



Research on scale adaptive particle filter tracker with feature integration

Yuqi Xiao^{1,2} · Difu Pan¹

Published online: 6 May 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

This research proposes an improved particle filter tracking algorithm based on SGA (the adaptive genetic algorithm supervised by population convergence). In order to improve the robustness and efficiency of the particle filter tracker in various tracking scenarios, this study proposes an adaptive feature selection strategy based on Harris corner detection, SIFT features and colour features. In addition, the tracking frame scale of the traditional target tracking algorithm is fixed in the tracking process, which leads to many problems such as more invalid features and lower positioning accuracy. To solve these problems, this study proposes an adaptive tracking frame scale adjustment model based on the spatial position of particles. Furthermore, considering that the scale adaptive model cannot accurately reflect the target rotation deformation, this paper proposes an adaptive tracking frame scale and direction adjustment model based on the covariance descriptors to accurately track the rotation of the target and further reduce the invalid features of the rectangle frame. The extensive empirical evaluations on the benchmark dataset (OTB2015) and VOT2016 dataset demonstrate that the proposed method is very promising for the various challenging scenarios.

Keywords Computer vision technology · Particle filter · Target tracking · Fusion feature

1 Introduction

Target tracking has been applied in many cutting-edge technologies, such as intelligent traffic supervision, automatic driving, human-computer interaction, crime inspection, weapon guidance and other fields. Visual target tracking is a challenging and valuable study [1–3]; however, it still needs to deal with the following major challenges: 1) a variety of interference factors including low resolution, in-plane rotation, out-of-plane rotation, scale variation, occlusion, deformation, motion blur, fast motion and out-of-view; 2) improvements to the efficiency, accuracy and stability of its tracking performance [4]. This study therefore focuses on addressing the above problems.

In recent years, a great deal of research has been carried out to address these issues. Ref. [5] proposed a mean shift tracking algorithm using 3D colour histograms. This method

overcomes the influence of low illumination and the interaction of similar targets. But it fails to track when the external illumination changes drastically. Ref. [6] proposed a tracking method that uses foreground probability and the histogram of the candidate model in weight to increase the weights of the foreground while suppressing background interference. This method is adapted to changes of background and external illumination. However, it fails to track when the target is occluded or moving fast.

Considering that visual objects often exhibit random behaviour, and target tracking is therefore a nonlinear state estimation problem, particle filters are widely used in visual target tracking because of their great advantages in solving the nonlinear state estimation problem [7, 8]. However, particle filter algorithms still suffer from particle degeneracy and problems with sample impoverishment. Furthermore, their accuracy, efficiency and stability need to be improved when solving complex nonlinear problems [9–11]. For example, Ref. [12] brought forward a labelled particle filter tracking algorithm which describes each image patch with a binary label and can handle real-time fast moving object tracking with less time cost while maintaining high tracking accuracy. However, it fails to track the interaction of similar targets on complicated and variable backgrounds. Ref. [13] presented a tracking method based on mean shift and particle filter (MSPF) which

✉ Yuqi Xiao
xiaoyuqi@csu.edu.cn

¹ School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

² CRRC Zhuzhou Locomotive Co. Ltd., Zhuzhou 412001, China

move particles to local peaks in the likelihood by incorporating the mean shift optimization into particle filtering. This method improves the estimated position of the particle by the mean shift algorithm, which makes the candidate regions of the particles much closer to the true target position. This algorithm has good real-time performance and can handle occlusions and the interaction of similar targets. However, in a low illumination environment, MSPF has a high probability of tracking failure. The tracking efficiency of MSPF also needs to be improved. Ref. [14] proposed an improved particle filter based on PSO (PSO-PF). The algorithm significantly improved the efficiency of the particle filter using the PSO's excellent iteration ability. It was applied to lane detection and human motion tracking respectively, and achieved good tracking results. However, PSO-PF did not solve the problem of sample impoverishment, so the calculation accuracy and stability of the algorithm are low.

