Journal of Information Hiding and Multimedia Signal Processing ©2019 ISSN 2073-4212
Ubiquitous International Volume **10**, Number **1**, January **2019**

Robust Target Tracking Based on Spatio-Temporal Context Learning

Huimin Xiao*, Xiao Liu

Shandong Institute of Space Electronic Technology, Yantai, China
*Corresponding author: xiaohuimin2006@126.com

Received April 2018; revised September 2018

Abstract. *Accuracy and real-time are two important indicators of target tracking. While most trackers can meet the real-time requirement, the important factor that affects tracking accuracy is occlusion, and it is also a key index to evaluate the robustness of the tracking algorithm. In this paper, we propose a framework of object tracking based on STC (Spatio-Temporal Context) algorithm, which improves the tracking algorithm with an anti-occlusion scheme. During tracking process, calculate the perceptual hash values of target context regions between adjacent images to predict tracking state and extract the template. And then determine whether target is lost according to the matching of template with current target. When it encounters severe occlusion, we cannot extract valid information. Re-capture the target with template matching after leaving the occlusion. The experiments proved that the proposed framework can track rigid object effectively under the condition that the target is completely obscured for a while, which improves the accuracy of the algorithm while ensuring a real-time tracking speed.*
Keywords: Robust target tracking, Spatio-temporal context learning, Perceptual hash, Template matching.

1. **Introduction.** The two important evaluation indicators of target tracking are real-time and accuracy [1]. A good tracker should not only meet the requirement of real-time, but also try to improve the accuracy of tracking as much as possible. However, the occlusion will reduce the accuracy greatly [2]. So improving the occlusion problem is of great help for good tracking performance. The occlusion process can be divided into three stages [3]. The first stage is getting into occlusion. During this time, the information of the target is gradually lost. Secondly, the target is in the block and the information remains a missing state. In the third stage, target leaves the occlusion and its information is restored by degrees [4]. Occlusion causes the instability of the target information or even lost, while the key to tracking is to search for enough information to locate where the target is [5]. Therefore, the algorithm should judge the occlusion accurately and keep tracking with the residual information, even if the target is completely blocked [6]. In this paper, a framework of target tracking is proposed to improve the tracking performance from the perspective of accuracy. It's difficult to solve the problem by using the existing tracking algorithm [7]. A target tracking with a blocked scene may lose target or make other tracking errors. As a result, this paper puts forward a solution for the occlusion problem and integrate it with a high speed tracking solution to improve the algorithm for better tracking performance.

2. **DESIGN OF THE FRAMEWORK.** The framework of object tracking is shown in Fig. 1, consisting of two parts: tracking solution and occlusion solution. Meanwhile, the occlusion solution includes three parts namely state judgement, trajectory prediction and target recapture.

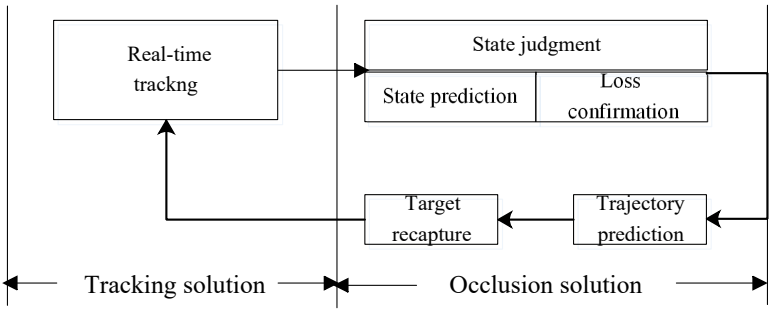

FIGURE 1. Framework of object tracking

Fig.2 shows the flow chart of the tracking framework composed of solutions in Fig.1. Detect the target state while tracking and if there is an abnormal state, extract target area as a template. After that, confirm whether the target was lost. If not, continue the tracking process, but if the target was lost indeed, try to predict the target movement and recapture it while it appears again. If there's no catch, the target is still in loss, then continue to predict the trajectory until find the target.

