Railway Foreign Object Tracking Based on Correlation Filtering of Optimized Regularization Model

Tao Hou[1], Yannan Chen[*], Caiwen Bao[1], and Yuhu Chen[1]

[1] School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China
[*] Corresponding author. E-mail: 2282446083@qq.com

Received: May. 14, 2021; Accepted: July. 29, 2021

Aiming at problems such as the untrustworthy association between spatial regularization weight and intrusive foreign object in complex railway scenes, as well as the degradation of correlation filter model, fully excavate the expressive ability of deep space features, and a foreign object tracking algorithm based on correlation filtering with depth space and time perception regularization is put forward. Firstly, selecte the fifth-level convolution feature of the Visual Geometry Group (VGG) network to extract the spatial area information of the foreign object, which is used to solve the regularization guide weight.

Secondly, a regularization term based on depth space is added to the objective function, whose aim is to establish a more reliable association between the spatial regularization weight and the invading foreign object. Thirdly, the time perception term is added to establish the connection between the filters in time. Finally, based on the depth space, a simple and effective model update strategy is proposed. On the public OBT datasets and complex railway scenes, the tracking results of the algorithm in this paper and the existing multiple algorithms are compared and analyzed. The results show that in complex railway scenes, the algorithm in this paper is superior to other algorithms in distance accuracy and success rate. The tracking speed is 23.1FPS, which basically meets the real-time requirements. Therefore, the correlation filtering algorithm of the improved regularization model is of great significance to railway safety.

Keywords: railway foreign object tracking; correlation filtering; depth space; time perception; spatial regularization; model update

http://dx.doi.org/10.6180/jase.202204_25(2).0015

1. Introduction

With the rapid development of railway industry, the mileage of operating lines has also been increasing. But most of the routes are in more complex environments, such as jungles, deserts, and villages. People, animals, cars, rolling rocks, etc. may invade the railway boundaries and pose a threat to the safe operation of trains [1–4]. Therefore, it is an important prerequisite to detect and track railway intrusive foreign objects accurately in real time for ensuring train safety. Because the grating fiber detection method and the power grid detection method ,which detect foreign objects through contact alarms, need to be scattered along the line, the preliminary engineering volume is relatively large [5].

The method of installing radar, ultrasonic and infrared sensors along the railway is susceptible to strong electromagnetic interference from electrified sections [6]. Therefore, the video detection method is favored. Zhou Ruilin applied the scale-adaptive kernel correlation filtering algorithm to the fast tracking of railway intrusion foreign objects. This method uses fast histogram of oriented gradient features and color names features to describe railway foreign objects, but the algorithm itself is affected by the boundary effect caused by cyclic shift density sampling,

Table 1. Filter training process chart

Input: image sequence $F_1, F_2, .., F_t$
Output: The tracking result ,that is, the target position in each frame of image.
Initialization: Co-process the initial continuous M frame images, and initialize ω. Filter training: 1. Solve ω_h, By $q_{(l,v)}$ solving the space indicator matrix $\mathbf{q}$ as in Eq. (12), solve ω_h by Eq. (15). 2. Solve ω . By the ADMM method, the solution is as Eq.(18). 3. Solve ϕ. Substitute into (Eq. (7)) to solve ϕ optimally. Model update: According to $\mathbf{q}$ solving N_q, by (Eq. (19))determine the update strategy.

and its robustness is poor [7].

In recent years of research, many algorithms have been combined depth features with correlation filtering, and their performances have been significantly improved on a large number of datasets. Ma et al. combined multi-layer depth features with correlation filtering algorithms, and studied the expressive capabilities of different layers of depth features [8]. Dai Jiashu et al. fully utilized the advantages of correlation filter and siamese network, proposed a robust single-object visual tracking framework based on fully convolutional siamese network and correlation filter. However it is still affected by the boundary effect generated by the periodic hypothesis of cyclic sampling [9].

In order to better solve the boundary effect problem, spatial-variation regularization correlation filters [10], the correlation filtering tracking algorithm based on adaptive spatial regularization [11] and correlation filtering algorithm for spatio-temporal regularization [12] has been proposed one after another. The above algorithms have effectively suppressed the boundary effect, but the weights of the spatial regularization term of [10] have not been associated with the continuous changes of the object. And [11] fails to establish the relationship between the filters in time. When the object has a large deformation, the filter is easy to overfit to learn the current inaccurate target. The spatial regularization weight of [12] has no learning ability, and tracking drift phenomenon is prone to occur when encountering background interference.

Therefore, in view of the shortcomings of the above-mentioned foreign object tracking methods, a correlation filtering foreign object tracking algorithm based on depth-space and temporal perceptual regularization is proposed in this paper. The regularization model in this paper can make the spatial regularization weight update continuously and accurately as the object changes. The time perception term can effectively constrain the filter learned in the current frame to be as similar as possible to the filter in the previous frame, using the filters learned at the current moment and the previous moment, and thereby effectively mitigating model pollution.

2. Regularization model based on depth space and time perception

The objective function of the algorithm in this paper is shown in Eq. (1).

$$\begin{aligned} \mathrm{L}(\phi,\omega,\tau) = \frac{1}{2}\left\|\sum_{d=1}^{D} x_d \otimes \phi_d - y\right\|^2 + \frac{1}{2}\left\|\sum_{d=1}^{D} \omega \odot \phi_d\right\|^2 \\ + \frac{\eta}{2}\left\|\sum_{d=1}^{D} \omega - \omega_h\right\|^2 + \frac{\mu}{2}\left\|\phi - \phi_{t-1}\right\|^2 \end{aligned} \tag{1}$$

Where represents the total number of feature channels; x_d represents the feature of the channel d; ϕ_n represents the filter of the channel d; $\otimes$ means space domain loop related operations, the same as the full text; $\odot$ means product by element, the same as the full text; y represents the desired output, which sets as a two-dimensional Gaussian distribution centered on the target; ω_h represents the regularization guide weight with the depth space information of the foreign object; η represents the spatial regularization parameter; $\frac{\eta}{2}\left\|\sum_{d=1}^{D}\omega - \omega_h\right\|^2$ represents the regularization term based on the depth space. It can be seen that the spatial regularization weight ω combines the prior guidance weight ω_h and can learn more accurate spatial penalty weight while changing with the target; μ represents the time perception parameter; $\frac{\mu}{2}\left\|\phi - \phi_{t-1}\right\|^2$ represents the time perception term, which is used to limit the sudden change between adjacent filters.