In conclusion, most of the traditional tracking algorithms are well suited for certain specific tracking scenarios but lack general adaptability to a variety of scenarios. What's more, in the traditional target tracking algorithm, the size of the target tracking box is determined by the initial size of the target, and the size of the tracking box is fixed in the whole tracking process. But in the actual situation, the size of the target area in the video image is not constant. If the target size becomes larger and beyond the initial size, it will lead to loss of target features, resulting in reduced target location accuracy, and increase the probability of target loss in the tracking process. If the target size becomes smaller than the initial size, the tracking frame will be mixed with excessive background images and useless features. To solve this problem, Ref. [30] proposed a scale adaptive kernel correlation filter tracker. Ref. [31] proposed the DSST method with adaptive multi-scale correlation filters using HOG features to handle the scale variation of target objects. However, both of Ref. [30] and Ref. [31] suffer from the defect of low computational efficiency and need to improve their robustness of various complex interference. Meanwhile, the traditional tracking algorithm can't effectively adapt to deformation and scale changes of the target, and thus results in many problems such as useless features in the tracking box, inaccuracy of target location and higher probability of target loss.

1.1 Motivation

Traditional trackers still suffer from lack of accuracy, efficiency and robustness due to issues mentioned above. Thus the main task of this paper is to solve the drawbacks of traditional trackers and improve the robustness and efficiency of the particle filter tracker in various tracking scenarios. We identify three key factors for solving these issues: sample diversity, target features and frame adjustment model.

Sample diversity: When image blur or serious occlusion happens, the candidates of the target becomes unreliable and the tracking process ends with false recognition. Therefore, it is necessary to enrich sample diversity in the tracking process to deal with the problem of insufficient validity of candidates under corresponding circumstances.

Target features: Some trackers use a single feature, which can achieve better tracking results for specific scenarios, but lack of adaptability for multiple scenarios. Some trackers adopt fusion features, but their real-time performance is poor because of heavy computation.

Frame adjustment model: If the scale of tracking frame can not be adjusted adaptively according to the scale variation of the target, the tracking frame will be mixed with too many invalid features, which will seriously affect the accuracy of target location and lead to an increase in computation.

1.2 Contributions

This paper proposes a scale adaptive particle filter tracker based on SGA (Adaptive genetic algorithm supervised by population convergence) and feature integration. After the resampling progress of PF, the individuals are ranked according to their fitness values and the particles with lower fitness are eliminated and replaced with the same number of particles with higher fitness. Then, genetic operations supervised by population convergence are introduced to PF to enrich its sample diversity and the accuracy. Meanwhile, this paper proposes the HSIFT feature selecting strategy based on the fusion of Harris, SIFT and colour features to improve the tracker's efficiency, accuracy and robustness against multiple tracking interference. Furthermore, this paper proposes a scale adaptive adjustment model of a rectangular box based on the space position of the particles to better adapt to scale changes of the target. Considering that the scale adaptive adjustment model of the rectangular tracking box cannot accurately reflect the rotation deformation of the target, this paper proposes the scale and direction adaptive adjustment model of an elliptical tracking box based on the covariance descriptor.

The main innovations and contributions of this paper are as follows:

- (1) In traditional genetic algorithm, the probability of crossover and mutation are fixed and can not be adjusted properly with the evolution of samples. In this paper, a genetic operating method supervised by population convergence is proposed and applied to the resampling process of particle filter. This adaptive genetic algorithm supervised by population convergence (SGA for short) can greatly improve the sample diversity of our tracker.
- (2) Traditionally, the feature selected by traditional tracking algorithms is for specific tracking scenarios, or the calculation of fusion features is too complex. This problem

greatly weakens the tracking speed, stability or adaptability of traditional tracking algorithms. In this paper, we propose the HSIFT feature extraction method, which performs SIFT operation (scale invariant feature transform) on the features obtained from Harris corner detection. It can select effective target features intelligently based on the degree of similarity between candidate models and target models. The experimental results prove that our tracker based on HSIFT fusion tracking strategy saves tracking time, improves the tracking efficiency and enhances the robustness of the algorithm on the premise of ensuring tracking accuracy and stability. HSIFT feature selection method overcomes the disadvantages of traditional tracking algorithms, such as their inflexible tracking form, and the lack of robustness against multiple tracking interference.

