doi:10.3772/j.issn.1006-6748.2020.04.007

Anti-occlusion pedestrian tracking algorithm based on location prediction and deep feature rematch[①]

Hu Zhentao(胡振涛)[*], Mao Yihao[*], Fu Chunling[**], Liu Xianxing[②][*]
([*] College of Computer and Information Engineering, Henan University, Kaifeng 475004, P. R. China)
([**] School of Physics and Electronics, Henan University, Kaifeng 475004, P. R. China)

Abstract

Aiming to the problem of pedestrian tracking with frequent or long-term occlusion in complex scenes, an anti-occlusion pedestrian tracking algorithm based on location prediction and deep feature rematch is proposed. Firstly, the occlusion judgment is realized by extracting and utilizing deep feature of pedestrian's appearance, and then the scale adaptive kernelized correlation filter is introduced to implement pedestrian tracking without occlusion. Secondly, Karman filter is introduced to predict the location of occluded pedestrian position. Finally, the deep feature is used to the rematch of pedestrian in the reappearance process. Simulation experiment and analysis show that the proposed algorithm can effectively detect and rematch pedestrian under the condition of frequent or long-term occlusion.

Key words: pedestrian tracking, correlation filter, Kalman filter, deep feature

0 Introduction

Pedestrian tracking is an important research area of computer vision and pattern recognition. It has been applied in many fields such as video monitoring, automatic driving, and unmanned aerial vehicle. Especially in the field of security, pedestrian tracking is considered as the most fundamental technique to complete trajectory analysis, traffic monitoring, gait recognition, and so on[1,2]. It is well known that pedestrian is difficult to be detected accurately when they are occluded, even caused the phenomenon of losing track. In terms of the duration of occlusion, pedestrian occlusion is mainly divided into the short-term occlusion and the long-term occlusion. Besides, in terms of the area of occlusion, it can be divided into the partial occlusion area and the complete occlusion[3].

In general, the tracking processes of pedestrians that are completely occluded can be mainly divided into 4 stages: (1) Target tracking before occlusion; (2) Occlusion judgment of pedestrian; (3) Prediction of pedestrian position under complete occlusion; (4) Rematch of pedestrian under reappearance or loss. Bolme et al.[4] used the correlation filtering to solve the pedestrian tracking problem. The filter is designed according to the minimum output sum of squared error (MOSSE), which can localize the pedestrian according to the criterion of maximum of the response. But given the fact MOSSE only utilizes the gray feature, it is easy to cause tracking drift problem in some complex situations. Henriques et al.[5] proposed a novel kernelized correlation filter (KCF), the ridge regression of linear space is mapped to the high-dimensional space by kernel function. The real-time and accuracy of pedestrian tracking are effectively improved. And on that basis, Li and Zhu[6] designed a new scale adaptive kernel correlation filter tracker with feature integration (SAKCF), which furtherly improves the accuracy of target tracking. Aiming to the problem of partial occlusion, Huang et al.[7] proposed an anti-occlusion and scale adaptive kernel correlation filter (ASAKCF), the occlusion judgment mechanism of ASAKCF can effectively deal with partial and short-term occlusion problems. In Ref. [8], the Kalman filter[9] and the camshift strategy[10] were combined. Kalman filter and camshift strategy are separately utilized to the position prediction and identification of occluded target. Aiming to the problem of complete occlusion, Ma and Wang[11,12] introduced the online target detection mechanism after the tracking failure. Among them, Ref. [11] used support vector machine (SVM) online detection method to

[①] Supported by the National Natural Science Foundation of China (No. 61976080, 61771006) and the Key Project of Henan Province Education Department (No. 19A413006).

[②] To whom correspondence should be addressed. E-mail: liuxxhd@126.com

Received on Dec. 30, 2019

deal with occlusion or loss of tracking rematch problem. In Ref. [12], the single shot multi box detector (SSD)[13] was applied into the correlation filter to identify and locate the target, its advantage is the ability to long-term track the target. Although the above algorithms can solve partial and complete occlusion problems to some extent, they still have some defects such as weak detection ability and poor matching effect.

A novel anti-occlusion pedestrian tracking algorithm based on location prediction and deep feature rematch (ALPDFE) is proposed in this paper. Its goal is to solve the tracking problem of pedestrians under frequent or long-term complete occlusion. The algorithm uses the deep feature of pedestrian's appearance to judge the occlusion. When the pedestrian is not occluded, SAKCF is used to estimate pedestrian location and scale. When the pedestrian is occluded, the location is predicted by Kalman filter. When pedestrian reappears, the deep feature and YOLOv3 method[14] are used to realize the judgment and rematch of pedestrian tracking. The main contributions of this paper are as follows: Firstly, we design a new occlusion judgment method which uses the deep learning strategy to extract pedestrian features. Secondly, in order to accurately estimate the pedestrian location in the occlusion or non-occluded conditions, we propose a location prediction structure by combining correlation filter with Kalman filter. Thirdly, aiming to the pedestrian reappear process, the deep features and target detection method are introduced to realize the pedestrian rematch process.