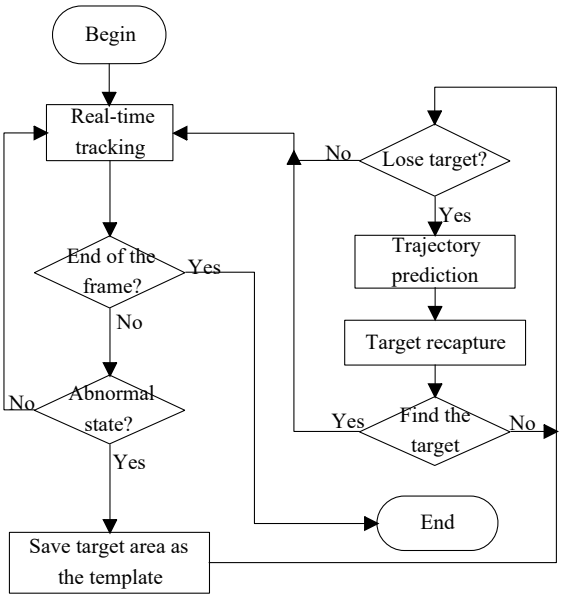

FIGURE 2. Flow chart of the tracking framework

2.1. **Target tracking algorithm.** We study and test a variety of trackers based on neural network, kernel learning, subspace learning, SVM and so on [8]. According to the three evaluation criteria of real-time, accuracy and robustness [9], we get the conclusion that the method of tracking using CNN (Convolution Neural Network) model has high accuracy, but it cannot meet the requirements of real-time. Kernel learning tracker has a greater advantage in tracking speed, but are slightly less accurate in tracking accuracy.

The STC tracking method based on subspace learning is much faster than other solutions and has good tracking accuracy. After a comprehensive analysis, we use the STC algorithm to make the tracker.

STC algorithm is a simple and fast visual tracking solution based on Bayesian framework [10, 11, 12]. This approach establishes the spatio-temporal relationship between the object interested and its local context [13]. It transforms the tracking into calculating a confidence map [14] as in (1).

$$m(x) = P(x|o) = \sum_{c(z)\in X^c} P(x, c(z)|o) = \sum_{c(z)\in X^c} P(x|c(z), o)P(c(z)|o) \tag{1}$$

Where x represents the target position, o represents target, c(z) represents the context feature. From the formula, we can see that the confidence map is composed of two probability model functions, which is convoluted by two models. One is spatial context model in (2).

$$P(x|c(z), o) = h^{sc}(x - z) \tag{2}$$

$h^{sc}(x-z)$ is a function representing the relative distance and direction of target position x and its local context position z, and it establishes the spatial relationship between target and its surrounding [15].

The other model in (3) is related to the appearance of target and is modeled by the principle of focus of attention, which concentrates on certain image regions requiring detailed analysis [11].

$$P(c(z)|o) = I(z)\omega_\sigma(z - x^*) \tag{3}$$

Where I(.) represents image grayscale and indicates the context appearance. is a Gaussian weighting function, defined as (4) and a is a normalized constant which limits the probability to 0 to 1 [15].

$$\omega_\delta(z - x^*) = ae^{-\frac{z-x^*}{\delta^2}} \tag{4}$$

In tracking process, the spatial model of t-th fame is in (5).

$$h_t^{sc}(x) = F^{-1}(\frac{F(m(x))}{F(I(x)\omega_\sigma(x - x^*))}) \tag{5}$$

The spatio-temporal context model of (t+1)-th frame is updated by the spatial model in t-th frame [16], where is the learning parameter.

$$H_{t+1}^{stc} = (1 - \rho)H_t^{stc} + \rho h_t^{sc} \tag{6}$$

Finally, the maximum likelihood function for locating target position in (t+1)-th frame is defined as:

$$m_{t+1}(x) = H_{t+1}^{stc}(x) \bigotimes (I_{t+1}(x)\omega_{\sigma_t}(x - x_t^*)) \tag{7}$$

The optimal target position is obtained by maximizing the object location likelihood function [13, 17]. In this tracking solution, it uses Fast Fourier Transform to speed up the spatial context learning and target detection [18].