- (3) The tracking frames of traditional trackers are fixed or the trackers can not adapt to the real-time change of the scale of the target. Therefore, a scale adaptive adjustment model based on the spatial position of the particles and an adaptive adjustment model of scale and direction in elliptical tracking box are proposed for the adaptive adjustment of the tracking frame. They can adjust the tracking frame in real time according to the change of target scale or direction of movement. Experiments show these two models can better capture and describe the target scale change or rotation deformation. And they can capture the real area of the target more accurately and help reduce the interference of the invalid feature in the candidate region.

2 Particle filter based on SGA

As mentioned in the introduction, particle filters are widely used in visual target tracking because of their great advantages in solving the nonlinear state estimation problem. However PF still suffer from particle degeneracy and problems with sample impoverishment. In this paper, a genetic operating method supervised by population convergence (SGA for short) is proposed and applied to the resampling process of particle filter to help improve its sample diversity. Then the PF algorithm based on SGA genetic method (SGAPF for short) is developed for our tracker.

2.1 Adaptive genetic algorithm supervised by population convergence (SGA)

The purpose of this section is to solve the problem of particle degradation and sample dilution in particle filter by SGA genetic operation. Our strategy is to filter out particles with weights below the average weight, and then randomly

replicate a corresponding number of particles from the retained particles for the eliminated particles. This method can improve the sample validity, but the sample dilution problem of traditional particle filter algorithm is still unsolved. Therefore, in order to further solve the problem of sample dilution, we do not simply replicate the effective particles to fill the obsolete particles, but do SGA genetic operation on the randomly selected effective particles and then replicate them into the new population.

Definition 1 The average fitness of individuals, which refers to particles in particle filter, is set as f_t (t in f_t means the current moment). f_{\max} represents the optimal individual fitness (fitness function corresponds to the weight function of particles) and f represents the average fitness value of all particles with fitness greater than f_t . The degree of population convergence is defined as $\Delta = f_{\max} - f$. Then the formula for calculating the SGA crossover and mutation probability is:

$$\begin{cases} P_c = -1/(1 + \exp(-k_1 gV)) + 1.5 \\ P_m = -1/(1 + \exp(-k_2 gV)) + 1 \end{cases} \quad (1)$$

where k_1 and k_2 are positive constants, and Δ is non-negative. As a result, the range of the crossover probability P_g is $[0.5, 1]$, while the range of the mutation probability P_m is $[0, 0.5]$. In the traditional algorithm, control parameters such as the crossover and mutation probabilities are independent of the population evolution process, and thus this fixed parameter control mechanism can lead to premature convergence [15, 16]. In the proposed algorithm, however, the crossover probability P_g and mutation probability P_m are adjusted automatically according to the degree of population convergence Δ .

Genetic operator The crossover and mutation of the SGA proposed in this paper are based on arithmetic crossover and non-uniform mutation, as expressed in Eqs. (2) and (3), respectively:

$$\begin{cases} x_{t+1}^i = \alpha \cdot x_t^i + (1-\alpha) \cdot x_t^j \\ x_{t+1}^j = \alpha \cdot x_t^j + (1-\alpha) \cdot x_t^i \end{cases} \quad (2)$$

$$\begin{cases} x_{t+1}^k = \begin{cases} x_t^k + f(t, m_t - x_t^k), p < 0.5 \\ x_t^k - f(t, m_t - l_t), p \geq 0.5 \end{cases}, x_t^k \in [l_t, m_t] \\ f(t, y) = y \cdot (1 - p^{(1-t/T)^b}) \in (0, y), p \in U(0, 1) \end{cases} \quad (3)$$