1 Scale adaptive kernelized correlation filter

1.1 Kernelized correlation filter

(1) Filter train

Giving the training sample of tth frame $(\boldsymbol{u}_t, \boldsymbol{y}_t)$, the goal is to train the filter $\boldsymbol{h}_t$ which minimizes the squared error between sample $\boldsymbol{u}_t$ and its regression target $\boldsymbol{y}_t$. $\boldsymbol{u}_t$ can be obtained by the circulant matrix based on the pedestrian's appearance feature, $\boldsymbol{y}_t$ is considered as the filter response, which will take the maximum response value in the pedestrian location. The mathematical expression for the above model is as follows.

$$\min_{\boldsymbol{h}_t} \sum\nolimits_{i=1}^{n} (f(\boldsymbol{u}_t^i) - \boldsymbol{y}_t^i)^2 + \lambda \| \boldsymbol{h}_t \|_2 \qquad i = 1, 2, \cdots, n \tag{1}$$

where i denotes the index of training samples, λ denotes the regularization parameter used to prevent overfitting. According to the basic knowledge of kernel function, $\boldsymbol{h}_t$ is represented by the feature mapping function φ.

$$\boldsymbol{h}_t = \sum\nolimits_{i=1}^{n} \boldsymbol{\alpha}_t^i \varphi(\boldsymbol{u}_t^i) \tag{2}$$

where $\boldsymbol{\alpha}_t$ denotes the filter parameter after mapping. According to the properties of circulant matrix and Fourier transform, the solution of filter $\boldsymbol{\alpha}_t$ is quickly calculated in the frequency domain.

$$\tilde{\boldsymbol{\alpha}}_t = \tilde{\boldsymbol{y}}_t / (\widetilde{\boldsymbol{K}}(\boldsymbol{u}_t, \boldsymbol{u}_t) + \lambda) \tag{3}$$

where ~ denotes the frequency domain form after Fourier transform, $\tilde{\boldsymbol{y}}_t$ is the ideal filter response, $\widetilde{\boldsymbol{K}}$ denote the kernel matrix in the frequency domain. The inverse Fourier transform of $\tilde{\boldsymbol{\alpha}}_t$ is the required solution $\boldsymbol{\alpha}_t$.

The expression of $\boldsymbol{K}$ corresponds to the kernel function. For Gaussian kernel function, the kernel matrix[5] $\boldsymbol{K}$ is express as

$$\boldsymbol{K}(\boldsymbol{u}_t, \boldsymbol{u}_t) = \exp\left(-\frac{1}{\sigma^2} (\| \boldsymbol{u}_t \|^2 + \| \boldsymbol{u}_t \|^2 - 2F^{-1}(\tilde{\boldsymbol{u}}_t^* \odot \tilde{\boldsymbol{u}}_t)) \right) \tag{4}$$

where F^{-1} denotes the Fourier inversion operation, * denotes the complex conjugate operation, $\boldsymbol{\sigma}$ denotes the parameter of Gaussian kernel, ⊙ denotes the dot-product operation.

(2) Pedestrian location

Let $\boldsymbol{u}'_t$ and $\boldsymbol{\alpha}'_t$ represent the sample model and filtering parameter model of tth frame, respectively. Input $\boldsymbol{u}_{t+1}$ for the $t+1$ frame, the filter response $\tilde{\boldsymbol{y}}_{t+1}$ can be calculated in the frequency domain.

$$\tilde{\boldsymbol{y}}_{t+1} = \widetilde{\boldsymbol{K}}(\boldsymbol{u}_{t+1}, \boldsymbol{u}'_t) \odot \tilde{\boldsymbol{\alpha}}'_t \tag{5}$$

The Fourier inversion of $\tilde{\boldsymbol{y}}_{t+1}$ is $\boldsymbol{y}_{t+1}$. The coordinate corresponding to the maximum value of $\boldsymbol{y}_{t+1}$ is the pedestrian location.

(3) Filter update

As the appearance of pedestrian changes, update $\boldsymbol{u}'_t$ and $\tilde{\boldsymbol{\alpha}}'_t$ with new sample $\bar{\boldsymbol{u}}_{t+1}$ and filter parameter $\bar{\tilde{\boldsymbol{\alpha}}}_{t+1}$ based on the $t+1$ frame pedestrian location.

$$\begin{cases} \boldsymbol{u}'_{t+1} = (1-\eta)\boldsymbol{u}'_t + \eta \bar{\boldsymbol{u}}_{t+1} \\ \tilde{\boldsymbol{\alpha}}'_{t+1} = (1-\eta)\tilde{\boldsymbol{\alpha}}'_t + \eta \bar{\tilde{\boldsymbol{\alpha}}}_{t+1} \end{cases} \tag{6}$$

where η denotes the learning rate.

1.2 Scale adaption

KCF is fixed size on the sample during the tracking process, so it is unable to deal with the scale variations of target. Based on the scale pyramids strategy, SAKCF[6] can realize the adaptive regulation of pedestrian scale by sampling different-sized candidate regions. Define the tth frame pedestrian size and the scale pool as $W \times H$ and s.

$$\boldsymbol{s} = \{s_j\} \quad j = 1, 2, \cdots, k \tag{7}$$

here, j denotes the index of scale. In the $t+1$ frame, sampling the image according to $\boldsymbol{s}$, we can obtain some image patches $\boldsymbol{u}^{s_j}_{t+1}$ of size $s_jW \times s_jH$. And $\boldsymbol{u}^{s_j}_{t+1}$ is adjusted to the fixed size by bilinear interpolation, the filter response of $\boldsymbol{u}^{s_j}_{t+1}$ can be obtained by Eq. (5).

$$\boldsymbol{y}(\boldsymbol{u}^{s_j}_{t+1}) = F^{-1}(\tilde{\boldsymbol{K}}(\boldsymbol{u}', \boldsymbol{u}^{s_j}_{t+1}) \odot \tilde{\boldsymbol{\alpha}}) \tag{8}$$

The different filtering responses $\boldsymbol{y}(\boldsymbol{u}^{s_j}_{t+1})$ for different s_j can be obtained. s' denotes s_j corresponding to the maximum value of filter response.

$$s' = \operatorname{argmax}\{\boldsymbol{y}(\boldsymbol{u}^{s_j}_{t+1}) \mid s_j \in s\} \tag{9}$$

Therefore, according to the t frame pedestrian scale and s', the $t+1$ frame pedestrian scale is considered as $s'W \times s'H$.