This scheme uses the relationship between the target and the surrounding environment to make the tracking. When the target is blocked gradually, tracking result also changes into an occlusion region. As the target and its surrounding area change no longer, the tracker doesn't follow the targets movement to track continuously. We need a solution to improve this situation and continue to track the target. According to the analysis of the occlusion process in article. Occlusion caused the instability of the target information or even lost, while the key to tracking is to search for enough information to locate where the target is [19]. Therefore, judging the moments of occlusion is the key to dealing with this problem. Only the correct prediction of obstacles, can we save the target information in time and recapture the target when it appears again.

2.2. **State judgement.** To make a judge of whether the target is lost accurately, we use two methods to detect the movement state. As shown in Fig.3, the target loss detection process consists of two parts, which are abnormal state detection and loss confirmation respectively. Since the context is twice the target area, we use the context area to check whether the target is in a normal tracking state in advance. After several frames running, check the movement state in another way to confirm whether the target is lost.

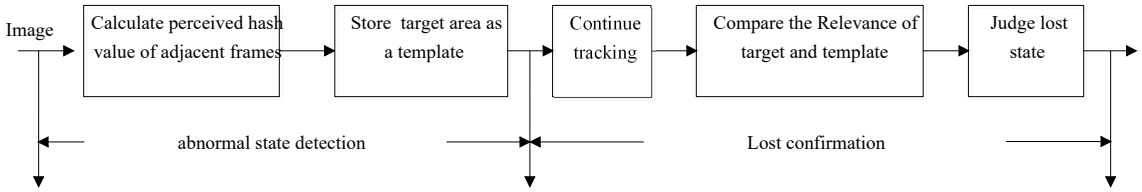

FIGURE 3. Block diagram of the loss detection state

2.2.1. *Abnormal state detection.* We determine the correlation between the front and rear frames by calculating perceptual hash of context between adjacent images to determine whether the motion state of the similarity of two images. The closer the results, image is more similar. Normally, you can judge the images are similar when Hamming distance is less than 5. The operation is as follows:

a) Remove high-frequency features by reducing image size.

b) Simplify the color to 64 grayscale

c) Calculate the DCT coefficient in (8) to extract low-frequency information

d) Calculate the average of DCT coefficient to make hash value and create a fingerprint by it

$$F(u,v) = \frac{2}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\frac{(2x+1)\mu\pi}{2N}\cos\frac{(2y+1)\nu\pi}{2N} \tag{8}$$

2.2.2. *Loss confirmation.* Use the matching degree of two images to judge the motion state again. Common matching methods includes squared difference matching method, correlation matching method, correlation coefficient matching method and normalization method. We choose the normalized correlation matching method which is more accurate with an empirical threshold obtained by a large number of experiments to judge whether the target is lost. The formula is (9).

$$R(x,y) = \frac{\sum_{x',y'}(T'(x',y') \bullet I'(x+x',y+y'))}{\sqrt{\sum_{x',y'}(T'(x',y')^2 \bullet \sum_{x',y'} I'(x+x',y+y')^2)}} \tag{9}$$

2.3. **Trajectory prediction.** When the unusual state is detected, the objects is not blocked immediately. Save the location as the movement target is normal. Perceptual hash algorithm uses the DCT (Discrete Cosine Transform) to extract the low-frequency components of the image to generate a fingerprint string for each image and compare the fingerprint of different images [20]. Use Hamming distance to determine the information and take the target region as a template for matching calculation and target recapture. We can use the remaining target information to predict the motion trajectory. A tracking algorithm based on trajectory prediction has a good effect for occlusion of linear motion. In the target loss stage, the target information is completely gone. Assuming that the movement has not change a lot and the target trajectory is a uniform straight line, use the target positions before loss to establish the linear predicting function as the movement direction, and take the average distance between target locations as the speed. The movement formula is as follows, where pPoint represents the target location before it lost.

$$\frac{y - pPoint1.y}{x - pPoint1.x} = \frac{pPoint5.y - pPoint1.y}{pPoint5.x - pPoint1.x} \tag{10}$$

$$\Delta x = \frac{\sum_{i=1}^{5} pPoint_i.x}{5}, \Delta y = \frac{\sum_{i=1}^{5} pPoint_i.y}{5} \tag{11}$$

