracker (CXT)[9], Local. Sparse Appearance Model and K-Selection (LSK)[10], Circulant Structure of Tracking-by-Detection with Kernels (CSK)[11], and Struck algorithm[3].
4.1 Experimental Parameters
Throughout the experiment, the parameters in the algorithm are set as follows: the formula parameters appearing in the original STC algorithm are the same, the parameters in formula (7) are \( \alpha = 2.25, \beta = 1 \), and the parameters in formula (13) are \( \rho = 0.075 \). Other algorithms for comparison use default parameters. After the parameters were set, they remained unchanged throughout the test.

4.2 Analysis of Results
In this paper, the accuracy and accuracy of this algorithm are used to evaluate the performance of the proposed algorithm. The accuracy measurement mainly refers to the Euclidean distance between the predicted position center point and the center position marked in the benchmark, and is calculated in units of pixels. The success rate measurement mainly refers to the degree of coincidence of the benchmark in which the target is predicted. Figure 2 shows the accuracy and success rate of the tracking algorithm, STC tracking and other seven mainstream tracking algorithms in the OBT-100 with 51 large interference videos.

\[ S = \frac{|r_t \cap r_a|}{|r_t \cup r_a|} \] (15)

Figure 2 Average overall performance comparison in OTB-100

In general, the coincidence rate score (OS) is defined as follows:

\[ S = \frac{|r_t \cap r_a|}{|r_t \cup r_a|} \] (15)

Where \( r_t \) and \( r_a \) correspond to the target block obtained by the tracking and the target frame obtained actually. When the coincidence rate OS obtained by the tracking result in a certain frame is greater than the set threshold, the tracking of the frame is considered to be successful, and the general threshold is set to 0.5. The ratio of the number of successfully tracked frames in the video to the total number of frames is the tracking success rate.

Figure 2 uses the one-pass evaluation (OPE) evaluation method, considering the robustness evaluation, by moderately and sparingly, using temporal robustness evaluation (TRE) and spatial robustness evaluation (SRE) evaluation methods as shown in Figure 3. The average accuracy and success rate shown are shown.
It can be seen from the comparison of the experimental results in Fig. 2 and Fig. 3 that in the video sequence provided by the common data set OTB-100, the improved algorithm proposed in this paper is superior to the original STC algorithm under both measurement standards. In addition, compared with the other seven mainstream tracking algorithms, the algorithm has great advantages, and the best results are obtained in most experimental videos. Under the two measurement standards, the average of the algorithm in all videos is the results are the best. This shows that the tracking algorithm proposed in this paper is reasonable and effective. Figures 4 and 5 show screenshots of partial tracking results during the experiment.
5. CONCLUSION
In this paper, the correlation filtering algorithm is added to the original STC target tracking algorithm, so that the target can be used to correct the output of the STC target tracking algorithm when it is occluded, which effectively increases the robustness of the algorithm. The shortcoming of the algorithm in this paper is that the target tracking frame scale cannot be changed. After losing the target, the target cannot be found, and multiple targets are not considered. Subsequent work will further improve these deficiencies.

REFERENCES
[1] Kwon J, Lee K M. Visual tracking decomposition[C]// Computer Vision & Pattern Recognition. 2010.
[2] Kalal Z, Mikolajczyk K, Matas J. Tracking-Learning-Detection[J]. IEEE Trans Pattern Anal Mach Intell, 2012, 34(7):1409-1422.
[3] Hare S, Saffari A, Torr P H S. Struck: Structured output tracking with kernels[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 38(10):2096-2109.
[4] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[J]. 2010.
[5] Zhang K, Zhang L, Liu Q, et al. Fast Visual Tracking via Dense Spatio-temporal Context Learning[J]. 2014.
[6] Mueller M, Smith N, Ghanem B. Context-Aware Correlation Filter Tracking[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.
[7] Wu Y, Lim J, Yang M H. Online Object Tracking: A Benchmark[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013.
[8] Xu J. Visual tracking via adaptive structural local sparse appearance model[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2012.
[9] Dinh T B, Vo N, Medioni G. Context tracker: Exploring supporters and distracters in unconstrained environments[C]// Computer Vision & Pattern Recognition. 2011.
[10] Liu B, Huang J, Lin Y, et al. Robust tracking using local sparse appearance model and K-selection[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2011.
[11] João F. Henriques, Caseiro R, Martins P, et al. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels[M]// Computer Vision – ECCV 2012. Springer Berlin Heidelberg, 2012.