2 Anti-occlusion pedestrian tracking algorithm

Although SAKCF has a better tracking performance, it cannot handle the tracking problem of complete occlusion effectively. According to 4 stages characteristics of complete occlusion tracking process, a novel anti-occlusion tracking algorithm is proposed. Specifically, it is divided into the following 3 steps. In the 1st step, a new occlusion judgment approach is designed. In the 2nd step, the pedestrian position is predicted during tracking failure and occlusion by Kalman filter. In the 3rd step, a new rematch strategy is presented for pedestrian reappearance.

2.1 Occlusion judgment

Appearance feature will be changed when pedestrians are occluded. Therefore, we calculate the tracking quality according to appearance features to determine whether there is occlusion. In recent years, the deep learning techniques have emerged as effective methods for the representation of appearance feature, which can learn features automatically from data. Wojke and Bewley[15] designed the light weight convolutional neural network (LWCNN) that its architecture is shown in Fig. 1, and used the deep cosine metric learning method to encode similarity directly into the training objective. The algorithm can obtain the better results for pedestrian re-identification. Using the extracting idea of pedestrian appearance features in Ref. [15], the tracking quality is calculated as follows.

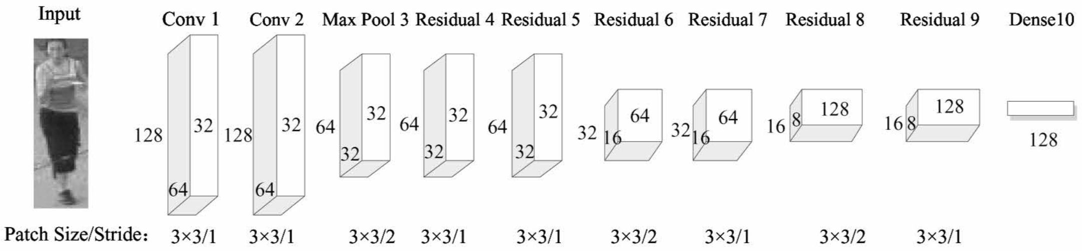


Fig. 1 The network architecture of LWCNN

Let a 128 × 64 RGB color image patch enter into LWCNN, and the feature size is mapped as 16 × 8 RGB color image through a series of convolutional layers. A global feature vector $\boldsymbol{m}$ of length 128 is extracted by fully-connected layer.

$$\boldsymbol{m} = [m_1, m_2, \cdots, m_{128}] \tag{10}$$

The appearance features $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ of the above 2 image patches are extracted separately. Then the similarity between the feature vectors can be calculated as

$$\Phi(\boldsymbol{m}_1, \boldsymbol{m}_2) = \frac{\boldsymbol{m}_1 \cdot \boldsymbol{m}_2^{\mathrm{T}}}{\|\boldsymbol{m}_1\|_2 \|\boldsymbol{m}_2\|_2} \tag{11}$$

where $\|\cdot\|_2$ denotes the 2-norm of vector.

As shown in Fig. 1, the same pedestrian has a higher degree of similarity in different frames when there is no occlusion as Fig. 2(a). The similarity is low when different person or existing occlusions as Fig. 2(b) and Fig. 2(c). Thus, the method used to calculate the tracking quality is effective.

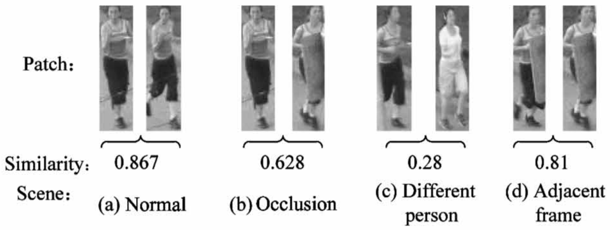


Fig. 2 Cosine distance of deep features for different patch

Constructing the feature template set $\boldsymbol{M}$. The maximum cosine distance by calculating between $\boldsymbol{m}$ and $\boldsymbol{M}$ is used to represent the tracking quality.

$$q = \Psi(\boldsymbol{m}, \boldsymbol{M}) = \max\{\Phi(\boldsymbol{m}_r, \boldsymbol{m}) \mid \boldsymbol{m}_r \in \boldsymbol{M}\} \quad r = 1, 2, \cdots, N \tag{12}$$

where, N denotes the size of $\boldsymbol{M}$.

It is worth noting that suppose only pedestrian deep feature form the previous frame is used as template, because the transformation of pedestrian appearance is slow in the adjacent frame, even if there is occlusion, they will have a high similarity. The template is updated at a fixed interval in Ref. [16], which can not only decrease calculation amount, but also avoid the problem of template drift. Thus, the deep features $\boldsymbol{m}$ of pedestrian every Υ apart frame can be extracted and then the tracking quality q is calculated. If q is more than the threshold q_t, it can be considered that the tracking result is normal and $\boldsymbol{m}$ can be used to update the feature template.

$$\boldsymbol{M} = (\boldsymbol{M} \oplus \boldsymbol{m})[N] \qquad (13)$$

here, $\oplus$ denotes add the $\boldsymbol{m}$ template to the feature template set $\boldsymbol{M}$. $[N]$ is for selecting the latest N features, the update process is shown as Fig. 3.