Since the objective function is convex and there is no closed solution, Alternating Direction Method of Multipliers (ADMM) [11] is used to solve the optimal solution. Therefore, the penalty parameter γ is introduced, the auxiliary variable ψ is added, and the constraint condition $\phi = \psi$ is introduced, then the Lagrangian equation of Eq. (1) is as Eq. (2).

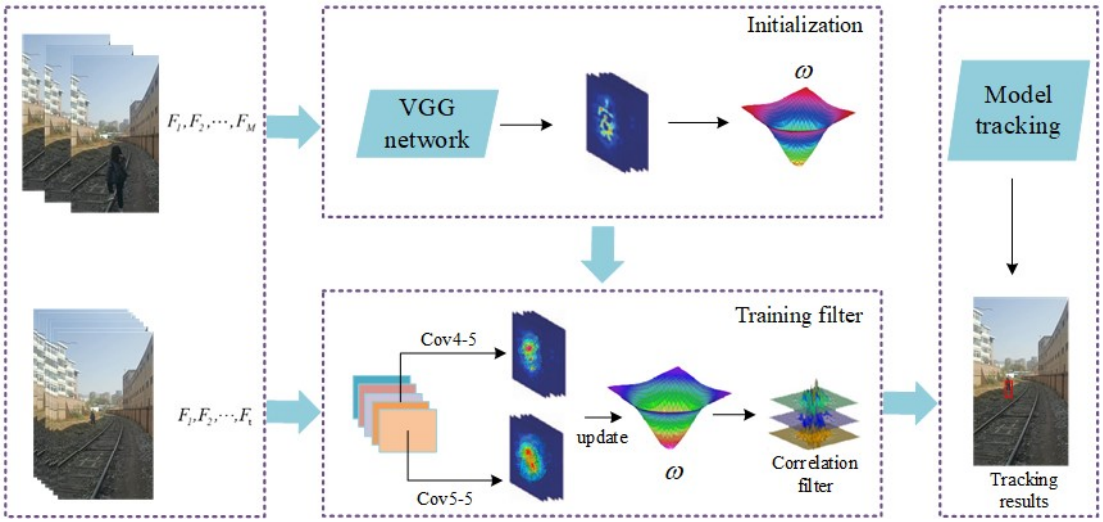


Fig. 1. The algorithm flow of this paper

$$\begin{aligned} \mathrm{L}(\phi,\psi,\omega,\tau) = \frac{1}{2}\left\|\sum_{d=1}^{D} x_d \otimes \phi_d - y\right\|^2 + \frac{1}{2}\left\|\sum_{d=1}^{D} \omega \odot \psi_d\right\|^2 \\ + \frac{\eta}{2}\left\|\sum_{d=1}^{D} \omega - \omega_h\right\|^2 + \frac{\gamma}{2}\sum_{d=1}^{D}\|\phi_d - \psi_d + \tau\|^2 \\ + \frac{\mu}{2}\|\phi - \phi_{t-1}\|^2 \end{aligned} \tag{2}$$

Among them, τ represents the Lagrangian multiplier; the size of the sample x_d is $D_m \times D_n$. Through ADMM iterative solution, the problem can be resolved into the following sub-problems.

Question 1: solve ϕ. Using Parseval's theorem, it is converted to the frequency domain to solve:

$$\begin{aligned} \mathrm{L}(\hat{\phi}) = \left\|\sum_{d=1}^{D} \hat{x}_d \otimes \hat{\phi}_d - \hat{y}\right\|^2 \\ + \gamma\left\|\hat{\phi}_d - \hat{\psi}_d + \hat{\tau}\right\|^2 + \mu\left\|\hat{\phi} - \hat{\phi}_{t-1}\right\|^2 \end{aligned} \tag{3}$$

Among them, $\hat{}$ represents the discrete Fourier transform operation corresponding to the physical quantity, the full text is the same; the size of $\hat{x}_d$ is $D_m \times D_n$. From Eq. (3) the value of the i-th element of $\hat{y}$ depend on the i-th element of samples $\hat{x}$ and the filter $\hat{\phi}$ in all chanels. So let $R_i(\bullet) \in \mathbb{R}^{D\times 1}$ denote the value of the i-th element in D chanels, and solve the Eq. (3) into $D_m \times D_n$ sub-problems:

$$\begin{aligned} \mathrm{L}\left(R_i(\hat{\phi})\right) = \left\|R_i^{(\mathrm{T})}(\hat{x})R_i(\hat{\phi}) - \hat{y}_i\right\|^2 \\ + \gamma\left\|R_i(\hat{\phi}) - R_i(\hat{\psi}) + R_i(\hat{\tau})\right\|^2 \\ + \mu\left\|R_i(\hat{\phi}) - R_i\left(\hat{\phi}_{t-1}\right)\right\|^2 \end{aligned} \tag{4}$$

Therefore, the closed-form solution of $R_i(\hat{\phi})$ is:

$$\begin{aligned} R_i(\hat{\phi}) = \frac{1}{\mu+\gamma}\left(\mathbf{I} - \frac{R_i(\hat{x})R_i^{(T)}(\hat{x})}{\mu+\gamma+R_i^{(T)}(\hat{x})R_i(\hat{x})}\right) \\ \times\left(R_i(\hat{x})\hat{y}_i + \gamma R_i(\hat{\psi}) - \gamma R_i(\hat{\tau}) + \mu R_i\left(\hat{\phi}_{t-1}\right)\right) \end{aligned} \tag{5}$$

Finally, $\hat{\psi}$ is performed the inverse discrete Fourier transform to obtain the solution of ϕ.

