IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 31, 2019, accepted August 24, 2019, date of publication September 4, 2019, date of current version September 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939435

Scale Self-Adaption Tracking Method of Defog-PSA-Kcf Defogging and Dimensionality Reduction of Foreign Matter Intrusion Along Railway Lines

ZHIJIAN QU[1], WANZHUO YI[1], RUILIN ZHOU[1,2], HANLING WANG[1], AND RUI CHI[1]
[1]School of Electrical Engineering, East China Jiaotong University, Nanchang 330013, China
[2]CRRC Times Signal & Communication Co., Ltd., Changsha 410199, China

Corresponding author: Zhijian Qu (08117324@bjtu.edu.cn)

This work was supported in part by the Foundation Plan for Distinguished Young Scholars in Jiangxi Province under Grant 20162BCB23045, and in part by the Applied and Cultivation Program of Science and Technology Deparment of Jiangxi Province under Grant 20181BBE58010 and 2019BBEL50006.

ABSTRACT In foggy weather, the occlusion of smoke dust and the variation of imaging scale of foreign objects will mislead the existing algorithm to learn the wrong target information, causing the drift of the tracking box. A kernel correlation filtering target tracking dimensionality reduction algorithm combining dark channel prior and scale estimation is proposed, effectively improve target tracking accuracy in foggy weather. Firstly, in the process of detecting intruding foreign objects along the railway with visual background extractor ViBe, the atmospheric scattering model based on dark channel prior was used to defog video sequence. After that, FHOG features of the initial tracking box were extracted by dense cyclic sampling and scale pyramid technology, and then a Kcf position filter, as well as a scaling filter with PCA dimensionality reduction was trained to realize the scale-adaptive rapid tracking of railway foreign objects in foggy weather. The experimental results show that in terms of tracking accuracy, the proposed Defog-PSA-Kcf algorithm is superior to the non-scale estimation link the generation algorithm Mean Shift devoid of scale estimation and the native Kernel Correlation Filter algorithm (Kcf), and while linear kernel algorithm Dual Correlation Filter algorithm (Dcf), which is higher than the scale-adaptive SA-Kcf and the SAMF algorithm; in terms of tracking speed, it's faster than the Mean Shift, SA-Kcf, and SAMF algorithms, and in tracking speed. The algorithm can achieve fast tracking results comparable to as fast as the Kcf algorithm.

INDEX TERMS Foggy weather, foreign object intrusion, dark channel prior, FHOG features, kernelized correlation filter, PCA dimensionality reduction.

I. INTRODUCTION

Railway clearance is the contour scale line which can not be crossed, used to separate buildings, equipment and rolling stocks so as to ensure the safety of the railway transportation. In addition, it also serves as a barrier to prevent pedestrians, vehicles, animals and foreign objects from intruding into the railway. In most cases, protective nets are used to protect the railway clearance. However, foreign object intrusion still occasionally happens at the junction of driveways and sidewalks when there is a lack or damage of protective nets, or when pedestrians and vehicles cross them. The rapid development of industry has brought highly dense social and economic activities, but also global warming when most parts of the world are affected by severe weathers such as heavy fog, rains and snows. Especially in heavy foggy weather, fog can absorb and scatter light, which will reduce the visibility. The video or images collected and captured will also have low saturation, low contrast and other characteristics, thereby worsening the visibility conditions of train drivers and affecting the imaging quality of integrated video surveillance system used in the railways. It is difficult to identify and track foreign objects intruding the railway clearance on a foggy day with the existing approaches, which therefore

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaoqing Pan.

will greatly affect the safety of railway transportation because train drivers often cannot get early warning of such foreign object intrusion. Therefore, the author strives to put forward a method of removing fog and reducing dimensionality for detecting and tracking foreign object intrusion which often happens in severe weather, especially in foggy weather.

The existing methods for detecting and tracking foreign object intrusion on railways can be divided into two categories: the sensor-based detection and tracking [1], as well as the machine vision-based detection and tracking. The former can detect and warn foreign objects intruding the railway clearance by using laser, microwave, ultrasonic or infrared sensors installed along the railway lines. However, it's susceptible to strong electromagnetic interference in electrified sections, and therefore not suitable for the detection and tracking of foreign object intrusion.

The machine vision-based detection and tracking method [2] is more intelligent and more suitable for the detection and tracking of foreign object intrusion as it integrates the machine learning into the image processing. This method can learn and predict unstructured data via a generative model or a discriminative model. The generative model exhibits the distribution of data from a statistical perspective, and can estimate the joint probability distribution. It can be obtained by enhanced learning and can be used even when the data are incomplete. Reference [3], [4] proposed a Mean Shift pedestrian tracking method combining target color and edge feature histograms, which can reduce the influence of local occlusion on the algorithm. However, the generative model has higher requirements on the data distribution assumption. If the data distribution does not meet the assumption, the model will exhibit poor robustness and inferior performance compared to that of the discriminative model [5]. The discriminative model can find the optimal separating hyperplane between different categories to reflect the difference of different categories of data and estimate the conditional probability distribution. The model can clearly distinguish the discriminative features of multiple categories, or between one category and another category, so it is more suitable for identifying multiple categories and simpler to use [6], [7]. At present, the discriminative tracking method based on deep learning or correlation filtering [8] for model training is a hot research topic in the target tracking field. As for the former, MDNet Convolutional Neural Network (CNN) [9] is a case in point. It uses several tracking image sequences to train the overall performance of the target in the shared layer and reconstructs the tracking network by connecting the branch layer of the image sequence to the shared layer, with higher algorithm tracking accuracy. However, this method features poor real-time performance because it requires thousands of overlapping candidate areas to be entered the CNN network one by one for calculation, about 0.5s for one frame. For the latter, the MOSSE [10] correlation filter was developed by David S. Bolme and others to improve the target tracking speed. The MOSSE filter resorts to a self-adaptive training method to introduce a correlation filter into the target tracking field, and the tracking speed of the algorithm is improved. As the correlation filter is a single-channel filter and only has manual features, its tracking performance is not ideal and its application scope is limited. Henriques *et al.* [11] greatly improved the target tracking accuracy by fully extracting the target information through intensive cyclic sampling. However, due to the lack of necessary scale estimation, the algorithm is apt to learn excessive background information or local texture information due to the change in the imaging size of the intruding foreign objects, which causes the drift of tracking box and affects the tracking accuracy of the target.

In severe weather, especially on foggy days, the tracking accuracy of the generative model is not quite ideal because the background information is not distinguished in the modeling process. Discriminant models can distinguish between foreground and background information. However, due to the influence of fog, the low saturation and low contrast characteristics of video and images lead to blurring of the tracking target, resulting in reduced tracking accuracy. However, most of the defogging algorithms studied by scholars both at home and abroad are based on the defogging of a single image [12]–[17]. For example, Tan [18] proposed a fog removal method based on local contrast maximization. As the contrast of the fog-free image is higher than that of the foggy image, it constructs the function of image edge information by assuming that the local ambient light is a constant and restores the local contrast of the image to achieve a defogging effect. Still, the problem is that over-saturated local color of fog-free image will lead to image distortion. Tarel and Hautière[19] assumed that the atmospheric dissipation function changes slowly in the local part, and tried to achieve a defogging effect by estimating the atmospheric dissipation function with median filter. However, halo phenomenon may occur at the edge after fog removal with such algorithm. In severe weather, the target tracking algorithms above affect the tracking accuracy of intruding foreign matters on railway, and the defogging algorithms above are mostly based on single image defogging. How to ensure the tracking speed level of foreign matters in heavy fog and improve the tracking accuracy have not been reported in the literature yet.

In view of the situation above, this paper proposes a detection and tracking method of Defog-PSA-Kcf dimensionality reduction based on dark channel prior and kernel correlation filtering for scale-adaptive foreign matter intrusion on railway. The dimensionality reduction detection and tracking algorithm uses dark channel prior to defog and uses the special application scenario of railway integrated video monitoring system to improve the estimated value of transmittance. After the defogging is realized by optimizing the transmittance, the visual background extractor ViBe [20], [21] is applied to the moving target detection stage in the video sequence in order to more accurately capture the intruding foreign matter. After that, foreign matter scale estimation is added to the original kernel correlation filter framework, which effectively improves the tracking accuracy

(a) the 160th frame image (b) the 327th frame image

FIGURE 1. Dark channel image.

of the intruding foreign matter with the change of scale. Principal component analysis (PCA) is used to greatly reduce the dimensionality of scale filter features, reduce the calculation time of the added scale estimation, and improve the tracking speed of the intruding foreign matter. The accuracy and real-time performance of the proposed algorithm in foggy, snowy and snowy weather are verified by several sets of comparative experiments.