According to the size of pedestrian and background complexity, q_t is taken usually between 0.77 and 0.83. Combined with the video frame rate and the target moving speed, Υ is taken usually between 6 and 12.

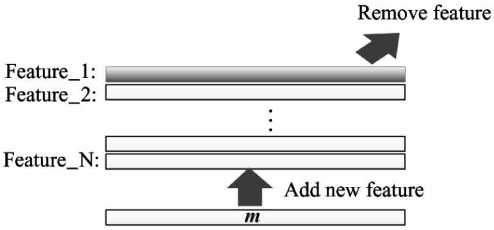


Fig. 3 The update process of feature template

2.2 Occlusion prediction

When q is less than q_t, the pedestrian is occluded, the appearance information of image is not available. The pedestrian location can be determined by the dynamic model of pedestrian motion. Assuming that the pedestrian motion model is known in adjacent frames, and Kalman filter is used to predict the pedestrian location. The model of state transition and observation is

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{\omega} \qquad (14)$$

$$\boldsymbol{z}_t = \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{v} \qquad (15)$$

where, $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ denote the state vector and observation vector of pedestrian at the t time, respectively. $\boldsymbol{A}$ and $\boldsymbol{H}$ denote state transfer matrix and observation matrix, respectively. Process noise $\boldsymbol{\omega}$ and observation noise $\boldsymbol{v}$ meet Gaussian noise with covariance $\boldsymbol{Q}$ and $\boldsymbol{R}$, respectively. The state of pedestrian can be defined as

$$\boldsymbol{x} = (c_x, v_x, c_y, v_y)^{\mathrm{T}} \qquad (16)$$

where, c_x and c_y denote the coordinate of the location center point of pedestrian motion, v_x and v_y correspond to the horizontal speed and vertical speed respectively. Defining the observation vector as $\boldsymbol{z} = [c_x, c_y]^{\mathrm{T}}$, and the concrete realization of Kalman filter is described as

$$\begin{cases} \boldsymbol{x}_{t+1|t} = \boldsymbol{A}\boldsymbol{x}_{t|t} \\ \boldsymbol{P}_{t+1|t} = \boldsymbol{A}\boldsymbol{P}_{t|t}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{Q} \\ \boldsymbol{K}_{t+1} = \boldsymbol{P}_{t+1|t}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{P}_{t+1|t}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R})^{-1} \\ \boldsymbol{x}_{t+1|t+1} = \boldsymbol{x}_{t+1|t} + \boldsymbol{K}_{t+1}(\boldsymbol{z}_{t+1} - \boldsymbol{H}\boldsymbol{x}_{t+1|t}) \\ \boldsymbol{P}_{t+1|t+1} = \boldsymbol{P}_{t+1|t} - \boldsymbol{K}_{t+1}\boldsymbol{H}\boldsymbol{P}_{t+1|t} \end{cases} \qquad (17)$$

where, $\boldsymbol{x}_{t+1|t}$ and $\boldsymbol{P}_{t+1|t}$ are the prediction value of pedestrian state and pedestrian state error covariance at the t time. $\boldsymbol{x}_{t|t}$ and $\boldsymbol{P}_{t|t}$ are the estimation value of pedestrian state and pedestrian state error covariance at the $t+1$ time, respectively, $\boldsymbol{K}_{t+1}$ is the filter gain at the $t+1$ time.

2.3 Rematch

In order to rematch pedestrian, it is necessary to consider pedestrian detection method. Due to the slow operation speed, low precision, and poor anti-interference ability, the traditional detection method is extremely limited in practical applications. YOLOv3 is considered as a general object detection algorithm based on deep learning[14], which can determine the spatial location and scale of persons based on the given image. In addition, because of the special algorithm structure of YOLOv3, it has better real-time characteristics.

YOLOv3 is introduced to implement pedestrian detection when the number of consecutive occlusion frames θ_n is more than the threshold θ_t. Defining the output $\boldsymbol{E}$ of detection as

$$\boldsymbol{E} = \{\boldsymbol{e}_\zeta\} \qquad \zeta = 1,2,\cdots,l \qquad (18)$$

where, l denotes the number of persons in the current frame image, $\boldsymbol{e}_\zeta$ denotes the position and scale information of the ζth person.

In order to determine whether there is a tracked target in $\boldsymbol{E}$, it is necessary to extract the deep features of pedestrian to obtain the feature matrix $\hat{\boldsymbol{M}} = \{\boldsymbol{m}_1, \boldsymbol{m}_2, \cdots, \boldsymbol{m}_l\}$. The maximum similarity is computed according to Eq. (12).

$$q' = \max\{\Psi(\boldsymbol{m}_\zeta, \boldsymbol{M}) \mid \boldsymbol{m}_\zeta \in \hat{\boldsymbol{M}}\} \qquad (19)$$

Suppose the maximum similarity q' corresponds to the ζth person. When q' is greater than q_t, the match is considered successful. Otherwise, the match fails.

2.4 The implementation steps of ALPDFE

The flowchart of ALPDFE is shown in Fig. 4, and the implementation is summarized in Algorithm 1.