Question 2: solve ψ It can be solved directly in the time domain, and the objective function can be obtained from Eq. (3):

$$\mathrm{L}(\psi) = \|\operatorname{diag}(\boldsymbol{\omega}) \odot \boldsymbol{\psi}\|^2 + \gamma\|\phi - \boldsymbol{\psi} + \tau\|^2 \tag{6}$$

Among them, $\boldsymbol{\omega} \in \mathbb{R}^{DD_{\mathrm{m}}D_{\mathrm{n}}\times DD_{\mathrm{m}}D_{\mathrm{n}}}$ represents the spatial regularization weight matrix of D channels; $\boldsymbol{\psi} \in \mathbb{R}^{DD_{\mathrm{m}}D_{\mathrm{n}}\times 1}$ represents the ψ superimposed and vectorized matrix of channels. The solution that can be obtained is:

$$\psi = \left(\boldsymbol{\omega}^{(\mathrm{T})}\boldsymbol{\omega} + \gamma\mathbf{I}\right)^{-1} \odot (\gamma\phi + \gamma\tau) \tag{7}$$

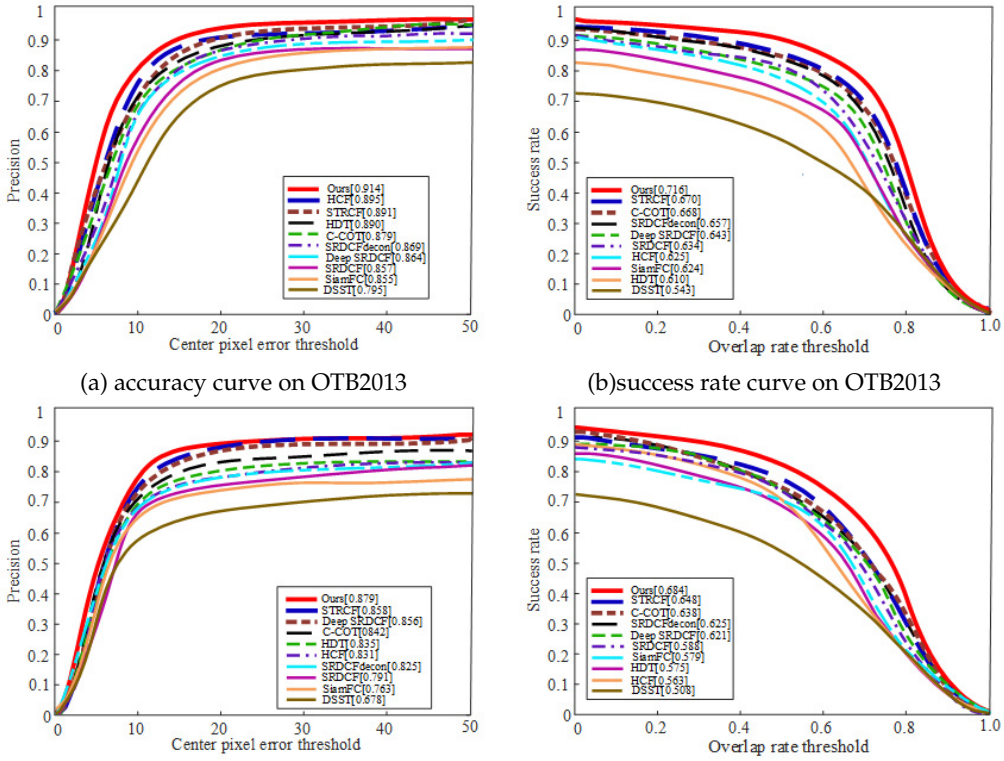
Question 3: The update method of $\hat{\tau}$ is

$$\hat{\tau}_{j+1} = \hat{\tau}_j + \hat{\phi}_{j+1} - \hat{\psi}_{j+1} \tag{8}$$

Among them, $\hat{\phi}_{j+1}$ and $\hat{\psi}_{j+1}$ are the values obtained by iterations of Question 1 and Question 2. Under normal circumstances, the update scheme of γ is shown in Eq. (9)

$$\gamma_{j+1} = \min\left[\gamma_{\max}, \zeta\gamma_j\right] \tag{9}$$

Among them, γ_{max} is the maximum value of γ, ζ is the scale factor.



(a) accuracy curve on OTB2013 (b)success rate curve on OTB2013

(c)accuracy curve on OTB2015 (d)success rate curve on OTB2015

Fig. 2. Comparison curve of distance accuracy and success rate on the OTB2013 and OTB2015 datasets

3. Optimal solution of spatial regularization weight ω

3.1. Extraction of invading foreign object space area

In order to fully explore the good representation and location ability of the depth feature to the railway foreign object, the spatial range of foreign object is extracted by using the high level feature. On the one hand, the depth features extracted from the pre-training model are used to build the appearance model of foreign object and on the other hand, they are used to constrain the regularization guide weight.

The fourth layer feature of the VGG network is extracted for foreign object location, and the fifth layer feature x is extracted for extracting spatial regions, which can be composed of $D_m \times D_n$ D-dimensional vectors and expressed as $x \in \mathbb{R}D_{\mathrm{m}} \times D_{\mathrm{n}} \times D$. Each vector is called a depth descriptor [13] and denoted by h.

3.2. Solution of the guidance weights ω_h of the depth space regularization

In this paper, starting from the first image of the object, it selects consecutive M frames for collaborative processing, and uses the space indicator matrix with the depth feature information of the foreign object to initialize ω_h.Principal component analysis (PCA) algorithm is used to perform correlation analysis on the depth feature descriptor $(H_{1,\ldots,}H_m, \ldots, H_M)$ of M frames continuous, where, $H_m = \left\{h_{m,(l,v)} \in \mathbb{R}^D\right\}$, $l = 1, 2, \ldots, D_m, v = 1, 2, \ldots, D_n$ The mean vector and covariance matrix of all descriptors are shown in Eq. (10) and Eq. (11).

$$\bar{h} = \frac{1}{K}\sum_{m}\sum_{l,v} h_{m,(l,v)} \tag{10}$$

$$\mathrm{Cov}(h) = \frac{1}{K}\sum_{m}\sum_{l,v}\left(h_{m,(l,v)} - \bar{h}\right)\left(h_{m,(l,v)} - \bar{h}\right)^{(\mathrm{T})} \tag{11}$$

Where, $K = D_m \times D_n \times M$ Then calculate the eigenvalue of the covariance matrix as $\lambda_l \geq \ldots \geq \lambda_z \geq 0$ and the corresponding to the eigenvector as $\zeta_l, \ldots, \zeta_z$ Since the first principal component has the largest variance, the feature vector ζ_l is used as the main projection direction. In a certain frame, calculate the principal component $q_{(l,v)} = \xi_1^{(\mathrm{T})}\left(h_{(l,v)} - \bar{h}\right)$ of the spatial depth descriptor at a certain position. Then $q_{(l,v)}$ of the frame is combined into a two-dimensional matrix $\mathbf{q}$ with dimension $D_m \times D_n$ named as the space indicator matrix, as shown in Eq. (12).