II. FOG REMOVAL ALGORITHM BASED ON DARK CHANNEL PRIOR

A. THEORY OF DARK CHANNEL PRIOR

Studies on a large number of fog-free images indicate that in the vast majority of images, most regions except non-sky regions have relatively low color value of RGB channels compared with that in other regions. In others words, the light intensity of such regions is lower than that in other areas. Therefore, the dark channel of any fog-free image can be defined mathematically as [17]:

$$M^{dark}(x) = \min_{y\in\Omega(x)} [\min_{c\in\{r,g,b\}} M^c(y)] \tag{1}$$

where, M^c in Equation (1) represents each channel of a color image. $\Omega(x)$ is a window centered on pixel X; c belongs to R, G and B color channels. $M^{dark}(x)$ is the dark channel of the image.

The meaning of Equation (1) is to firstly find the minimum value of each channel pixel in the color image, then put it into a gray image of the same size as the original fog-free image, and finally figure out the minimum value filter of the gray image.

According to the dark channel prior theory, in the non-sky area, the dark channel of the fog-free image has a low brightness value, which is almost zero.

In real life, there will be low values of dark primary colors in the image, such as leaves, buildings, cars, shadows caused by the object blocking the spread of light, making the shadows a lower value of the dark primary colors in the image. Fig.1 (a) is a dark channel image in the 160th frame, Fig.1 (b) is a dark channel image in the 327th frame.

1) ATMOSPHERIC SCATTERING MODEL

After the light is scattered by atmospheric particles, the observed light intensity effect can be expressed by the atmospheric scattering model proposed [22], [23]:

$$I(x) = M(\mathrm{x})t(x) + A(1 - t(x)) \tag{2}$$

In Equation (2), x means the coordinates of the pixels in the image; M(x) refers to the fog-free image to be restored, which is unknown; I(x) denotes the existing image to be defogged, which is known; t(x) is the transmittance of the image, and A represents the global atmospheric light component. The collected light are divided into two categories in this model: a. light that reaches the camera after atmospheric light scattering in the environment; b. light that reaches the camera after being reflected and decayed by atmospheric particles.

In addition, to solve the problem that the rain and snow weather may affect the accuracy of the algorithm tracking. We use a median filtering and etching operation to preprocess the image frame to eliminate the noise that raindrops and snowflakes form on the binary image. The specific content is as follows: the pixel value of each pixel is replaced with the median value of the domain pixel value to filter out the salt and pepper noise in the binary image, and then the isolated point of the target edge is removed by the etching operation to achieve the effect of shrinking the target size.

B. FOG REMOVAL PRINCIPLE OF DARK CHANNEL PRIOR

1) ESTIMATION OF TRANSMITTANCE

The transmittance t(x) has consistency in local areas, the estimated transmittance is defined as $\tilde{t}(x)$. Assuming that A is a known constant and the transmittance of a certain area on the image remains unchanged.

The visual effect of the video shot in the railway integrated video monitoring system is relatively dark, so that the intrusion foreign matter and the background cannot be distinguished well. In order to ensure that the restored image is close to the real scene, an image fidelity adjustment factor $w(0 \le w \le 1)$ is introduced. w_o is taken as the best fidelity adjustment factor for typical images in this application scenario.

Then, the estimated value of transmittance according to the dark channel prior theory is shown in equation (3):

$$\tilde{t}(\mathrm{x}) = 1 - w_o \min_{y\in\Omega(\mathrm{x})} [\min_{c\in\{r,g,b\}} \frac{I^c(\mathrm{y})}{A^c}] \tag{3}$$

The estimated transmittance is obtained according to Equation (3). When the transmittance value is small, the value of M is large, which may lead to noise interference in M(x). Therefore, it is necessary to limit t(x) by setting a threshold. To prevent t(x) being too small, when t(x) is smaller than the threshold value, let $t(x) = 0.1$ according to the algorithm proposed in this paper.

2) IMAGE RESTORATION

The value of atmospheric light component A is also required after calculating the estimated transmittance according to Equation (3). The above algorithm is based on the assumption that A is a constant. However, in this application scenario, if the selected atmospheric light value A is larger than its

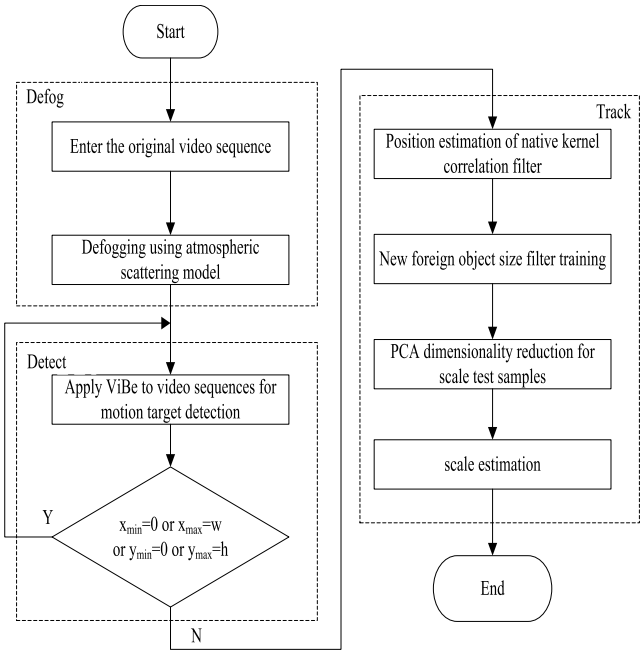


FIGURE 2. Flow diagram of detection and tracking of foreign object intrusion in foggy weather.

true value, the restored image will be darker and the image details will be lost. If the selected atmospheric light value A is less than the true value, the restored image will be brighter. In order to select an appropriate atmospheric light value, the pixels with brightness ranking top 0.1% can be selected in the dark channel image, and then the brightest point in foggy image I(x) can be selected as A value. After an appropriate A value is selected, the M(x) can be calculated to restore the fog-free image. The expression for the final restoration of the image is:

$$M(\mathrm{x}) = \frac{I(\mathrm{x}) - A}{\max(t(x), 0.1)} + A \tag{4}$$

III. PRINCIPLES FOR DETECTION OF FOREIGN OBJECT INTRUSION

The ViBe [20], [21] is applied to the moving target detection stage of video sequence to capture the intruding foreign objects. The scale estimation of new foreign objects is added to the genetic Kernel Correlation Filter algorithm framework to effectively improve the accuracy in tracking. The specific steps are shown in Fig.2.

In the initial processing stage, the first frame image in the video is used to establish a background sample set for each pixel. The sample set is composed of domain pixel values. Then compare the values of all pixels on images in the video, except the first frame, with those values in the established background sample set. In this way, the foreground of foreign objects can be segmented. Finally, in order to adapt the algorithm to the complex changes of monitoring scenarios, use random selection mechanism and neighborhood spread mechanism to update the background model online. The detailed steps are as follows:

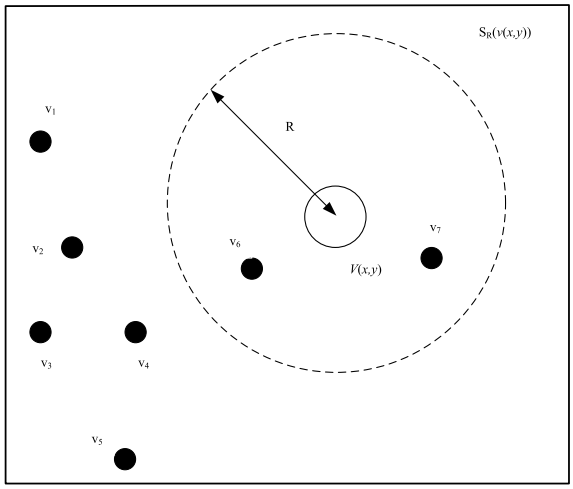


FIGURE 3. Schematic diagram of pixel classification with Vibe algorithm.