3 Simulations and analysis

Two scenarios are selected to verify the feasibility and validity of the proposed algorithm based on different occlusion scenes. Video 1 is the Human 3 of visual

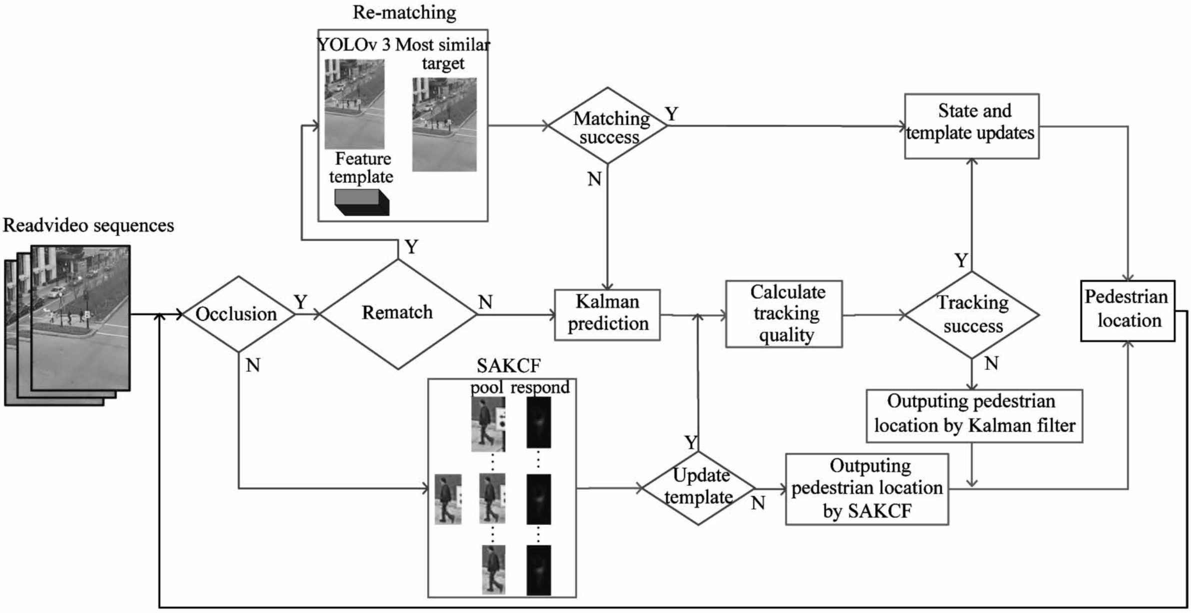


Fig. 4 The flowchart of ALPDFE

Algorithm 1 The implementation of ALPDFE

Input: Video sequence with a total of V frames
Output: Tracks of pedestrian.

```
Step 1:  For $t = 1, 2, \cdots, V$
Step 2:  Read the $t$th frame image.
Step 3:  If $\theta_n == 0$
Step 4:    According to SAKCF, we firstly solve the pedestrian location.
Step 5:    If $(\mathrm{mod}(t, \Upsilon) == 0) \&\& (\theta_n == 0)$
Step 6:      Calculate the tracking quality $q$ by Eq. (12), and then perform Step 11 to Step 14
Step 7:  Else
Step 8:    If $\theta_n > \theta_t$
Step 9:      Use YOLOv3 to solve the feature matrix $\hat{\boldsymbol{M}}$;
Step 10:       Calculate $q'$ by Eq. (19)
Step 11:       If $q' > q_t$ or $q > q_t$
Step 12:         Let $\theta_n$ is equal to zero, we can update SAKCF
                 and the feature template set by Eq. (6) and
                 Eq. (13), respectively. And then go to step 18.
Step 13:       Else
Step 14:         $\theta_n = \theta_n + 1$;
Step 15:         Go to step 17.
Step 16:     Else
Step 17:         Calculate the position of pedestrian by
                 Eq. (17) and solve $q$ by Eq. (12), and then
                 perform Step 11 to Step 14.
Step 18: Get the current prediction position.
Step 19: end
```

tracker benchmark[17]. Video 2 is a real time tracking of pedestrian in the campus. Pedestrians are occluded frequently or completely in the video. Using one pass evaluation (OPE) mode evaluates performance from both qualitative and quantitative aspects. Qualitative analysis is the following 2 aspects: frequent occlusion and long-term occlusion. And Quantitative analysis is from the following 2 aspects: distance accuracy and overlap accuracy. The ALPDFE is compared with SiamFC, KCF, ASAKCF, ALP and SAKCF. And in order to compare the effectiveness of the location prediction step and rematch step, the anti-occlusion pedestrian tracking algorithm based on location prediction (ALP) is introduced into ablation experiment. Unlike ALPDFE, ALP includes only location prediction part by Kalman filter. The simulation parameters are set as follows: the occlusion threshold $\theta_t = 4$, the tracking quality threshold $q_t = 0.8$, the model updating interval $\Upsilon = 10$, the size of feature template $N = 4$.

3.1 Qualitative analysis

(1) Frequent occlusion

Testing video uses a typical Human 3 video segment, pedestrians are frequently occluded by obstacles and reappearance. The tracking effect is shown in Fig. 5. At the 10th frame, the pedestrian is not occluded, and all algorithms can track accurately. At the 35th frame, the pedestrian is occluded by other fast moving pedestrians. As SiamFC lacks the occlusion processing mechanism, first tracking drift phenomenon occurs. However, because the occlusion is small and it has similar appearance color to the pedestrian, the SAKCF and KCF still can continue to track. The pedestrian is occluded frequently from the 60th frame to the 150th frame. It can be seen that both SAKCF and KCF fail to track. At this point, ASAKCF, ALP and ALPDFE still can track continuously. At the 750th frame, ALP and ALPDFE show better tracking performance after a fast focal length change. As the tracking time increases, due to error accumulation and complex background, ASAKCF fails to track. Although ALP does not include the rematch part, it can effectively predict target location by Kalman filter when the frequent occlusion occurs, so it almost does not lose target in the whole tracking process. Because ALPDFE adds the rematch strategy for the phenomenon of tracking failure, pedestrian location is revised constantly. Therefore, ALPDFE has better tracking performance than the other 4 algorithms.