Table 2. Performance comparison of tracking methods based on deep learning

	Ours	DeepSRDCF	HDT	HCF	SiamFC
accuracy	**0.879**	0.858	0.835	0.831	0.763
success rate	**0.684**	0.621	0.575	0.563	0.579
speed (FPS)	21.7	0.2	2.8	10.7	**82.5**

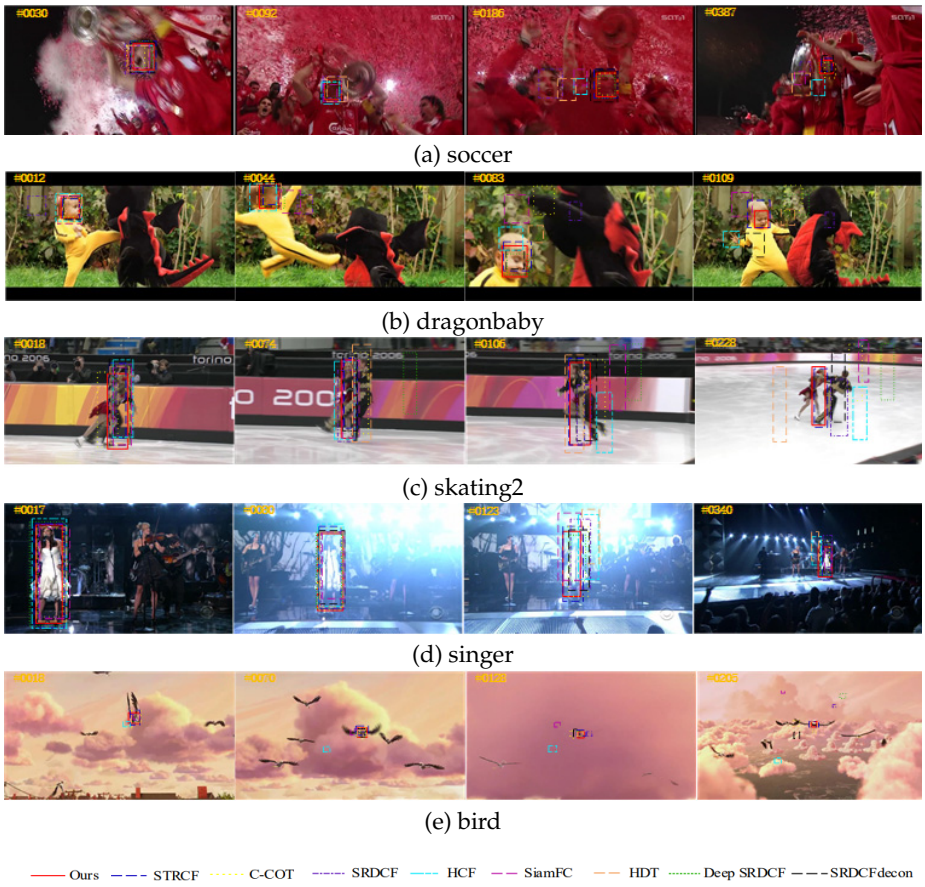


Fig. 3. Comparison of the tracking results of different algorithms on the OTB datasets

$$\mathbf{q} = \begin{bmatrix} q_{(1,1)} & q_{(1,2)} & \cdots & q_{(1,D_n)} \\ q_{(2,1)} & q_{(2,2)} & \cdots & q_{(2,D_n)} \\ \vdots & \vdots & \ddots & \vdots \\ q_{(D_m,1)} & q_{(D_m,2)} & \cdots & q_{(D_m,D_n)} \end{bmatrix} \tag{12}$$

Since ζ_1, is obtained on multiple frames of images, in the $\mathbf{q}$ matrix, when the value of $q_{(l,v)}$ is a positive, it means that there is a positive correlation between the depth descriptors. The positively correlated area is the intrusion foreign object area. Similarly, when the value of $q_{(l,v)}$ is a negative, it means that there is a negative correlation between the depth descriptors, and the negative correlation area is the background area. The larger the absolute value of $q_{(l,v)}$ the higher the correlation. In addition, if there is no positive value in the $\mathbf{q}$ matrix almost, it means that the target is occluded or lost, which is used as a criterion for model update.

The positive value in the matrix $\mathbf{q}$ is set to 1, and the negative value area is set to 0, and it is named the weight indicator matrix, that is $\mathbf{q}_{(l,v)}$, as shown in Eq. (13).

$$\mathbf{q}_{(l,v)} = \begin{cases} 1, q_{(l,v)} > 0 \\ 0, \text{ otherwise} \end{cases} \tag{13}$$

Calculate the mean value of all elements of the weight indicator matrix $\mathbf{q}_{(l,v)}$, as shown in Eq. (14), then $\overline{\mathrm{q}} \in [0,1]$. The closer $\overline{\mathrm{q}}$ is to 1, the closer it is to the middle area of the target, on the contrary, the closer it is to the edge area of the target.