where α and p are random numbers with the value range $(0, 1)$; x_t^i, x_t^j and x_t^k represent particles that intersect at time t ; x_{t+1}^i and x_{t+1}^j represent new particles generated by crossover operation at time $(t + 1)$; x_{t+1}^k represent new particles generated by non-uniform mutation operation at time $(t + 1)$; T is the maximum number of iterations; parameter b determines the non-uniformity of the mutation operation; and $f()$ is an adaptive mutation operator that can adaptively adjust the step size. $f()$

is used to search potential regions in the whole domain, but in the later stages of iteration only a small neighbouring area of the current solution is searched so as to ensure the accurate positioning and locking of the optimal solution.

2.2 SGAPF algorithm

The calculation accuracy and optimization ability of SGA are better than the traditional GA algorithm, and this paper presents the SGAPF algorithm. Specifically, after the resampling progress of PF, the individuals are ranked according to their fitness values, and the particles with lower fitness are eliminated and replaced with the same number of particles with higher fitness. Then, the SGA arithmetic crossover and non-uniform mutation are introduced to enrich the sample diversity and the accuracy of PF.

(1) Sample initialization

Set S as the sample number and T as the maximum number of iterations. Randomly generate S samples x_0^i ($i = 1, 2, \dots, S$).

(2) Important density sampling [17]

a) Calculation of important density function

$$x_t \sim q(x_t^i | y_t) = p(x_t^i | x_{t-1}^i) p(x_{t-1}^i | y_{t-1}) \tag{4}$$

b) Particle weight update

$$w_t^i = w_{t-1}^i \frac{p(y_t | x_t^i) p(x_t^i | X_{t-1}^i)}{q(x_t^i | X_{t-1}^i, Y_t)} \tag{5}$$

c) Particle probability density update

$$p(x_t | y_{1:t}) = \sum_{i=1}^S w_t^i \delta(x_t - x_t^i) \tag{6}$$

where δ refers to the Dirac function. Firstly, according to Eq. (4), S particles are randomly generated, and the particle weight and probability density are then updated according to Eqs. (5) and (6).

(3) Determination of the degree of particle degeneracy

$$N_{eff} = \frac{1}{\sum_{i=1}^S (w_t^i)^2} \tag{7}$$

N_{eff} represents the degree of particle degradation. If N_{eff} is greater than the threshold N_{th} , the particle degradation is not obvious and we can move directly to Step 5. Otherwise, there may be

serious sample degradation and resampling must be carried out on the current particle set before updating the particle position.

(4) Resampling

All weighted particle weights are reset to $w_t^i = 1/S$.

(5) Adaptive genetic operation

Eliminate individuals with lower fitness and replace them with the same number of individuals with higher fitness. Then, introduce the SGA genetic operation to enrich sample diversity and the accuracy of PF.

(6) Output results

It is necessary to determine whether to stop the iteration according to the iterative termination conditions. If not, then return to Step 2); otherwise, terminate the iteration and output the results as follows:

$$x_t = \sum_{i=1}^S w_t^i x_t^i \tag{8}$$

3 Target tracking algorithm based on SGAPF

When image blur or serious occlusion happens, the candidates of the target becomes unreliable and the tracking process ends with false recognition. Therefore, it is necessary to enrich sample diversity in the tracking process to deal with the problem of insufficient validity of candidates under corresponding circumstances. Based on Section 2, we introduce the SGA genetic method into our tracker to enrich the sample diversity and ensure the validity of candidates. In this part, the SGAPF tracker is introduced in detail from the following aspects: mathematical model of target state, target model, similarity measure, updating of particle weight and target model as well as target location.