Step 1: initialize the background model

The background model is M(0). As shown in Equation (5), a sample set is established for each pixel. The background sample set at pixel (x, y) is $G_v(\mathrm{x}, \mathrm{y})$, and $G_v(\mathrm{x}, \mathrm{y})$ contains n background sample pixel values $v_i (i \in \{0, 1, \ldots, n\})$.

$$G_v(x, y) = \{v_1, v_2, \ldots, v_{n-1}, v_n\} \tag{5}$$

As adjacent pixels have the similar features in the spatial and temporal distribution, the initialization of background model $G^0(x, y)$ is completed based on the first Frame M(0) of the video sequence. The sample pixel values v_i are all selected from the 8 neighborhoods $N_G(x, y)$ of pixel (x, y) with equal probability, as shown in Equation (6). In the second Frame M(1) of the video sequence, the second step of foreign object intrusion detection can be taken.

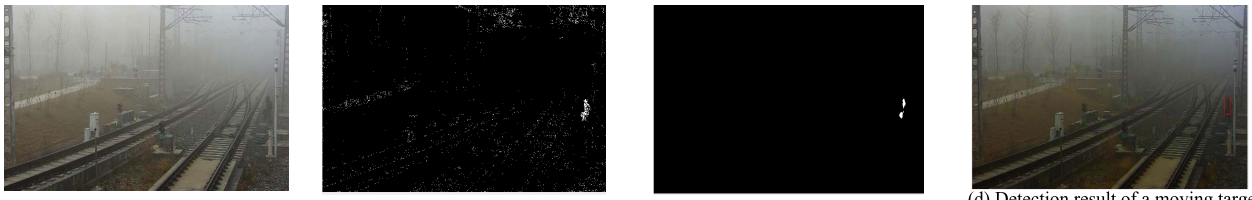
$$G^0(x, y) = \{v^0(x_0, y_0) | (x_0, y_0) \in N_G(x, y))\} \tag{6}$$

Step 2: detect foreign object intrusion

As shown in Fig.3, take the pixel value at pixel (x, y) in the Frame M(t) ($t \geq 1$) to be detected as $v(x, y)$, then calculate the number of elements intersected by the background sample set $G^{t-1}(x, y)$ and the circle $S_R(v(x, y))$ with $v(x, y)$ as its center and R as its radius. If the number is not less than the threshold T_{min}, $v(x, y)$ is deemed as similar to the background sample set $G^{t-1}(x, y)$, and the pixel (x, y) is regarded as the background point or otherwise the foreground point, as shown in Equation (7).

$$v(x, y) = \begin{cases} 0, & |S_R(v(x, y)) \cap G^{t-1}(x, y)| \geq T_{\min} \\ 255, & |S_R(v(x, y)) \cap G^{t-1}(x, y)| < T_{\min} \end{cases} \tag{7}$$

M(t) is converted into binary image B(t) with Equation (7), as shown in Fig.4 (a) and Fig.4 (b). Then calculate the Euclidean distance between v_i and $v(x, y)$ of each background sample point. If the distance is less than or equal to R, v_i is called an intersected element. The statistical task can be completed in this way.



(a) Original image of Frame M(78) (b) B(78) after pixel classification (c) B(78) after morphological processing (d) Detection result of a moving target after fog removal

FIGURE 4. Foreign object intrusion detection with ViBe algorithm.

In order to reduce the influence of external interference such as withered grass and shaking shadows, median filtering and expansion corrosion processing are applied to binary image B(t). The results are shown in Fig.4 (c). The foreground connected domain in B(t) where the minimum outer rectangular area is larger than the threshold value is regarded as the intruding foreign object. Then, according to the location of the smallest outer rectangle in the foreground connected domain on the image, the subsequent work to be performed is selected.

Let the resolution of M(t) image be w × h, and the upper, lower, left and right boundary values of the minimum outer rectangle be $x_{\min}$, $x_{\max}$, $y_{\min}$ and $y_{\max}$ respectively. When $x_{\min} = 0$ or $x_{\max} = \mathrm{w}$ or $y_{\min} = 0$ or $y_{\max} = \mathrm{h}$, the foreign object does not fully enter the video surveillance range and can not provide complete target information for the follow-up tracking, so it is necessary to continue to update the background model and implement Step 3. Else, the foreign object has completely entered the video surveillance area. The minimum outer rectangular area is used as the initial tracking box (box_0) for intensive cyclic sampling in the subsequent tracking.

Step 3: update background model

In the ViBe algorithm, the background sample set $G^{\mathrm{t}}(x, y)$ must be updated in four steps, so as to adapt to various changes in the space-time dimension of complex monitoring scenarios.

(1) Recovery of the background area: Since the foreground pixels do not participate in the update of the background model, it is possible to generate ''ghosts'' and be erroneously detected as pixel blocks of the foreground area, which is difficult to re-integrate into the background. In real application, the target moving based on the foreground will not be in the same location for a long time. The continuous frame pixels detected as the foreground are reset as the background pixels so that the algorithm have certain self-healing abilities.

(2) Random selection of update points: take an integer in an arbitrary interval [1, (2Φ-1)] as a random step, and select pixels (x, y) in the binary image B(t) row by row. If the selected pixels are background pixels, $G^{\mathrm{t}}(x, y)$ is updated with the pixel value $v(x, y)$. There is only $1/\Phi$ chance for the background pixels to update the sample set, so the update frequency of the background model is lower. As a result, it is not easy to miss the detection of slow moving objects.

(3) No memory updating of background sample set: if the randomly selected pixel (x, y) is a background pixel, in order to avoid long-term retention of samples in the sample set and the influence on the accuracy of the model, the pixel value $v(x, y)$ in M(t) is used to replace any sample in $G^{\mathrm{t}}(x, y)$, so that the life cycle of the sample decrease exponentially.

(4) Updating of sample set in spatial neighborhood: based on the assumption that neighborhood pixels have the spatial consistency and in order to eliminate ''ghosting'' effect as soon as possible, one pixel (x_0, y_0) is selected from $N_G(x, y)$ randomly and $v(x, y)$ is used to replace any sample in $G^{\mathrm{t}}(x_0, y_0)$ so that background sample information gradually spread outward.

After the above updating steps are completed, return to Step 2 to detect the foreign objects in Frame M(t + 1). The final detection results are shown in Fig.4 (d). Since the boundary value of the minimum outer rectangular of the intruding foreign object in Fig.4 (d) does not meet $x_{\min} = 0$ or $x_{\max} = \mathrm{w}$ or $y_{\min} = 0$ or $y_{\max} = \mathrm{h}$, it is considered that the foreign object has completely entered the video monitoring area in Frame M(78). The FHOG features of Frame M(78) are sampled intensively and cyclically to train a location filter to track the foreign object from Frame M(79).

IV. SCALE-ADAPTIVE TRACKING ALGORITHM WITH DIMENSIONALITY REDUCTION

FHOG features are extracted from the original samples. The description of FHOG features aims to merge the local unsigned and signed histograms of oriented gradient (HOG), which is insensitive to the geometric deformation and illumination changes of the target. It is suitable to construct the appearance model [24] for foreign objects such as pedestrians, animals and vehicles. The FHOG features of the initialization tracking box (box_0) in Fig.4 (d) are extracted as the original samples to generate the sample set for the location training.

A. CONSTRUCTION OF KERNELIZED-CORRELATION SHIFT FILTER

1) DENSE CYCLIC SHIFT OF THE ORIGINAL SAMPLES

Since the FHOG feature matrix in the initialization tracking box (box_0) contains less information, not enough to train a kernelized-correlation location filter. Therefore, when tracking the starting Frame M(78), set the center of the minimum

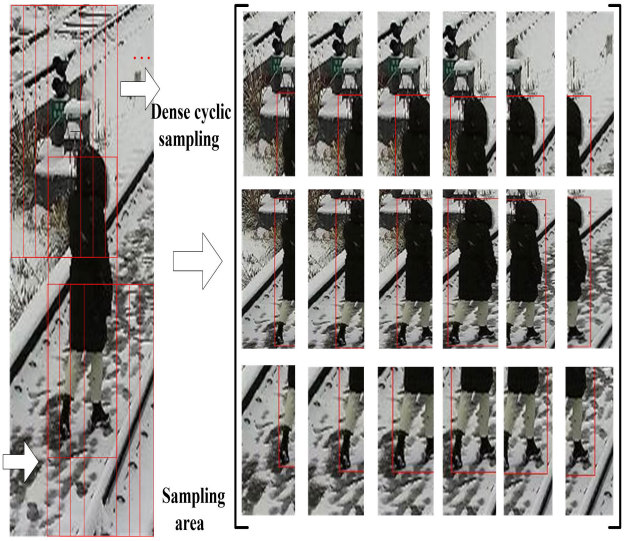


FIGURE 5. Dense cyclic sampling for location training samples.

outer rectangular of the foreign object as the center, and then conduct the dense and cyclic sampling in the sampling area which is twice as big as that of the box$_0$. Finally several training samples with the same size as box$_0$ can be obtained. The sampling process is shown in Fig.5.