Table 3. Scene description

scenes	description	frames
v1	complex background, obstructed by power distribution equipment	800
v2	scale change, rapid movement, deformation, light change, shade shading	800
v3	light weak, scale change, sandstorm weather, fuzzy appearance	600
v4	scale changes, lighting changes, weed occlusion, complex backgrounds	1000

$$\overline{\mathrm{q}} = \frac{1}{D_m \times D_n} \sum_{l,v} \mathrm{q}_{(l,v)} \tag{14}$$

Initialize the regularized guidance weight ω_h, namely

$$\omega_h = \omega_0 \odot \frac{1}{1 + \vartheta \cdot \overline{\mathrm{q}}} \tag{15}$$

Among them, ϑ is the fixed parameter, and ω_0 is the original penalty weight. In this way, the spatial regularization guide weight ω_h brings the content information of the invading foreign object on the basis of ω_0, which can better punish the background.

3.3. Solving the spatial regularization weight

In order to speed up the convergence speed, the ADMM method is still used to solve the spatial regularization weight ω. First, the spatial regularization weight ω is vectorized as $\omega^{'}$, then the objective function is:

$$\mathrm{L}\left(\omega'\right) = \sum_{d=1}^{D} \left\|\mathrm{diag}\left(\psi_k\right) \odot \omega'\right\|^2 + \eta \left\|\omega' - \omega_h\right\|^2 \tag{16}$$

Introducing the constraint condition $\omega^{'} = v$, the Lagrangian equation of the above formula becomes:

$$\begin{aligned} \mathrm{L}\left(\omega', v, \sigma\right) = &\sum_{d=1}^{D} \left\|\mathrm{diag}\left(\psi_k\right) \odot \omega'\right\|^2 \\ &+\eta \left\|u - \omega_h\right\|^2 + \delta \left\|\omega' - v + \sigma\right\|^2 \end{aligned} \tag{17}$$

Among them, δ is the penalty parameter; σ is the Lagrangian multiplier vector. The solution of the sub-problem is shown in Eq. (18).

$$\begin{cases} \omega' = \frac{v-\sigma}{\frac{1}{\delta}\sum_{d=1}^{D} \psi_k \odot \psi_k + \mathbf{I}} \\ v = \frac{\eta\omega_h + (\omega' + \sigma)\delta}{\eta + \delta} \\ \sigma_{j+1} = \sigma_j + \omega'_{j+1} - v_{j+1} \\ \delta_{j+1} = \min\left[\delta_{\max}, \rho\delta_j\right] \end{cases} \tag{18}$$

Substituting the solved spatial regularization weight into Eq. (7), solve the correlation filter ψ, so that the spatial regularization parameter can effectively suppress the boundary effect.

3.4. Algorithm flow

In summary of the description of the training process of the filter, the algorithm can be summarized as Table 1.

4. Model update strategy based on depth space

In the process of object tracking based on correlation filtering, the model update step is particularly critical. A good model update strategy can improve the algorithm's ability to process complex railway scenes, and effectively prevent tracking drift and even target loss.

In this paper, through the observation and analysis of the spatial indicator matrix **q** in Section 3.2, it can be seen that when the target is occluded or lost, most of the elements of **q** are negative. A model update strategy based on depth space is designed. Set the model update threshold th.

When the number of positive values N_q of the space indicator matrix **q** is greater than th , it indicates that the target is not occluded, and the model is updated with α as the update rate; When the number of positive values N_q of the space indicator matrix **q** is less than th, it indicates that the target is blocked or lost, and the update rate is reduced. After a large number of experiments in railway scenes, with $\frac{\alpha}{5}$ as the update rate, the tracking effect is the best. The model update strategy is shown in Eq. (19).

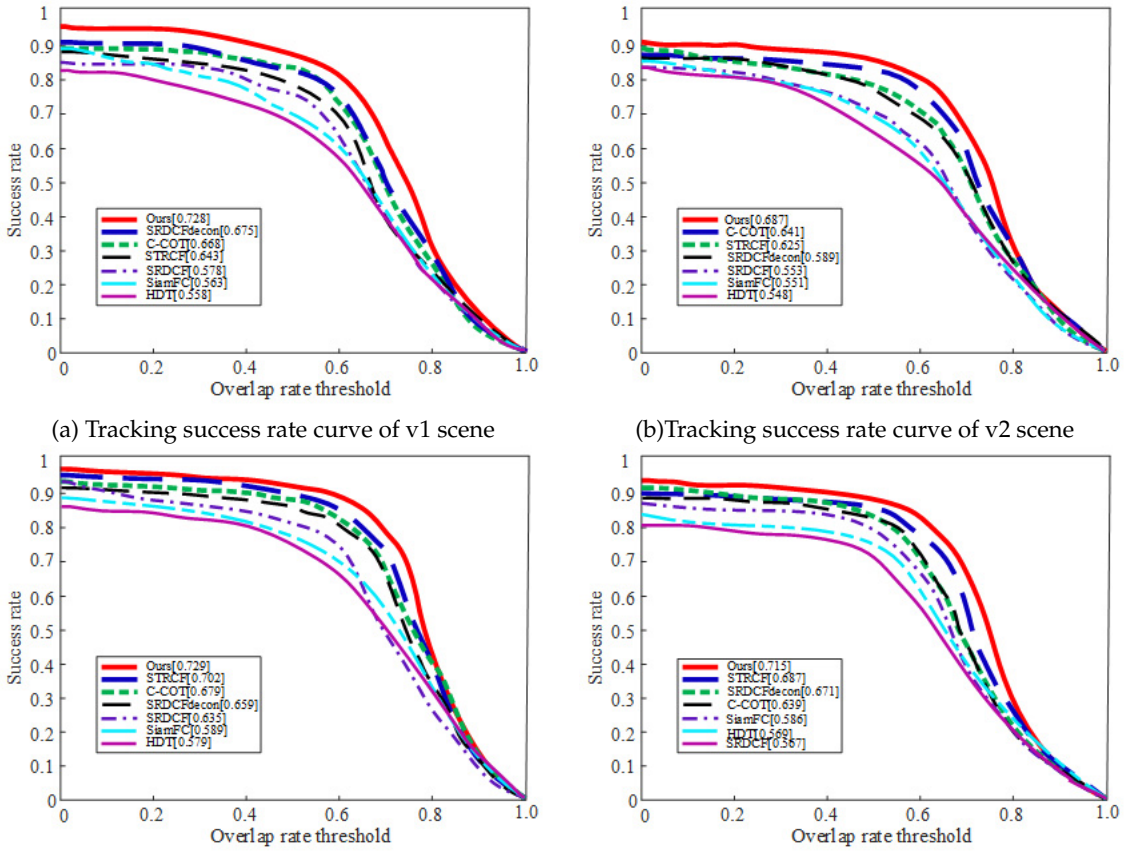
$$\phi_t = \begin{cases} (1-\alpha)\phi_{t-1} + \alpha\phi, N_{\mathrm{q}} > th \\ \left(1 - \frac{\alpha}{5}\right)\phi_{t-1} + \frac{\alpha}{5}\phi, N_{\mathrm{q}} < th \end{cases} \tag{19}$$

Among them, ϕ_t represents the filter template of the current frame; ϕ_{t-1} represents the filter template of the previous frame.