3.1 Mathematical model of target state

The mathematical model of the target state is as follows:

$$\begin{cases} X_t = AX_{t-1} + Bu_{t-1} + Cf_{t-1} \\ A = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} T^2/2 & 0 \\ T & 0 \\ 0 & T^2/2 \\ 0 & T \end{bmatrix}, C = \begin{bmatrix} T^2/2 & 0 \\ T & 0 \\ 0 & T^2/2 \\ 0 & T \end{bmatrix} \end{cases} \tag{9}$$

where $X_t = [x_t, x'_t, y_t, y'_t]$ is the coordinate of the motion state at time t ; x and y are the position coordinates; x' and y' are the velocities in the X and Y axis directions respectively [18, 19]; A is the state transition matrix; T is the sampling period; B and

C are input matrixes; and $u_{t-1} = [u_{t-1}^x, u_{t-1}^y]^T$ is the acceleration vector input at time (t-1) (if $u_t=0$, then this is a CV model). f_{t-1} is white Gaussian noise at time t, and the expression for this is given in Eq. (10) [20]:

$$\begin{cases} f_{t-1} = [f_{t-1}^x, f_{t-1}^y]^T \in N(0, Q) \\ Q = \text{diag}(\sigma_x^2, \sigma_y^2) \end{cases} \tag{10}$$

The target observation model is

$$\begin{cases} Y_t = h(X_t) + v_t, \\ h(X_t) = [r_t, \theta_t]^T \\ r_t = \sqrt{x_t^2 + y_t^2}, \\ \theta_t = \arctan\left(\frac{y_t}{x_t}\right) \end{cases} \tag{11}$$

In Eq. (11), $Y_t = [r_t, \theta_t]^T$ represents the distance and phase information of the target, and v_t is white Gaussian noise, which can be expressed as in Eq. (12):

$$\begin{cases} v_t = [v_t^r, v_t^\theta]^T \in N(0, R) \\ R = \text{diag}(\sigma_r^2, \sigma_\theta^2) \end{cases} \tag{12}$$

Target tracking involves estimating the real state of a target using the state model (9) and observation model (11).

3.2 Target model and candidate model

The region of the target in the previous video image is called the target area, and the most likely region of the target in the next frame is called the candidate region. The target and observation models of the system are used to express the characteristic probability density function of the eigenvalues in the target and candidate regions, respectively. The specific expressions of the target model and candidate model are given in Eqs. (13) and (14) [21]:

$$q(x_0) = C \sum_{i=1}^S k\left(\left\|\frac{x_i - x_0}{r}\right\|^2\right) \delta[b(x_i) - u] \tag{13}$$

$$p(y) = C \sum_{i=1}^S k\left(\left\|\frac{x_i - y}{r}\right\|^2\right) \delta[b(x_i) - u] \tag{14}$$

where S refers to the number of feature points, x_0 is the centre of the target region, x_i is the i-th feature point, y represents the centre of the candidate region, r is the bandwidth of the kernel function, C is the normalized constant of the characteristic probability density, and $k(x)$ is the outer contour function of kernel function K and is used to assign a greater weight to the feature point that is closer to the centre of the target model. The feature index interval mapping function b and Dirac function δ are used to determine whether the feature points belong to the u -th feature index intervals (if it belongs to the interval

u , then $\delta = 1$; otherwise, $\delta = 0$). The formula of $k(x)$ and C are as follows:

$$k(x) = \frac{3}{4} (1 - x^2) \tag{15}$$

$$C = 1 / \sum_{i=1}^S k\left(\left\|\frac{x_0 - x_i}{r}\right\|^2\right) \tag{16}$$

3.3 Similarity measure

The process of visual object tracking involves selecting the candidate region which is most similar to the target region of the previous image from all candidate regions of the current frame image. The similarity is measured using the similarity function. In this paper, the Bhattacharyya coefficient is chosen as the similarity function, as shown below [22];

$$\rho(y) \equiv \rho(p(y), q(x_0)) = \sum_{u=1}^S \sqrt{p(y)q(x_0)} \tag{17}$$

where $p(y)$ and $q(x_0)$ are the candidate model and the target model respectively. The larger the similarity $\rho(y)$, the greater the degree of matching of the candidate model and the target model.