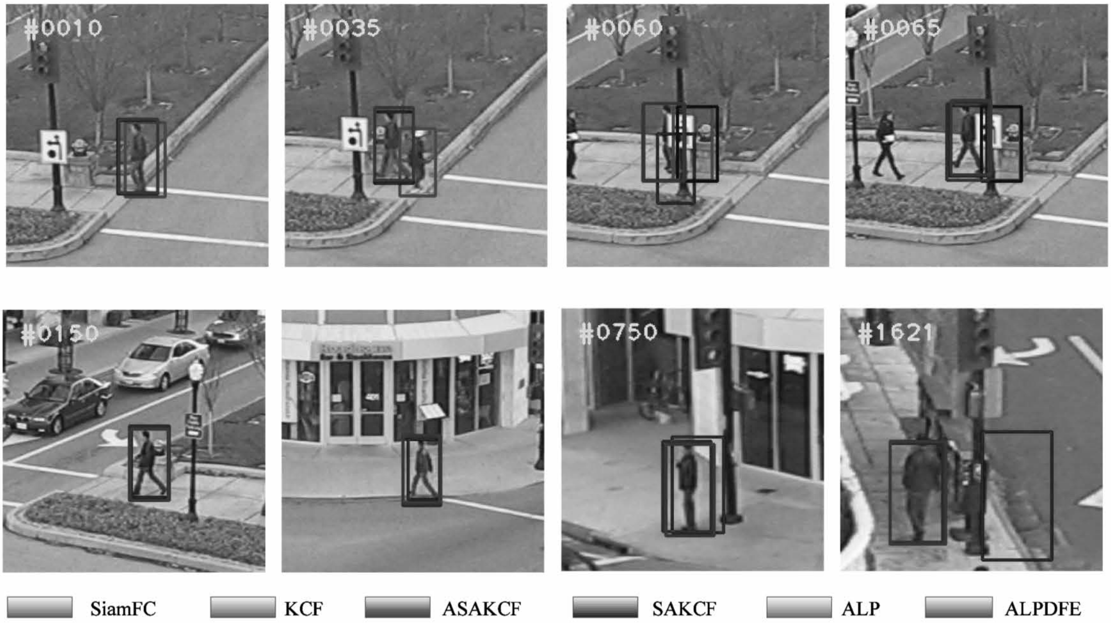


Fig. 5 The tracking result of video 1

(2) Long-term occlusion

Video 2 is a long-term occlusion process for pedestrians. In this video, pedestrians pass through the parking lot and have 100 frames of continuous occlusion. There is interference from vehicles and other moving pedestrians, which increases the difficulty of pedestrian tracking.

The same initial position for the above 6 algorithms in the first frame is set. Their tracking performances are shown in Fig. 6. At the 15th frame, all algorithms are tracking normally. At the 160th frame, pedestrian is occluded, and pedestrian is almost occluded completely at the 174th frame. After that, SiamFC, KCF, ASAKCF and SAKCF cannot continue to track pedestrian. ALP and ALPDFE can predict the pedestrian location by Kalman filter when pedestrian is occluded completely. It can be seen that only ALP and ALPDFE does not lose pedestrian track at the 195th frame. Due to the increase of occlusion time, the phenomenon of tracking drift appears for ALP. The rematch strategy of ALPDFE is executed when a continuous loss occurs, and the pedestrian can be tracked at the 271th frame, the other 5 algorithms fail to track moving pedestrian.

3.2 Quantitative analysis

(1) Center location error

Fig. 7 shows the center location error curve of the above 6 comparison algorithms. In Fig. 7(a), the errors of SiamFC, KCF, ASAKCF, and SAKCF are gradually increased at the 35th, 50th, 50th and 1600th frame, respectively. The center location error of all algorithms is small, and pedestrians can be tracked normally in the first 50 frames. At the 50th frame, pedestrian is occluded, ASAKCF, ALP and ALPDFE can continue to track pedestrian, and other 3 algorithms lead to large location errors. As tracking time goes on, the center location error of ASAKCF also increases obviously at the 1 600th frame, because there is similar interference in the background which indicates the failure of pedestrian tracking. And when there is pedestrian occlusion or pedestrian tracking drift, ALP can use the prediction mechanism of Kalman filter to predict pedestrian position, which maintains the continuity of