In summary, the flow chart of the algorithm in this paper is shown in the Fig. 1. The leftmost box is the video data along the railway, where the first M frames of images are used to initialize the spatial regularization weight. The two boxes in the middle are the initialization process and training process of the filter. The last box is the model update process and the tracking result display.

5. Experimental results and evaluation analysis

In order to fully verify the effectiveness of the algorithm in this paper, three indicators of distance accuracy, tracking success rate and tracking speed are used to compare the tracking performance of this algorithm and advanced target tracking algorithms on the OTB2013 [14] and OTB2015[15] datasets. Advanced target tracking algorithms are based on spatial regularization, such as SRDCF,



(a) Tracking success rate curve of v1 scene

(b)Tracking success rate curve of v2 scene

(c)Tracking success rate curve of v3 scene

(d) Tracking success rate curve of v4 scene

Fig. 4. Comparison curve of the success rate of different algorithms in v1∼v4 scenes

STRCF, SRDCFdecon [16], based on correlation filtering, suah as C-COT [17], DSST and based on deep learning , such as HCF [8] HDF [18], SiamFC [19], DeepSRDCF [20]. Then it is applied and verified in complex railway scenes.

Use MATLAB2018a programming, and use MatConvNet toolbox [21] to realize and train the forward propagation of the deep network. A computer with a main frequency of 2.4GHz and a memory of 64GB is used in the experiment . The GPU (1080Ti) is used for acceleration. In the experiment, the 3 times of the target area is used as the search area. The time regularization parameters and spatial regularization parameters are μ=12, η=0.1 respectively, and fixed parameter is $\vartheta = 2.5$ In the process of solving the correlation filter, the initial value, maximum value and scale factor of the penalty parameter are γ_0=1, γ_{max} =10 and ρ=5. The iteration times of ADMM are all twice, and the rest of the parameter settings refer to SRDCF. Model update threshold is th=10.

5.1. Experiments and evaluation analysis on the OTB datasets

The OTB datasets includes scenes in various situations such as fast motion, low resolution, motion blur, image blur, target occlusion, lighting changes, non-rigid deformation, in-plane rotation, and complex backgrounds.Fig. 2(a)-(d) show the distance accuracy and success rate of the 9 algorithms in the OTB2013 datasets and OTB2015 datasets, respectively. The accuracy and success rate of the algorithm in this paper are better than methods such as STRCF and SRDCF on the OTB datasets.

The accuracy and success rate on the OTB2013 datasets are 0.914 and 0.716, respectively. Compared with the DeepSRDCF algorithm, which is also based on depth features and spatial regularization, the accuracy on both datasets is improved by two percentage points, and the success rate is increased by six percentage points. Compared with the CNN-based method, the method in this paper is significantly better than the HCF and HDT algorithms, which based on multi-layer deep feature fusion on the OTB2015 data set. Compared with SiamFC based on Siamese net-

work, it also has obvious advantages.

Table 2 shows the accuracy, success rate and speed comparison of several tracking methods based on deep learning on the OTB2015 datasets. SiamFC has the best tracking speed, which is 82.5FPS. The tracking speed of this algorithm is 21.7FPS, but it is far better than SiamFC algorithm in distance accuracy and success rate. At the same time, compared with DeepSRDCF, this paper still has obvious advantages in speed.

Fig. 3(a)-(e) show the tracking results of this algorithm and other 9 algorithms in some typical video sequences. The main attribute of soccer is the complex background. The algorithm in this paper accurately punishes the background area and effectively solves the tracking problem of the model under the complex background. The bird, skating2 and soccer all have target occlusion phenomenon, especially the main difficulty of bird video sequence is the background occlusion, which leads to the tracking failure of many algorithms. From the 126th frame to the 196th frame of the bird video sequence, the target is blocked by clouds to varying degrees.

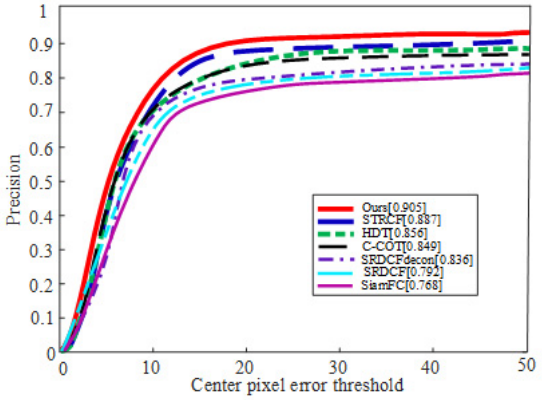
The algorithm in this paper can still track the target effectively. This shows that the temporal perception item considers the continuous change of the target in each frame, so that the filter can changed smoothly following the object. At the same time, the model update strategy can effectively avoid the problem of model degradation. Almost all video sequences have scale changes, especially dragonbaby video sequences have obvious scale changes. From the 26th frame to the 86th frame, the target increases from small to large, and after the 86th frame, it changes from large to small. The algorithm in this paper also considers the influence of scale while suppressing the filter., which can accurately and effectively track. The main attribute of the singer video sequence is the change of illumination, and the algorithm in this paper has a good ability to cope with it.