2.4. **Target recapture.** We use the template matching to re-capture the target. Template matching is one of the ways to find a target in an image [21]. The template matching works in the same way as the reverse projection of the histogram, and by matching the actual image block and the input image by sliding the image block on the input image, each pixel gets a matching metric [22]. Other algorithms on image feature extraction are presented in the privious works [23, 24, 25], and these methods also can be used to in the object tracking. Usually take its white region as the highest matching to get the position. Through a large number of test experiments, we select the standard coefficient matching as template matching criteria. According to the predicted target position, the matching degree is calculated in the region with the range of 3 times the target size, and the position with the largest correlation coefficient represents the point with the highest matching degree.

3. **RESULTS AND ANALYSIS.** Experiments were made and the test data are pktest02 datasets, which is a group of infrared date with a resolution of (320*256) and FaceOcc1 datasets with a resolution of (352*288). Pktest02 shows the situation that vehicles travel through some occlusion by trees and passing through shadows. FaceOcc1 shows a woman was blocked by a book.

FIGURE 4. Tracking effect of original algorithm

The tracking effect of original algorithm is as shown in Fig 4. In pktest02, the tracking target changes from a car to an occlusion and the spatial relationship between the target and the surrounding background is substantially unchanged, the rectangle no longer move, while the target continues to move. In FaceOcc1, when occlusion appears, the tracking box moves with the book and error happens. We cannot tracking the woman anymore.

We calculate Hamming distance varying with the frame numbers and the results are shown in Fig.5, which is corresponding to the sequences from frame0 to frame200 in pktest02. The abscissa is the frame number and the ordinate is the Hamming distance, while the area with the ordinate 0 is the occlusion time.

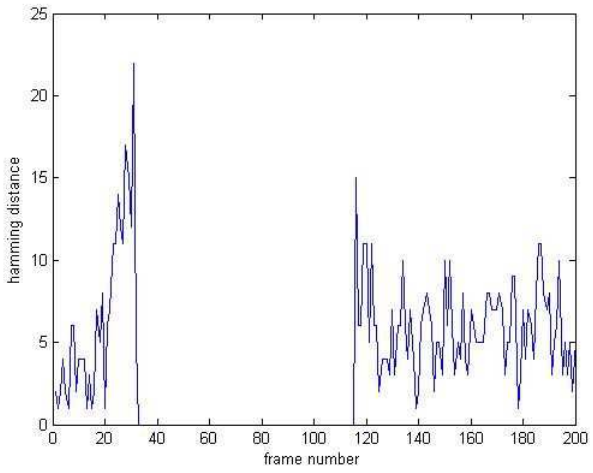

FIGURE 5. Hamming distance varying with frame numbers

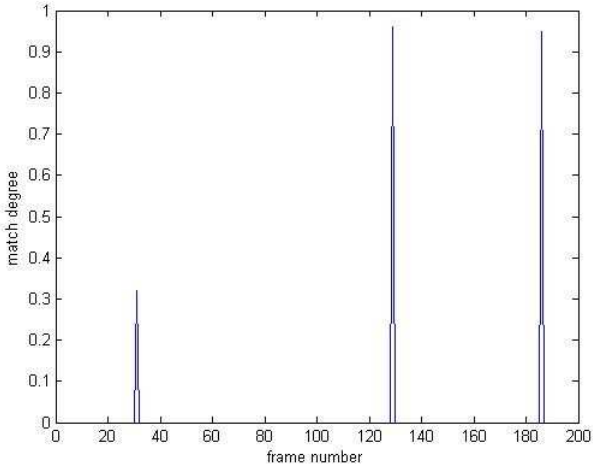

FIGURE 6. Matching degree after 10 frames of abnormal state

Fig.6 shows the matching degree corresponding to Fig.5, and the frame is shown in Fig.7. It can be seen from Fig.6 that there are three abnormal state. The first time Hamming distance is more than threshold and matching degree is low. We confirm the target is lost. For the second time, prediction makes a mistake but the matching degree is high. We confirm that target isn't lost. When other suspected targets appear, use hash

FIGURE 7. Frame images corresponding to abnormal state moments

value to make a prediction and matching degree to confirm the normal state. So that this solution can both detect the abnormal circumstances and ensure tracking as before.