Assuming that the size of the initialization tracking box is $w_0 \times h_0$, define a sliding sampling window of the same size with the original algorithm. Then, starting from the upper left corner of the area to be searched, slide the sampling window in an order from left to right and from top to bottom. For each pixel that has been slid, FHOG features in the window are extracted as training samples. This algorithm exhibits a poor real-time performance and needs to be improved because it takes a lot of time to collect $w_0 \times h_0$ training or test samples for each frame.

Specific improvement steps include: Firstly, use a 2D Hann window to filter the FHOG matrix of box0, so as to obtain the original sample S_g with the boundary effect eliminated. After that, multiply S_g right by the horizontal displacement matrix R, simulating the horizontal motion of the sliding sampling window, and multiply S_g left by the vertical displacement matrix Q, simulating the longitudinal motion of the sliding sampling window. In this way, a location training sample matrix S_P can be obtained. Displacement matrices R and Q are shown as below.

$$R=\begin{bmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ & & \dots & & \\ 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix} \quad Q=\begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ & & \dots & & \\ 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

The location training sample matrix S_P can be expressed as Equation (8), where $i \in [1, h_0/l]$, $j \in [1, w_0/l]$.

$$S_p(i,j) = Q^{(i-1)l} S_g R^{(j-1)l} \tag{8}$$

This matrix is a block circulant matrix, where the Fourier domain can be diagonalized to accelerate the training of a target location filter based on a kernel ridge regression function, thereby determining the target orientation.

2) TRACKING THE LOCATION OF THE FOREIGN OBJECT BASED ON KERNEL RIDGE REGRESSION

The kernel ridge regression algorithm is integrated into the filtering process of the area to be searched in Frame M(79), and the candidate area with the greatest response is determined to be the location to track the foreign objects [11]. Firstly, map the location training sample set S_p to a linear separable high-dimensional feature space with a non-linear mapping function $\Phi(S_p)$. Then, train a linear regression function $Y_p = \omega^T \Phi(S_p)$ with $\Phi(S_p)$, where Y_p is the score matrix corresponding to the location training sample set, and ω is a vector in the high-dimensional feature space spanned by $\Phi(S_p)$ and satisfy Equation (9).

$$\omega = \sum\nolimits_{i,j} m_{ij} \phi(S_p)_{ij} \tag{9}$$

where, m is the parameter to be adjusted, substituting Equation (9) into the linear regression function can be used to obtain a kernel correlation position filter Y_p. Where m is the parameter to be adjusted.

The optimal parameter group d makes the position filter cost loss function minimum, as shown in equation (10).

$$m = \min_m \left\| Y_p - Y \right\|_2^2 + \lambda \left\| \phi(S_p) m \right\|_2^2 \tag{10}$$

where, Y is a two-dimensional Gaussian matrix, also the ideal output of the location filter. In order to make the least squares fitting which might has multiple sets of solutions have a unique solution, set the value of λ, which is the regularization coefficient of the cost loss function. Then, obtain the partial derivative of the cost loss function against m, and determine that m is the optimal solution when the partial derivative value is 0. Finally, obtain m as shown in Equation (11).

$$m = (Y_p m^{-1} + \lambda I)^{-1} Y \tag{11}$$

where, $\Phi(S_p)\Phi(S_p)^T$ is the inner product kernel matrix of location sample set S_p in the high-dimensional feature space, also can be expressed as $Y_p m^{-1}$, abbreviated as $K(S_p, S_p)$. The Equation (11) can be converted into the Equation (12).

$$m = (K(S_p, S_p) + \lambda I)^{-1} Y \tag{12}$$

The matrix inversion operation in Equation (12) takes up a lot of computing resources, and reduces the real-time performance of the algorithm. If S_p is still a block circulant matrix after being mapped to the feature space, the complex matrix inversion operation can be simplified to point multiplication based on the truth that the Fourier domain can be diagonalized in a block circulant matrix In this paper, the Gaussian kernel matrix as shown in Equation (13) is used to ensure that S_p is still a block circulant matrix after being mapped into $K(S_p, S_p)$.

$$K(X, Y) = \exp(-\frac{1}{\sigma^2}(\|X\|_2^2 + \|Y\|_2^2 - 2(X \bullet Y)^T)) \tag{13}$$

The block circulant matrix $K(S_p, S_p)$ is diagonalized with Equation (14).

$$K(S_p, S_p) = Fdiag(\hat{K}(S_g, S_g))F^H \tag{14}$$

where, F is a discrete Fourier matrix, diag () is used to generate diagonalization matrix, and "^" represents the Fourier transform of the matrix. By substituting Equation (14) into Equation (12) and according to the reciprocal of elements on the diagonal line inverse to the diagonalization matrix, the fast parametric tuning equation (15) of the location filter can be obtained.

$$\mathrm{m} = ifft(\frac{\hat{K}(S_g, S_g) \bullet \hat{Y}}{\hat{K}(S_g, S_g)^T \bullet \hat{K}(S_g, S_g) + \lambda}) \tag{15}$$

Then, take the sampling area of Frame M(78) as the area to be searched for Frame M(79), and the tracking box of Frame M(78) as the sliding sampling window of Frame M(79); and extract the test sample matrix S_{pt} of Frame M(79) in the video sequence via intensive cyclic sampling and substitute it into the kernelized-correlation location filter to obtain the response Y_{pt} as shown in Equation (16).

$$Y_{pt} = \phi(S_p)\phi(S_{pt})^T m = K(S_p, S_{pt})m \tag{16}$$

where, $K(S_p, S_{pt})$ is the Gaussian kernel matrix of the test sample set S_{pt} and the training sample set S_p. Finally, take the maximum coordinate in the response matrix Y_{pt} as the predicted location of the intruding foreign object of Frame M(79) to support the scale estimation of the foreign object. In this way, all intruding foreign objects can be identified completely.

B. CONSTRUCTION OF SCALE CORRELATION FILTER WITH PCA DIMENSIONALITY REDUCTION

1) CONSTRUCTION OF SCALE TEST SAMPLE SET

The center of the scale pyramid is the center of a series of minimal outer rectangles of the intruding foreign objects. The image blocks of different sizes form a sample set which can provide enough test samples for the scale estimation of foreign objects. Assume that the dimension of scale filter is d. For an integer $n \in [(1\text{-}d)/2,(d\text{-}1)/2]$, take the position of the intruding foreign object predicted with Equation (16) as the center, and the size of tracking box of Frame M(78) as the benchmark, intercept a series of image blocks with the size of $R^n w_0 \times R^n h_0$ from Frame M(79), and let R be the scale factor between pyramid layers. Then, we can obtain a d-layer image scale pyramid as shown in Fig.6.

After that, adjust the size of each layer of the scale pyramid through bilinear interpolation, so that the size of each layer is consistent with that of the tracking box of Frame M(78). Finally, take the processed FHOG information of each layer of the scale pyramid as the scale test sample set S_{st}.

2) CORRELATION FILTER OF SCALE TEST SAMPLE SET

According to the principles of correlation filtering, the signal convolution response value of similar images is high [25].

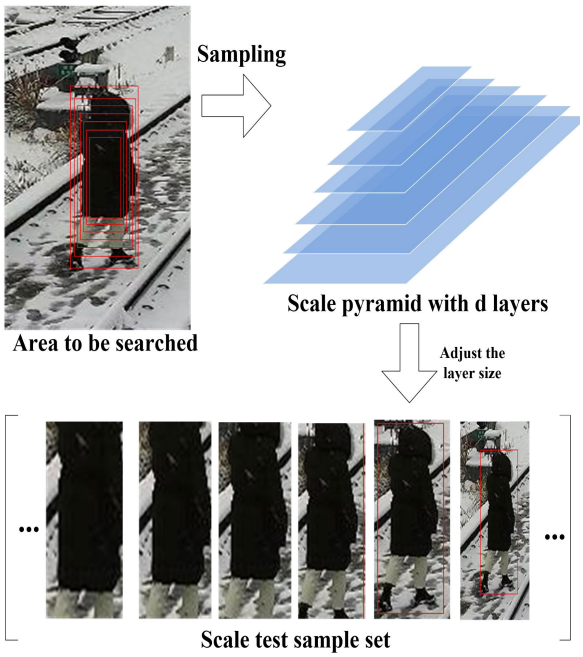


FIGURE 6. Scale pyramid of foreign object.