5.2. Experiments and analysis in complex railway scenes

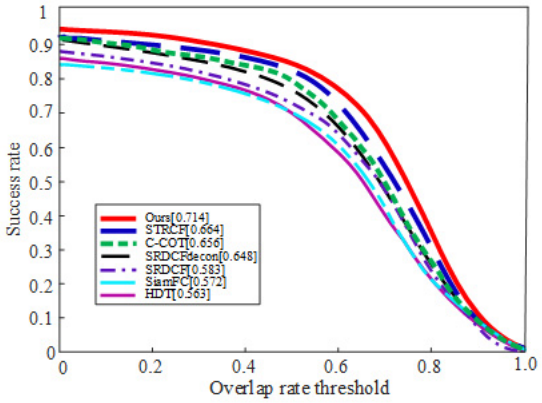
A camera with a frame rate of 31 frames/s is used to collect 4 sets of video sequences along a certain railway track, which are v1∼v4, as shown in Table 3.

According to the tracking results of the OTB datasets, several algorithms with relatively high accuracy, success rate and shorter tracking time are selected: SRDCF, STRCF, SRDCFdecon, SiamFC, C-COT and HDT. The algorithm in this paper is compared with the above 6 algorithms in v1 v4 scenes. The comparison chart of the tracking success rate curves of different tracking algorithms in the four scenes is shown in Fig. 4(a)-(d).

It can be seen from Fig. 4 that in a variety of complex



(a) accuracy



(b)success rate

Fig. 5. Comparison curve of average accuracy and average success rate

scenes, the tracking success rate of the algorithm in this paper ranks first. Only in the v2 scene, due to long-term occlusion, the success rate of this paper is lower than other scenes.

Fig. 5(a) and Fig. 5(b) respectively show the average distance accuracy and average tracking success rate of the 7 algorithms in v1 v4 video sequences. The accuracy and success rate of this paper are 0.905 and 0.714 respectively, which show obvious advantages compared with other algorithms. The average tracking time of the algorithm in this paper is 23.1 FPS, which basically meets the real-time requirements.

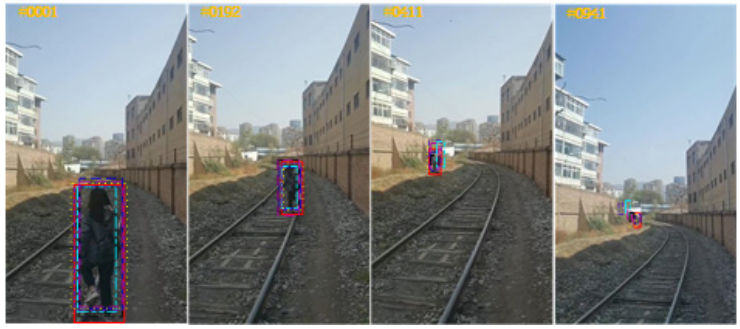
In four scenes with different tracking algorithms, the tracking results of some frames are shown in Fig. 6.

In the four scenes, almost all have the attributes of complex backgrounds. Because the algorithm in this paper initializes the regularization guide weights with object's depth space information, adaptively calculates the spatial regularization weight, and effectively overcomes the influence of complex backgrounds. Scenes v1, v2, and v4 all

(a) Tracking result of v1 scene

(b) Tracking result of v2 scene

(c) Tracking result of v3 scene



(d) Tracking result of v4 scene

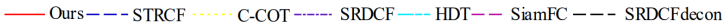


Fig. 6. Comparison of tracking results of different algorithms in v1∼v4

have foreign object occlusion, especially v2 from the 335th frame to the 410th frame, the target is severely occluded after moving into the shade. Until frame 415, the target reappeared. Due to the suppression of model degradation by the time perception term and the model update strategy based on depth space, only the algorithm in this paper successfully tracks the invading foreign objects again.

6. Conclusion

In the complex railway scenes, the existing regularization correlation filtering algorithms have some defects, such as limited learning ability of spatial regularization weight and model degradation, etc. Adding a deep space regularization term to the regularization model, the algorithm's ability to cope with complex backgrounds is significantly enhanced . Combined with the time perception item, the algorithm's robustness in scenes such as foreign body occlusion and foreign body deformation is effectively improved.

Besides, the model update strategy based on depth space information improves the long-term tracking performance of the algorithm. The experimental results show that in the OTB datasets and complex railway scenes, the algorithm in this paper has obvious advantages in distance accuracy and success rate under the premise of satisfying real-time performance. The next research will consider the relationship between time perception items and the characteristics of foreign objects, and improve the ability to recognize foreign objects invading railway.

Acknowledgements

Granted by the Natural Science Foundation of Gansu Province (Grant No.: 1606RJZA002); Research Funded by Universities and Colleges in Gansu Province (2017a -026); Funded by the Hundred Talents Training Program of Lanzhou Jiaotong University (2018-103).

Schedule

References

[1] Y. Wang. *"Gaosu tielu changjing fenge yu shibie suanfa yanjiu [Research on high-speed railway scene segmentation and recognition algorithm]"*. (phdthesis). Beijing, China, 2019.

[2] X. Li, L. Zhu, and Z. Yu, (2020) *"Zishiying tielu changjing qianjing mubiao jiance [Adaptive railway scene foreground target detection]"* **Jiaotong yunshu xitong gongcheng yu xinxi** ***v.20***(02): 87–94.

[3] T. Hou, H. Wu, and H. Niu, (2020) *"Gaijin MOG-LRMF de tiegui dongtai yiwu jiance [Real-time detection of rail dynamic foreign object intrusion based on improved MOG-LRMF]"* **Jiaotong yunshu xitong gongcheng yu xinxi** (2): 91–100.

[4] H. Shi, H. Chai, Y. Wang, and Z. Yu, (2015) *"Jiyu mubiao shibie yu genzong de qianrushi tielu yiwu qinxian jiance suanfa yanjiu [Research on Embedded Railway Foreign Body Intrusion Detection Algorithm Based on Target Recognition and Tracking]"* **Tiedao xuebao** ***37***(7): 58–65.

[5] Z. Qu, R. Zhou, X. Sun, S. Yuan, and L. Zou, (2019) *"Chidu zishiying de tielu yiwu qinxian PSA-Kcf jiangwei genzong fangfa [Scale adaptive PSA-Kcf dimension reduction tracking method foreign body intrusion]"* **Tiedao xuebao** ***041***(005): 71–81.