The trajectory prediction of the uniform linear motion is shown in Fig. 8. The target is completely blocked by tree, we use a uniform linear motion to predict the movement of the occlusion period, as shown in the dotted line, and the rectangle box indicates the predicted motion state.

FIGURE 8. Linear prediction in target loss time

The images of the recapture process are shown in Fig. 9. As shown in Figure 9(a), the template stored in the loss detection phase is used to find the predicted target position within 3 times the size of the target area to improve the matching speed and accuracy.

FIGURE 9. Process of capturing the target

Figure 9(c) shows the tracking results in image sequences with a block by using the original algorithm and the new tracking framework separately. The frame number is as shown in the upper left corner and the target is in the red rectangle.

The tracking results using the framework proposed is shown in Fig.10. By comparing the results between the original solution and the improved scheme, which is as shown in Fig.4 and Fig.10, we can see that the algorithm with occlusion solution can keep tracking in the presence of an obstruction, which improves the robustness of tracking and is still able to keep the tracking speed.

4. **Conclusions.** In this paper, a framework of object tracking based on STC algorithm is proposed. We design the solution that could improve the tracking results when the target encountered occlusion by improving the accuracy of tracking, while the speed keeps as fast as before. Blocking is an important factor in tracking accuracy. We combine the tracking scheme with an anti-occlusion solution with the method of loss detection, trajectory prediction and recapture process to improves the performance of occlusion

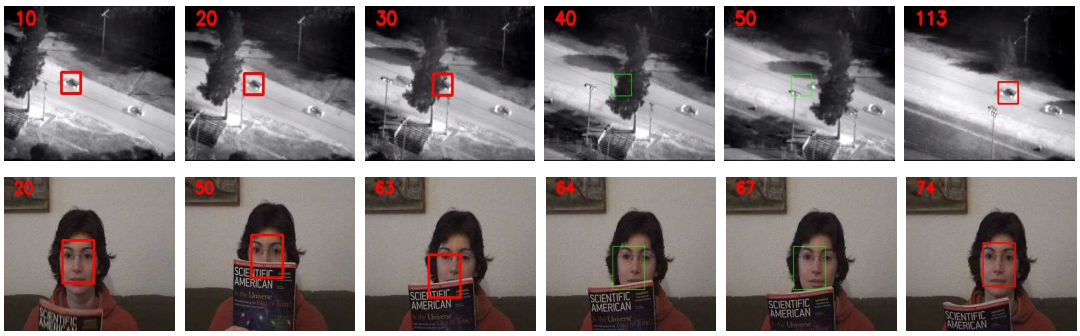

FIGURE 10. Tracking effect of the proposed framework

tracking of rigid target. By analyzing the results of occlusion process, we draw a conclusion that the occlusion scheme can use perceived hash value and template matching to judge the occurrence of occlusion effectively and accurately, predict the trajectory while target is vanishing and catch the target in time after the occlusion disappears. The proposed framework improved the performance of tracking while with occlusion in the tracking scene, thus improve the accuracy of the tracking process.

Acknowledgment. This work is supported by National Science Foundation of China under Grant No. 61671170, Natural Science Foundation of Heilongjiang under Grant No. F2015003, the Open Projects Program of National Laboratory of Pattern Recognition under Grant No.201700019.