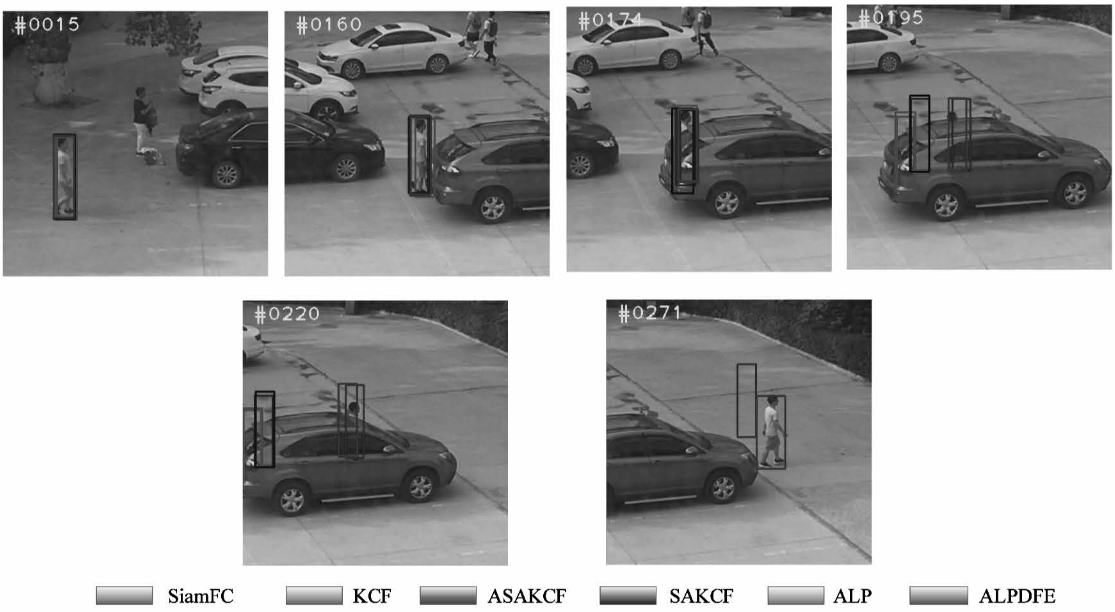
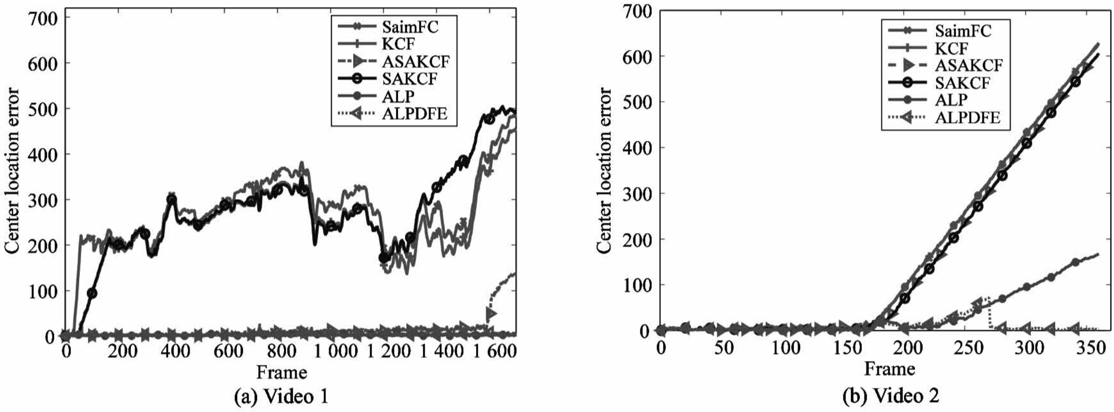


Fig. 6 The tracking result of video 2

tracking to some extent. Compared with ALP, ALPDFE can use not only the prediction mechanism but also the rematch strategy to reposition the pedestrian, so it obtains a low location error. In Fig. 7(b), SiamFC, KCF, ASAKCF and SAKCF have similar location error curves. The above 4 algorithms can track normally when pedestrian is not occluded before the 175th frames. After the 175th frame, SiamFC, KCF, ASAKCF and SAKCF stay in the initial occlusion location when the pedestrian is completely occluded. As the movement of pedestrian is linear approximately, the error curve of center location increases linearly. After the reappearance of pedestrian, pedestrian cannot be tracked again because the pedestrian location is greatly deviated before and after occlusion. ALP and ALPDFE utilize the location prediction by Kalman filter during occlusion, which can reduce the central location error of the occlusion process in the 175th – 270th frames. However, ALP can not rematch pedestrian after a long period of occlusion, Kalman prediction is prone to produce large tracking error. In addition, due to adding the rematch strategy in the ALPDFE, its tracking precision is effectively improved when pedestrian reappears by reducing the central location error after the 270th frame.