First, convolve the FHOG feature S_s of the tracking box for Frame M(78) with scale filtering template H to minimize the sum of squares of errors of cost loss function, as shown in Equation (17).

$$\varepsilon = \left\| g - \sum_{i=1}^{3m+4} H^i * S_s^i \right\|_2^2 + \lambda \sum_{i=1}^{3m+4} \left\| H^i \right\|_2^2 \tag{17}$$

where, g is a three-dimensional Gaussian matrix, representing the ideal output of the scale filter; and H is the scale filter template. Firstly, convert the convolution in time domain to point multiplication in frequency domain by discrete Fourier transform, and then obtain the partial derivative of ε with respect to H^i. The optimal solution of Layer i of the scale filter template H is H^i when the partial derivative is 0, as shown in Equation (18).

$$\hat{H}^i = \frac{conj(\hat{g}) \bullet \hat{S}_s^i}{\sum_{j=1}^{d} conj(\hat{S}_s^j) \bullet \hat{S}_s^j + \lambda} \tag{18}$$

The conj in the equation refers to the complex conjugation of matrices. With the scale pyramid technique, a d-dimensional test sample set S_{st} is established in Frame M(79) and substituted into the scale filter (19) to obtain response Y_{st}.

$$Y_{st} = ifft(\sum_{i=1}^{3m+4} \hat{H}^i \bullet \hat{S}_{st}^i) \tag{19}$$

Then, find the scale test samples which can make the Y_{st} reach the highest score in the test sample set S_{st}, and take the size of the corresponding layer on the scale pyramid as the

scale of the tracking box (box_1) for Frame M(79), to identify all the intruding foreign objects.

Finally, after updating the location filter parameter m with Equation (15), prepare to predict the target location in Frame M(79). At the same time, in order to update the scale filter template quickly, the molecular A and denominator B of the scale filter template H^i are iterated frame by frame with Equation (20). Where, constant η represents the learning rate of the scale filter and $S_s(t+1)$ represents the FHOG feature of the t + 1st frame tracking box.

$$A^i(t+1) = (1-\eta)A^i(t) + \eta conj(\hat{g})S_s^i(\hat{t}+1)) \tag{20a}$$

$$B(t+1) = (1-\eta)B(t) + \eta \sum_{j=1}^{d} conj(\hat{S}_s^j(t+1))\hat{S}_s^j(t+1) \tag{20b}$$

C. PCA DIMENSIONALITY REDUCTION OF SCALE FEATURES

In Equations (19) and (20), the tracking speed of the genetic Kcf algorithm is reduced to some extent when considering that the FHOG features of the scale test samples need to be subjected to 3m + 4 2D discrete Fourier transform during the scale estimation of new foreign objects. To reduce the time required for target detection and filter updating, and based on the truth that the dimensionality of scale test sample features is much higher than the quantity of scale test samples, Principal Component Analysis (PCA) will be used to reduce the dimensionality of the FHOG features of scale test samples without losing any image information.

PCA dimensionality reduction is a commonly used data analysis method. It projects a set of related variables into an unrelated low-dimensional subspace through translation and rotation of coordinates, to reduce the data dimensionality. The coordinate axes of the low-dimensional subspace is selected as the direction with the largest variance in the high-dimensional data to extract the main characteristic component of the original data.

Detailed steps include: firstly, move the coordinate origin to the center of the sample data with the Equation (21) so that those samples that are independent originally are roughly related now.

$$S_{st}(i) = S_{st}(i) - mean(S_{st}(i)) \tag{21}$$

where $S_{st}(i)$ is the expansion of the FHOG feature of the i^{th} test sample. Then, perform the eigenvalue decomposition on the d-dimensional covariance matrix U as shown in the Equation (22) and set each row of the projection matrix V_t as a feature vector corresponding to U, to eliminate data offset caused by image noise.

$$U(i,j) = S_{st}(i)S_{st}(j)^T \tag{22}$$

Finally, using $V_tS_{st}(i)$ to project $S_{st}(i)$ into the low-dimensional subspace, the test sample $S_{st}(i)$ is reduced from $(3m + 4) \times w_t/l \times h_t/l$ dimension to rank(U) dimension. $S_s(t + 1)$ in the Equation (20) also belongs to the test sample set S_{st}, so it can be replaced in a similar way by $V_tS_s(t + 1)$, thereby reducing the time required for the update of scaling filter.

V. EXPERIMENT RESULTS AND ANALYSIS

A. EXPERIMENT EXAMPLES AND INDICATORS OF PERFORMANCE EVALUATION INDICATORS

(1) Examples: two sets of video image sequences were collected from some railway test line by using a high-definition network dome at a frame rate of 25 fps. One set of the sequences are the video sequences T1 and T2 with almost no change in the scales of foreign objects; and the other set are the video sequences S1 and S2 with dramatic changes in the scales of foreign objects. The durations of the video sequences T1 and T2 are 20s and 24s, with 489 and 596 frames of images respectively. The durations of the video sequences S1 and S2 are both 28s, with the same 700 frames of images.

The computer used in the experiment, with an Intel i5-8250U CPU and 8GB of memory, is equipped with a win10 operating system. The performance of the Defog-PSA-Kcf algorithm was compared respectively with that of the Mean Shift and Kcf algorithms which are short of the scale estimation process, and with that of the SA-Kcf and SAMF algorithms which possess scale-adaptive functions, so as to analyze the differences in tracking performance of different algorithms.

In addition, the values of parameters used by the algorithm proposed by this paper are fixed, as shown in Tab.1.

When the variance σ of Gaussian kernel function is close to 0, the sample data mapped to the high-dimensional space is easy to be over-fitted by the linear function while when σ is close to 1, the data may be linearly inseparable, so σ is taken as 0.5— the intermediate value of 0-1.

In addition to keep the value of σ not being too close to 0 to avoid over-fitting, a relatively simple position filter should be obtained without under-fitting. Consider adding a regularization term using L2 norm for the cost function. We want the coefficient λ of the regularization term to be as small as possible, because when $\lambda > 1$, most of the parameters of the hypothesis function will be less than 1. With the increase of λ, the values of each parameter will decrease continuously to zero, resulting in under-fitting of the training samples. Therefore, λ should be taken as a smaller positive value in the interval of [0,1], and it is taken as 10^{-4} in this paper.

(2) Performance evaluation indicators: in terms of tracking accuracy, the overlap rate between the predicted tracking frame and the artificial labeled tracking frame is calculated, and the success rate is calculated by counting the proportion of the image frame exceeding the overlap rate threshold in the video sequence to measure the tracking accuracy [26]. It can simultaneously reflect the accuracy of the algorithm in position and scale estimation, as shown in equation (23).

$$overlap = \frac{|box_p \cap box_t|}{|box_p \cup box_t|} \tag{23}$$

TABLE 1. Parameter information.

Process	Parameter name	Symbol	Configured value
Target detection	Number of background model samples	n	20
	Sphere radius	R	30
	Segmentation threshold	T_{min}	2
	Random step	Φ	10
Feature extraction	Gradient partition number	m	9
	Side length of cell units	l	4
Location tracking	Variance of Gaussian kernel function	σ	0.5
	Regularization parameters	λ	10^{-4}
Scale Estimation	Learning rate	η	0.02

where box_p refers to prediction tracking box and box_t refers to the manually recorded tracking box. Then count the proportion of image frames whose overlapping rate exceeds the threshold {0.1, 0.2, 0.3, . . ., 1} in the video sequences to draw an AUC curve and obtain a success rate map. The area under the AUC curve is usually taken as the final score, which can reflect the tracking accuracy of the algorithms. The AUC score is proportional to the tracking accuracy. The tracking speed can be measured by calculating the average number of image frames processed by each algorithm per second.