[6] H. Wu. *"Tielu guidao yiwu ruqin de zhineng shibie ji zidong yujing yanjiu [Research on intelligent recognition and automatic early warning of foreign object intrusion on railway track]"*. (phdthesis). Lanzhou, China., 2020.

[7] R. Zhou. *"Jiyu jiqi shijue de tielu yiwu qinxian lubangxing jiance ji genzong fangfa yanjiu [Research on Robustness Detection and Tracking Method of Railway Foreign Body Intrusion Limit Based on Machine Vision]"*. (phdthesis). Nanchang, China.

[8] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang. "Hierarchical convolutional features for visual tracking". In: *Proceedings of the IEEE international conference on computer vision*. 2015, 3074–3082.

[9] J. Dai and N. Yan. "Robust Single-object Visual Tracking Framework via Fully Convolutional Siamese Network with Correlation Filter". In: *2020 13th International Symposium on Computational Intelligence and Design (ISCID)*. IEEE. 2020, 359–363.

[10] H. Fu, Y. Zhang, W. Zhou, X. Wang, and H. Zhang, (2020) *"Learning reliable-spatial and spatial-variation regularization correlation filters for visual tracking"* **Image and Vision Computing** ***94***: 103869.

[11] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li. "Visual tracking via adaptive spatially-regularized correlation filters". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, 4670–4679.

[12] D. Elayaperumal and Y. H. Joo, (2020) *"Visual object tracking using sparse context-aware spatio-temporal correlation filter"* **Journal of Visual Communication and Image Representation** ***70***: 102820.

Table 4. Symbol explanation table

notation	Meaning of notation	The meaning of subscripts and superscripts
D	the total number of feature channels	nothing
x_d	the feature of the channel d	x is deep space features, which is composed of $D_m \times D_n$, D-dimensional vectors; d is the number of the characteristic channel
ϕ_d	the filter of the channel d	d is the number of the characteristic channel
y	the desired output	nothing
ω_h	the regularization guide weight with the depth space information	h means deep space
ω	spatial regularization weight	nothing
η	the spatial regularization parameter	nothing
μ	the time perception parameter	nothing
γ	penalty parameter	nothing
ψ	auxiliary variable	nothing
τ	the Lagrangian multiplier	nothing
$\hat{}$	the discrete Fourier transform operation	nothing
$R_i(\bullet) \in \mathbb{R}^{D\times 1}$	the value of the i-th element in D	i is the number of feature channel element
$\boldsymbol{\omega} \in \mathbb{R}^{DD_{\mathrm{m}}D_{\mathrm{n}}\times DD_{\mathrm{m}}D_{\mathrm{n}}}$	the spatial regularization weight matrix of D channels	nothing
$\boldsymbol{\psi} \in \mathbb{R}^{DD_{\mathrm{m}}D_{\mathrm{n}}\times 1}$	the ψ superimposed and vectorized matrix of D channels	nothing
γ_{max}	the maximum value of γ	nothing
ζ	the scale factor	nothing
$\hat{\tau}_j$	the value of the j-th iteration of the Lagrange multiplier in the Fourier domain	j is the number of iterations, the full text is the same
H_m	depth descriptor of the m-th frame image	m is the number of the previous M frames
$h_{m,(l,v)}$	depth descriptor at position (l,v) in the m-th	(l,v) is the position in the image
$\mathbf{q}$	the space indicator matrix	nothing
$\lambda_1, ..., \lambda_z$	the eigenvalue of the covariance matrix	the subscript indicates the eigenvalue number
$\zeta_1, ..., \zeta_2$	the corresponding to the eigenvector	the subscript indicates the feature vector number
K	The total number of all positions of the M frames image	nothing
$q_{(l,v)}$	The principal component of the spatial depth descriptor at position (l,v)	(l,v) is the position in the image
ϑ	the fixed parameter when initializing ω_h	nothing
ω_0	the original penalty weight	nothing
$\omega^{'}$	the spatial regularization weight ω is vectorized	nothing
v	auxiliary variable	nothing
δ	the penalty parameter	nothing
σ	the Lagrangian multiplier vector	nothing
N_q	the number of positive values of the space indicator matrix $\mathbf{q}$	the subscript indicates the space indicator matrix
ϕ_t	the filter template of the current frame	t represents the current frame, the full text is the same
ϕ_{t-1}	the filter template of the previous frame	t-1 represents the previous frame, the full text is the same
α	the update rate	nothing
th	the model update threshold	nothing
superscript (T)	nothing	matrix transpose operation, the full text is the same

[13] X.-S. Wei, J.-H. Luo, J. Wu, and Z.-H. Zhou, (2017) *"Selective convolutional descriptor aggregation for fine-grained image retrieval"* **IEEE Transactions on Image Processing *26***(6): 2868–2881.

[14] Y. Wu, J. Lim, and M.-H. Yang. "Online object tracking: A benchmark". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013, 2411–2418.

[15] Y. Wu, J. Lim, and M.-H. Yang. "Object tracking benchmark". In: *Proceedings of the IEEE Transactions on Pattern Anaiysis and Machine Intelligence*. 2015, 1834–1848.

[16] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg. "Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, 1430–1438.

[17] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg. "Beyond correlation filters: Learning continuous convolution operators for visual tracking". In: *European conference on computer vision*. Springer. 2016, 472–488.

[18] Y. Qi, S. Zhang, L. Qin, H. Yao, Q. Huang, J. Lim, and M.-H. Yang. "Hedged deep tracking". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, 4303–4311.

[19] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr. "Fully-convolutional siamese networks for object tracking". In: *European conference on computer vision*. Springer. 2016, 850–865.

[20] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg. "Convolutional features for correlation filter based visual tracking". In: *Proceedings of the IEEE international conference on computer vision workshops*. 2015, 58–66.

[21] A. Vedaldi and K. Lenc. "Matconvnet: Convolutional neural networks for matlab". In: *Proceedings of the 23rd ACM international conference on Multimedia*. 2015, 689–692.