REFERENCES

[1] X. Xi. *Study on Target Tracking Algorithm in Aerial Video*. Xidian University, 2009.
[2] K. J. Liu. *Study on Target Tracking Algorithm Based on Particle Filter and Corner Matching*. Journal of University of Science and Technology of China.
[3] X. Chen. *Study on Algorithm and Application of Video Target Tracking Technology in Complex Environment*. Graduate School of Chinese Academy of Sciences (Changchun Institute of Optics and Fine Mechanics and Physics), 2010.
[4] Z. Hao. *Study on video object detection and tracking algorithm in complex scene*. Yunnan University, 2011.
[5] MA Li. *Study on occlusion in target tracking*. Shandong University, 2006.
[6] J. Biao, W. L Hu, H. Q. WANG. *Study on Moving Object Tracking Masking Based on Multi-Level Tracking Queue*. Acta Optica Sinica, 2011, (08): 219-226.
[7] W. G. Gong, X. Wang, Z. H. Li. *A Real-time Detection and Tracking Algorithm for Anti-blocking Infrared Multi-target*. Journal of Instrumentation, 2014, (03): 535-542.
[8] W.Gap, M. Zhu, B. G. He, X. T. Wu *Study on Target Tracking Technology* . Chinese Optics, 2014, (03): 365-375.
[9] L. L. Ma, J. G. Chen, X. M. Hu, P. Ma. *Application Criteria for Target Tracking Filter Performance*. Journal of Xi'an Engineering University, 2013, (03): 364-368.
[10] J. Q Xu, Y. Lu. *A robust visual tracking algorithm based on weighted temporal and spatial context*. Acta Automatica Sinica, 2015, (11): 1901-1912.
[11] K. Qian, X. H. Chen, B. W. Sun. *Spatiotemporal context a robust and fast tracking algorithm*. computer engineering and application, 2016, (12): 163-167.
[12] S. M. Wang, Y. T. Chen *In house, super pixel level spatiotemporal context weighted target tracking*. computer application research, 2017, (01): 270-274.
[13] K. H. Zhang, L. Zhang, M. H. Yang, D. Zhang. *Fast visual tracking via dense spatio-temporal context learning*. European Conference on Computer Vision(ECCV), Zurich: Springer,2014:147-141.
[14] S. Li. *Study on target tracking algorithm based on dense space-time context*. Chongqing University, 2016.
[15] Z. Zhao, P. F. Huang, L. Chen. *Improved space-time context tracking algorithm based on fusion Kalman filter*. Acta Aerospace Sinica, 2017, (02): 274-284.

[16] W. J. Liu, S. Y. Dong, H. C. Qu. *Space-temporal context anti-occlusion visual tracking.* Journal of Image and Graphics, 2016, (08): 1057-1067.
[17] Z. W. Zhang. *Study on dense sampling target tracking method based on space-time context and kernel correlation filter.* Northwest A F University, 2015.
[18] L. Zhang, Z. Q. Hou, Q. S. Yu, W. J. Xu. *Two-layer search target tracking algorithm using fast Fourier transform.* Journal of Xidian University, 2016 (05): 153-159.
[19] Y.J. Wang. *Study on occlusion problem in target tracking process.* Huazhong University of Science and Technology, 2004.
[20] K. Wen, J. H. Gao. *Study on Video-aware Hash Algorithm for Converting Time-Domain Domain Change Information.* Acta Electronic Journal, 2014, (06): 1163-1167.
[21] J. L. Chen, D. Miao, B. Kang, J. R. Shen. *Study on Target Tracking Technology Based on Kalman Filter and Template Matching.* Optics and Optoelectronic Technology, 2014, (06): 9-12.
[22] J. J. Hu. *Study on Multi-Template Matching Algorithm for Target Tracking in Complex Scene.* National Defense University of Science and Technology, 2010.
[23] D. N. Zhao, D. J. Guo, Z. M. Lu, H. Luo. *Tracking Multiple Moving Objects in Video Based on Multi-channel Adaptive Mixture Background Model.* Journal of Information Hiding and Multimedia Signal Processing, Vol. 8, No. 5, pp. 987-995, September 2017
[24] J. Wang, Y. Y Wang, K. Wang and C. Z. Deng,*l1-Regularized Hull Representation for Visual Tracking.* Journal of Information Hiding and Multimedia Signal Processing, Vol. 9, No. 2, pp. 313-324, March 2018
[25] Y. Y. Wang, J. Wang. *L1-L2 Norms Based Target Representation for Visual Tracking.* Journal of Network Intelligence, Vol. 3, No. 2, pp. 102-112, May 2018