B. EVALUATION OF DEFOGGING PERFORMANCE

In order to obtain a clearly separable restored image of the background and the person, an appropriate transmittance estimation value is determined. Based on this, two groups of typical foreign matter intrusion images on the railway are selected and adjusted to 720 ∗ 404. Accurately find the actual value of the best fidelity adjustment factor w_o for a typical image under the application scenario of the railway integrated video monitoring system, measure the definition of the typical image with the peak signal-to-noise ratio, and select the data point w between [0,1] to obtain the peak signal-to-noise ratio value corresponding to the w value of a typical image in a frame of video. By comparing linear and non-linear least squares data points for various fitting, the w-PSNR curve of a typical image is fitted by the best fitting method, and the maximum value of the fitting function is obtained to determine the best fidelity adjustment factor w_o of the typical image, and $w_o = 0.3571$ is obtained, as shown in Fig.7.

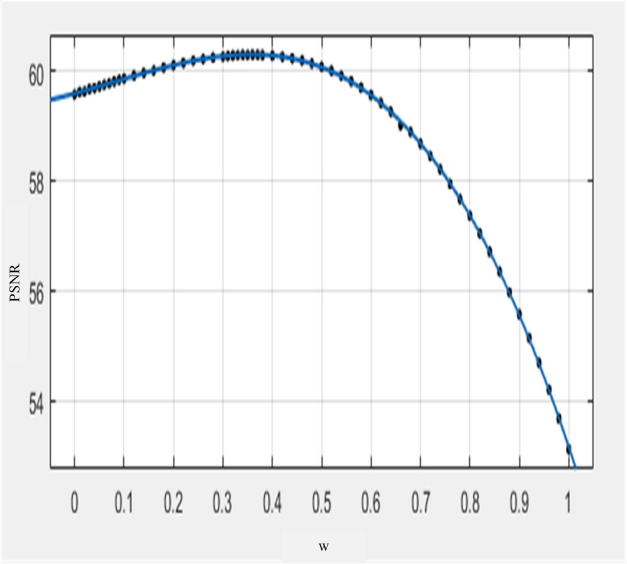


FIGURE 7. w-Peak Signal to Noise Ratio fitting curve.

TABLE 2. Information entropy of images.

IE	160th frame	270th frame
Original image	6.2182	6.2714
Defog-PSA-Kcf algorithm	7.5120	7.4307

To evaluate the performance of defogging, the objective indicators such as Information Entropy(IE) which is set to restore image, and Peak Signal to Noise Ratio(PSNR), are used to reflect the image quality after restoration. IE is an indicator of image details, and the larger the IE, the clearer the image. PSNR peak signal-to-noise ratio is the most widely used objective evaluation index for images. The PSNR is based on the error between corresponding pixel points, that is, based on error-sensitive image quality evaluation, and the lager the PSNR, the better the image quality and the higher the sharpness.

A typical image under the railway video integrated monitoring system is selected. Such as the 160th image and 270 frames in the video T1, calculate the information entropy and peak signal to noise ratio before and after defogging, and the results are shown in Tab.2 and Tab.3.

It can be seen from Tab.2 that the IE of defogged images are higher than those of original images, which indicates that the image information after defogging is increased. Tab.3 shows that the restoration image after defogged have higher Peak Signal to Noise Ratio. To sum up, the indicators

TABLE 3. Peak signal to noise ratio of images.

PSNR	160th frame	270th frame
Defog-PSA-Kcf algorithm	60.2798	61.2610

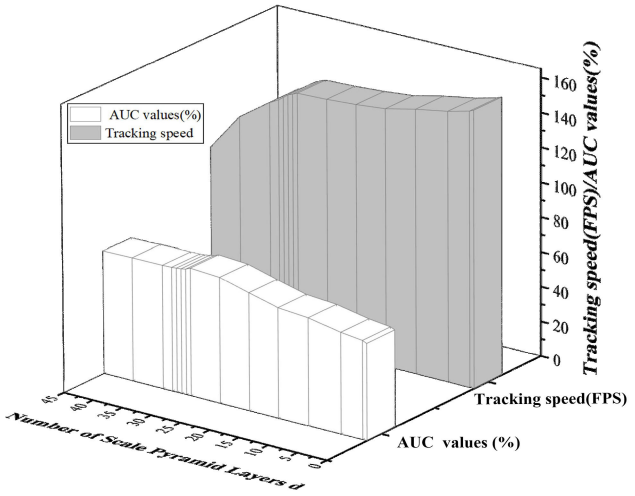


FIGURE 8. Performance impact test of the number of scale pyramid layers.

after defogging are all better than those before defogging, which can prove the superiority and effectiveness of its defogging ability.

C. TEST ON THE INFLUENCE OF SCALE PYRAMID LAYERS ON TRACKING PERFORMANCE

If the number of scale pyramid layers d is too small and there are less scale test samples, the accuracy of scale estimation will be lower; if d is too large, then the dimensionality reduction effect of the scale test samples will be reduced to some extent, and therefore the tracking speed will decrease. With different values of d (number of scale pyramid layers), AUC values and the tracking speed diagrams of the Defog-PSA-Kcf algorithm on the Sequence T and Sequence S are used to examine the effect of the values of d on the tracking performance of the Defog-PSA-Kcf algorithm, and the results are shown in Fig.8.

It can be seen from Fig.8 that: when $d = 0$, the tracking accuracy and tracking speed of the Defog-PSA-Kcf algorithm are consistent with those of genetic Kcf algorithm devoid of the scale estimation process; when $d < 25$, the tracking accuracy of the Defog-PSA-Kcf algorithm will continuously improve with the increase of d; and when $d \geq 25$, the AUC value will be around 73%. Therefore, $d = 25$ is the intermediate value. It can be concluded that if the Defog-PSA-Kcf algorithm need to track videos of foreign object intrusion in different scenarios, and at the same time its tracking accuracy is to be maximized, then the value of d should be greater than 25. It is also observed that the tracking speed of the Defog-PSA-Kcf algorithm will decrease with the increase of d, and 33 is the inflection point. When d is greater than 33, the tracking speed of the Defog-PSA-Kcf algorithm will drop precipitously. In order to improve the speed performance of the algorithm on the premise of ensuring tracking accuracy, the value of the scale estimation layer d of the Defog-PSA-Kcf algorithm was set to 33 layers, and subsequent experiments are performed, and the percentage in the average dimensionality reduction of each video sequence is shown in Tab.4.

TABLE 4. The percentage in the average dimensionality reduction of each video sequence.

sequence	T1	T2	S1	S2
percentage	97.82%	98.09%	98.6%	98.64%

In the case of original video sequence without dimensionality reduction, the average dimensionality of the video sequence is about 1000, which is larger than the average dimensionality after dimensionality reduction by principal component analysis. The average dimensionality of video sequence after dimensionality reduction is below 33 dimensions. Tab.4 shows the effect of dimensionality reduction, where the average dimensionality reduction percentage of each video sequence is more than 95%, and the data will not be sparsely distributed after the scale is compressed, while preserving the important information of the original data. After that, other performance comparison experiments are conducted.

D. COMPARISON EXPERIMENT OF TRACKING ACCURACY

The Mean Shift algorithm, the genetic Kcf algorithm, the scale-adaptive SA-Kcf and SAMF algorithm were used respectively for tracking foreign objects mentioned in the experiment examples, and their different tracking accuracies compared with that of the Defog-PSA-Kcf algorithm are shown in Fig.9. Among them, T1 and T2 are different video sequences without scale changes. S1 and S2 are different video sequences with obvious scale changes.