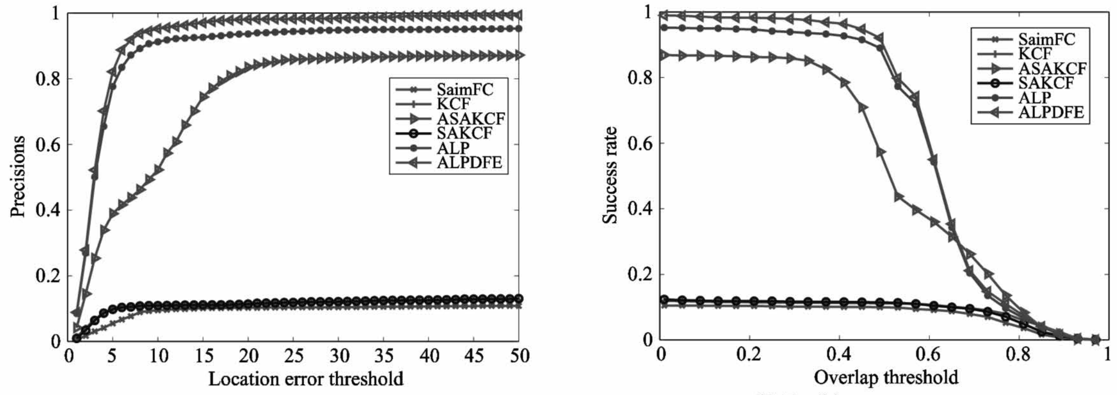
(a) Video 1

(b) Video 2

Fig. 7 Center location error

(2) Precision and success rate of tracking

Tracking precision is calculated according to the ratio which the number of center location error in the error threshold is relative to the total number of frames. Success rate is the ratio which the number of overlap within the threshold to total number of frames. Fig. 8 shows the normalized tracking precision curve and the success rate curve of test video for 6 algorithms. Fig. 8(a) shows the precision curve, it can be seen that the tracking precise of ALPDFE is better than that of ALP. The result also shows that the results in Fig. 8(a) illustrate the effectiveness of rematch strategy in ALPDFE, because the only difference between ALPDFE and ALP is that ALP lacks rematch steps. As the error threshold increases, the precision curve of SiamFC, KCF, SAKCF, ASAKCF, ALP, and ALPDFE gradually increases, then they flattens out at the location thresholds of 10, 5, 5, 20, 10 and 10, respectively. The SiamFC, KCF, and SAKCF locks the judgment and process for occlusion, which cause the failure of pedestrian tracking, or the tracking precision is low. ASAKCF can only deal with partial occlusion problems, and its tracking precision is lower than ALPDFE with rematch strategy. Fig. 8(b) shows the success rate curve of the above 6 algorithms. As the overlap threshold is larger, the requirements for successful tracking are more demanding. Therefore, the success rate curves of SiamFC, KCF, SAKCF, ASAKCF, ALP and ALPDFE show firstly flat and then downward tendency at the overlap thresholds of 0.7, 0.7, 0.7, 0.3 ,0.4 and 0.4, respectively. In Fig. 8(b), it can be seen that the tracking success rate of the ALPDFE is better significantly than other 5 algorithms.



(a) Tracking precision curve

(b) Tracking success rate curve

Fig. 8 Precision and success rate of tracking

4 Conclusions

Aiming to the problem of pedestrian tracking under occlusion, state prediction and rematch strategy is synthetically considered into the design of proposed algorithm. A novel anti-occlusion pedestrian tracking algorithm based on location prediction and deep feature rematch is proposed. In the realization of ALPDFE, the pedestrian appearance features are used to determine whether the occlusion phenomenon of pedestrian exists. When the occlusion phenomenon occurs, Kalman filter is used to predict the pedestrian position. Besides, when pedestrian reappears, pedestrians can be matched and repositioned by using YOLOv3 method. The simulation experiments and theoretical analysis show that ALPDFE can improve the anti-occlusion problem of pedestrian tracking. In addition, the SAKCF can be replaced by other existing trackers in the framework of ALPDFE, and ALPDFE has strong extensibility. In the real scene of pedestrian tracking, we need consider more complex environmental characteristics such as the pedestrian density, the rapid change of pedestrian scale and the lighting change of surrounding environment. Therefore, designing the accurate and rapid scale estimation method and enhancing the anti-interference capability of pedestrian tracking model will be the focus of the follow-up study.

References

[1] Lu M, Xu Y. A survey of object tracking algorithms[J]. *Acta Automatica Simica*, 2019, 45(7): 1245-1259

[2] Zhang X Y, Shi H, Li C, et al. Learning transferable self-attentive representations for action recognition in untrimmed videos with weak supervision[C] // Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, USA, 2019, 33: 9227-9234

[3] Zhang X Y, Li C, Shi H, et al. AdapNet: adaptability decomposing encoder-decoder network for weakly supervised action recognition and localization[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 1: 1-12

[4] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C] // Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 2544-2550

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

[6] Li Y, Zhu J K. A scale adaptive kernel correlation filter

tracker with feature integration[C]// European Conference on Computer Vision, Zurich, Switzerland, 2015: 254-265

[7] Huang Y P, Ju C, Hu X, et al. An anti-occlusion and scale adaptive kernel correlation filter for visual object tracking[J]. *KSII Transactions on Internet and Information Systems*, 2019, 13(4): 2094-2112

[8] Huang S L, Hong J X. Moving object tracking system based on camshift and Kalman filter[C]//Proceedings of International Conference on Consumer Electronics, Communications and Networks, Xianning, China, 2011: 1423-1426

[9] Simon D. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms[J]. *IET Control Theory and Applications*, 2010, 4(8): 1303-1318

[10] Wu H M, Zheng X S. Improved and efficient object tracking algorithm based on Camshift[J]. *Computer Engineering and Applications*, 2009, 45(27): 178-180

[11] Ma C, Huang J B, Yang X K, et al. Adaptive correlation filters with long-term and short-term memory for object tracking[J]. *International Journal of Computer Vision*, 2018, 126(8): 771-796

[12] Wang H Y, Wang L, Yin W R, et al. Multi-scale correlation filtering visual tracking algorithm combined with target detection[J]. *Acta Optica Sinica*, 2019, 39(1): 388-397

[13] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multi- box detector[C]// European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 21-37

[14] Paulius T, Artūras S. Automated image annotation based on YOLOv3[C]//Proceedings of IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering, Vilnius, Lithuania, 2018: 1-3

[15] Wojke N, Bewley A. Deep cosine metric learning for person re-identification[C]// IEEE Winter Conference on Applications of Computer Vision, Lake Tahoe, USA, 2018: 748-756

[16] Danelljan M, Bhat G, Khan F S, et al. Eco: efficient convolution operators for tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 6638-6646

[17] Wu Y, Lim J, Yang M H. Object tracking benchmark [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1834-1848

Hu Zhentao, born in 1979. He received his Ph. D degree in Control Science and Engineering from Northwestern Polytechnical University in 2010. He also received his B. S. and M. S. degrees from Henan University in 2003 and 2006 respectively. Now, he is an assistant professor of College of Computer and Information Engineering, Henan University. His research interests include complex system modeling and estimation, target tracking and particle filter, etc.