As it's shown in Fig.9, the Defog-PSA-Kcf algorithm can accurately track the foreign objects in the video sequences T1 and T2 in foggy weather, so can the SA-Kcf and SAMF algorithms. While, the Mean Shift algorithm could not update the target template, therefore it completely lost the target in Frame 83 and Frame 92 of T1 and T2 sequences respectively. Similarly, in Frame 462 of T2 sequence, the Kcf algorithm mistook a low-voltage switch box next to the rail in T2 as the tracking target, as it seems to have a similar FHOG feature with the foreign object due to the low definition of videos on the foggy day. The Kcf algorithm and the Mean Shift algorithm which are short of the scale estimation process both learned FHOG features and RGB color features of excess background due to the large scale changes in the

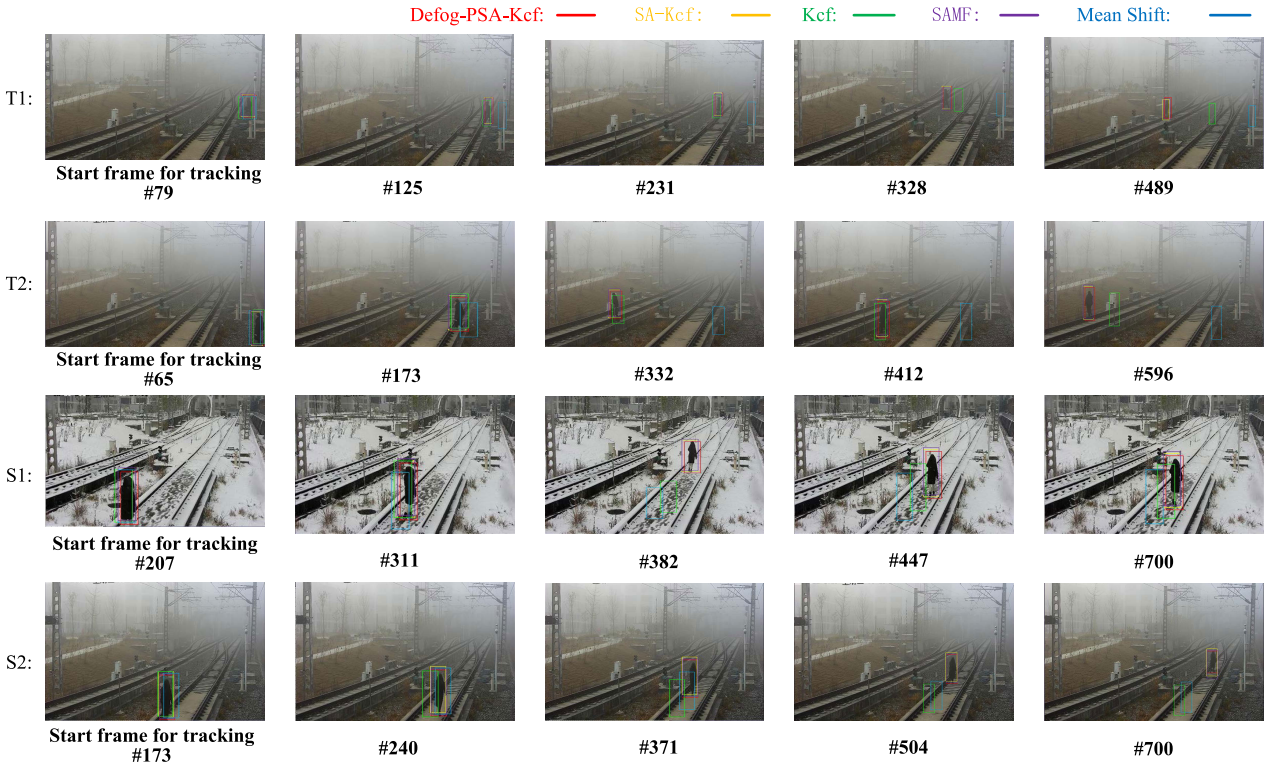


FIGURE 9. Tracking performance of different algorithms.

TABLE 5. Average accuracy of each algorithm within 20 pixels.

Sequence	Defog-PSA-Kcf	SAMF	SA-Kcf	Kcf	MS	Dcf
T	82.41%	75.3%	79.62%	74.45%	35.41%	72.9%

TABLE 6. AUC values of tracking algorithms.

Sequence	Defog-PSA-Kcf	SAMF	SA-Kcf	Kcf	MS	Dcf
T	75.32%	69.32%	71.08%	64.2%	29.7%	60.55%
S	74.3%	70.56%	72.5%	47.6%	21.78%	45.27%

video sequences S1 and S2 respectively, resulting in the drift of tracking boxes.

Fig.10 indicates the success rates of different algorithms in the video sequences T and S are mapped respectively to compare the accuracy performance of each tracking algorithm. In order to quantify the tracking accuracy of each algorithm, the AUC values in the video sequence S is calculated as shown in Tab.5. The average precision and AUC values in the video sequence T are calculated as shown in Tab.6. The threshold for calculating the average accuracy of each algorithm is 20 pixels.

From the Tab.5 and Tab.6, the average accuracy of Defog-PSA-Kcf algorithm proposed in this paper can reach 82.41%, which is better than other algorithms without defogging. When defogging process is performed before the detection, the tracking accuracy of the Defog-PSA-Kcf algorithm proposed in this paper is superior to that of the Dcf, as well as the Kcf and Mean Shift algorithms without the scale estimation process, especially when the scale of the video sequence S changes drastically, the tracking accuracy of the Defog-PSA-Kcf algorithm can reach 74.3%, while that of the Dcf, Kcf and Mean Shift algorithms are 45.27%, 47.6% and 21.78% respectively. While the average tracking accuracy of the SAMF algorithm can be 70.56% as it has only 7 scale estimation layers. When the Defog-PSA-Kcf algorithm defogged the video sequence before detection, PCA dimensionality reduction was also carried out at the same time on the scale test samples to optimize the images, improve the definition of images and shield a large amount of noise interference. Therefore, its average tracking accuracy is slightly higher than the tracking accuracy of 72.5% of the SA-Kcf algorithm.

E. COMPARISON EXPERIMENT OF TRACKING SPEED

The running time of the Defog-PSA-Kcf, SA-Kcf, Kcf, SAMF, Mean Shift and Dcf algorithms on the video sequences T and S were recorded and the number of image

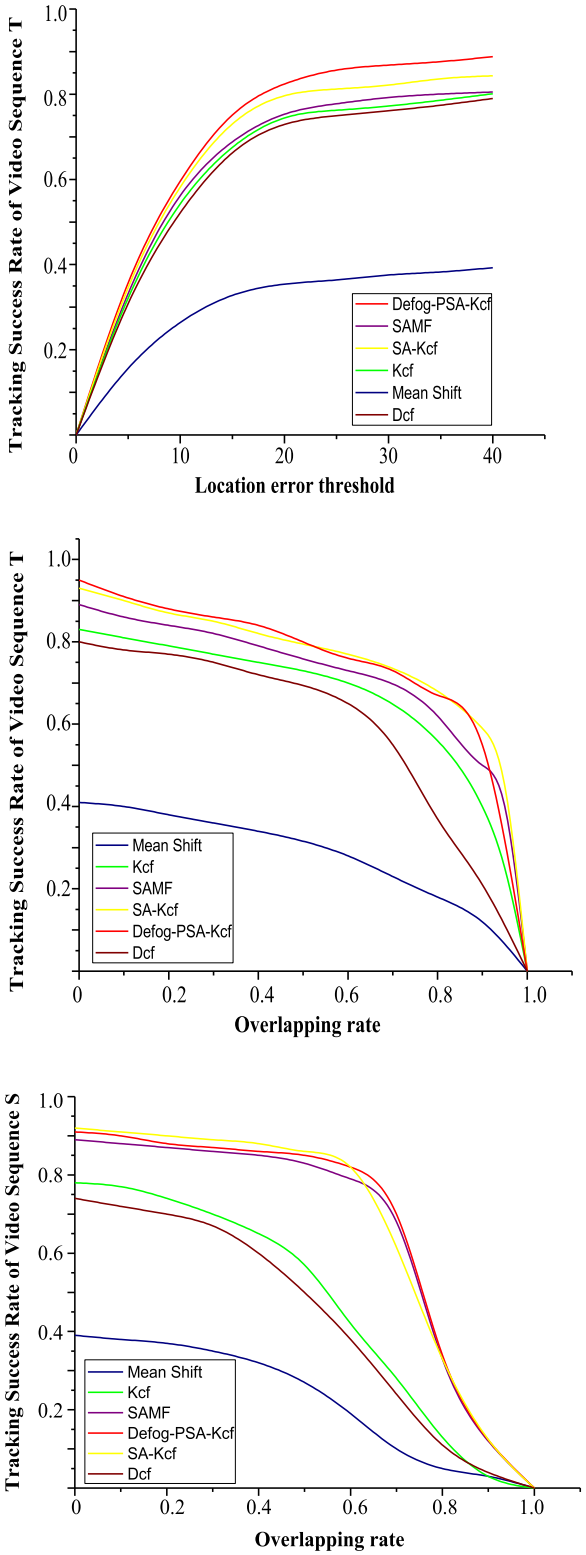


FIGURE 10. Tracking success rates of various algorithms.

frames processed by each algorithm in unit time were also calculated to reflect the tracking speed of different algorithms. The results are shown in Fig.11.

Fig.11 indicates that the tracking speeds of the correlation filter-based tracking algorithms are all above 90FPS,

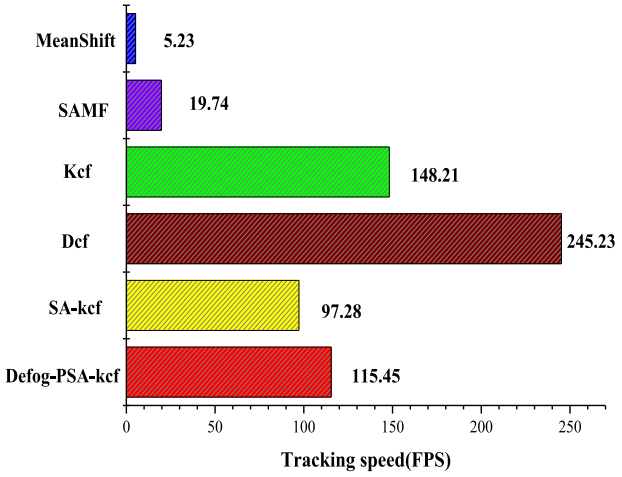


FIGURE 11. Tracking speeds of algorithms.

higher than those of SAMF algorithms (19.74FPS) and the Mean Shift algorithm (5.23FPS). Among others, the tracking speed of the Kcf algorithm could be as high as 148.21FPS because it requires no scale estimation of foreign objects. The tracking speed of the Dcf algorithm could be reach 245.23FPS. Although the SA-Kcf algorithm with added scale estimation was superior to the Kcf algorithm in tracking accuracy, its tracking speed has decreased to 97.28 FPS. The Defog-PSA-Kcf algorithm used principal component analysis to greatly reduce the times of two-dimensional discrete Fourier transform (2D-DFT) in the scale estimation process, which made the tracking speed rise to 115.45FPS. Despite of slight decrease in the tracking speed due to the defogging process before the detection, its tracking accuracy was improved. Compared with the Kcf algorithm, the tracking accuracy of the Defog-PSA-Kcf algorithm increased dramatically while only at the expense of slight decrease in tracking speed (32.76FPS). Although the speed of Dcf algorithm is faster than that of Kcf algorithm, the tracking speed of Kcf algorithm and Defog-PSA-Kcf algorithm has generally met the needs of this application scenario. Meanwhile, emphasis is placed on improving the tracking accuracy of the target. In this case, compared with the Kcf algorithm on the original basis, Defog-PSA-Kcf algorithm substantially improves the tracking accuracy of the target while only sacrificing the speed of 32.76FPS, and has the function of tracking foreign matters in foggy weather.

VI. CONCLUSION

The Defog-PSA-Kcf algorithm can integrate the dark channel prior, kernelized correlation filter and intensive cyclic sampling to accurately determine the location of foreign objects in foggy weather. Improve the original Kcf algorithm: Add dark channel prior to defogging the video sequence, and integrate scale filters into the scale estimation stage by error least squares sum function and scale pyramid technique. Moreover, by principal component analysis, the error caused by

redundant information is reduced. The improved algorithm effectively increases the tracking accuracy of foreign objects in railway intrusion.

Experimental results show:

(1) The average accuracy of the proposed algorithm in four video sequences can reach 74.81%, which is better than that of the Mean Shift, Dcf and Kcf algorithms without scale estimation of 25.74%, 52.91% and 55.9%, slightly higher than that of the SA-Kcf and SAMF algorithms with scale self-adaptation of 71.79% and 69.94%.

(2) The tracking speed of the algorithm is improved: and the tracking speed of the Defog-PSA-Kcf algorithm from97.28FPS to 115.45FPS, which is as fast as that of the genetic Kcf algorithm.

REFERENCES

[1] J. J. Garcia, J. Urena, Á. Hernandez, M. Mazo, J. A. Jimenez, F. J. Alvarez, C. De Marziani, A. Jimenez, M. J. Diaz, C. Losada, and E. Garcia, "Efficient multisensory barrier for obstacle detection on railways," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 702–713, Sep. 2010.

[2] H. Zhang, X. Li, H. Zhong, Y. Yang, Q. M. J. Wu, J. Ge, and Y. Wang, "Automated machine vision system for liquid particle inspection of pharmaceutical injection," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1278–1297, Jun. 2018.

[3] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. CVPR*, Jun. 2000, pp. 142–149.

[4] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–575, May 2003.

[5] X. Chen, S. Wang, C. Shi, H. Wu, J. Zhao, and J. Fu, "Robust ship tracking via multi-view learning and sparse representation," *J. Navigat.*, vol. 72, no. 1, pp. 176–192, 2019. doi: 10.1017/S0373463318000504.

[6] S. Hare, S. Golodetz, A. Saffari, V. Vineet, M.-M. Cheng, S. L. Hicks, and P. H. S. Torr, "Struck: Structured output tracking with Kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2096–2109, Oct. 2016.

[7] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.

[8] X. Yang, H. Zhang, L. Yang, C. Yang, and P. X. Liu, "A joint multi-feature and scale-adaptive correlation filter tracker," *IEEE Access*, vol. 6, pp. 34246–34253, 2018.

[9] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, CA, USA, Jun. 2016, pp. 4293–4302.

[10] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 2544–2550.

[11] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with Kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[12] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.

[13] D. Wang and J. Zhu, "Fast smoothing technique with edge preservation for single image dehazing," *IET Comput. Vis.*, vol. 9, no. 6, pp. 950–959, 2015.

[14] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[15] Z. Tufail, K. Khurshid, A. Salman, I. F. Nizami, K. Khurshid, and B. Jeon, "Improved dark channel prior for image defogging using RGB and YCbCr color space," *IEEE Access*, vol. 6, pp. 32576–32587, 2018.

[16] H. Yuan, C. Liu, Z. Guo, and Z. Sun, "A region-wised medium transmission based image dehazing method," *IEEE Access*, vol. 5, pp. 1735–1742, 2017.

[17] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[18] Robby T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[19] J.-P. Tarel and N. Hautière, "Fast visibility restoration from a single color or gray level image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Washington DC., USA, Sep./Oct. 2009, pp. 2201–2208.

[20] O. Barnich and M. V. Droogenbroeck, "ViBE: A powerful random technique to estimate the background in video sequences," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 945–948.

[21] O. Barnich and M. V. A. Droogenbroeck "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[22] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, no. 3, pp. 233–254, 2002.

[23] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 713–724, Jun. 2003.

[24] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with dicriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[25] M. Danelljan, G. Hãger, F. S. Khan, and M. Felsberg. "Accurate scale estimation for robust visual tracking," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2014, pp. 1–5.

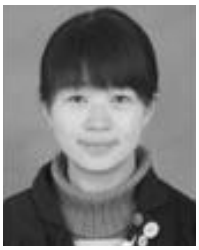
[26] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, Jun. 2013, pp. 2411–2418.

ZHIJIAN QU was born in China, in 1978. He received the M.S. degree from the School of Electrical Engineering, East China Jiaotong University, Nanchang, China, in 2004, and the Ph.D. degree from the School of Electrical Engineering, Beijing Jiaotong University, Beijing, China, in 2012. He has participated a special research in the University of New Brunswick, Canada, as a Visiting Scholar. He is currently a Professor with East China Jiaotong University. His recent research interests include smart distribution grid information networks and big data sets information systems, and intelligent monitoring systems.

WANZHUO YI was born in China, in 1996. She is currently pursuing the master's degree in image processing and computer vision with the School of Electrical and Automation Engineering, East China Jiaotong University.

RUILIN ZHOU graduated from East China Jiaotong University. He is currently with the CRRC Times Signal & Communication Co., Ltd., Changsha, China. His current research interests include image processing and computer vision.

HANLING WANG was born in China, in 1994. He is currently pursuing the master's degree with the School of Electrical and Automation Engineering, East China Jiaotong University. His research interests include image processing and big data monitoring of stream computing.



RUI CHI was born in China, in 1985. She received the M.S. degree in computational mathematics from Northern Nationalities University, Yinchuan, China, in 2012, and the Ph.D. degree in transportation information engineering and control from the School of Automation, Wuhan University of Technology, Wuhan, China, in 2017. She is currently a Lecturer with East China Jiaotong University, China. Her current research interests include intelligent optimization algorithm and neural networks.