3.4 Updating of particle weight and target model

The fitness function of the particle is the probability density function of the measurement, expressed as

$$p(y_t | x_t^i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1 - \rho(y_t)}{2\sigma^2}\right) \tag{18}$$

where σ is the standard deviation of Gaussian density. We can then update the particle weights on the basis of the fitness function:

$$w_t^i = w_{t-1}^i p(y_t | x_t^i) \tag{19}$$

where $w_0^i = 1/S$ refers to the initial weight, and S is the sample capacity. The normalized particle weight is expressed as:

$$\hat{w}_t^i = w_t^i / \sum_{i=1}^S w_t^i \tag{20}$$

The average weight of all particles in the current frame is defined as follows:

$$\bar{w}_t = \sum_{i=1}^S w_i / S \tag{21}$$

The average value of the characteristic probability density of the optimal candidate region in the current frame image is calculated using a normalized particle weight:

$$\bar{p}_t(y) = \sum_{i=1}^s \hat{w}_t^i \times p_t(y) \tag{22}$$

To better integrate the current frame’s optimal candidate model and the existing target model information, this scheme updates the target model according to $\bar{p}_t(y)$, $q_{t-1}(x)$ and \bar{w}_t :

$$q_t(x) = \bar{w}_t \times \bar{p}_t(y) + (1-\bar{w}_t)q_{t-1}(x) \tag{23}$$

where $\bar{p}_t(y)$ refers to the average probability density of the optimal candidate region in the current frame, $q_{t-1}(x)$ is the target model of the previous frame, and \bar{w}_t represents the average weight of all particles in the current frame.

3.5 Target localization and sample updating

The position of the target is determined by the weighted summation of the position of all particles, given by:

$$x_t = \sum_{i=1}^s w_t^i x_t^i \tag{24}$$

After the target is located, the extent of particle degradation is determined by:

$$N_{eff} = 1 / \sum_{i=1}^s (w_t^i)^2 < N_{th} \tag{25}$$

where N_{eff} represents the number of effective particles and N_{th} is the threshold.

If the number of effective particles is smaller than the threshold value, resampling is performed, and the weighted sample set $\{x_t^i | w_t^i\}_{i=1}^s$ is mapped to a sample set with equal weight $\{x_t^i | 1/N\}_{i=1}^s$. Otherwise, the SGA optimization is carried out immediately to form a new population in the next cycle.

4 HSSGAPF target tracking method

The most commonly used features of the traditional target tracking algorithm are the colour, corner, edge, contour and texture of the object. There are different effective features that can be used when encountering similar targets, complex background, target occlusion and so on. But the traditional tracking algorithm can not adapt to the complex tracking situation because of the single form of feature extraction and target tracking.

In this paper, we propose the HSSGAPF target tracking method: 1) First, a feature extraction method (called HSIFT) is proposed, which combines Harris corner detection and SIFT (scale invariant feature transform). 2) Then the SGAPF target tracking algorithm based on HSIFT (HSSGAPF for short) is obtained. Based on the decision of feature similarity between the candidate template and target template, the HSSGAPF algorithm can select appropriate features

according to the actual situation and complete intelligent switching in tracking mode.

4.1 Harris corner detection

Harris is a classic angular point detection algorithm which measures local variation by shifting in various orientations. It has a good performance on stability and robustness, so it has been widely used in the field of image processing. The following images show a Harris corner detection experiment.

As shown in Fig. 1, there are six images. The first one is the original image, the second is the Harris corner detection result of the original picture, and the remaining four images show the experiment results corresponding to the situations in which the original one is enlarged, narrowed, rotated 20 degrees and rotated 130 degrees. Compared with pictures 2, 5 and 6, we can see that the number of effective feature points and the matching locations are almost the same, regardless of the small rotation or large rotation of the image, which means that Harris has good rotation invariance. Compared with pictures 2, 3 and 4, when the scale of image is reduced or enlarged, the number and locations of the feature points are different from the original ones, which indicates that Harris does not have scale invariance.

4.2 SIFT descriptors

SIFT (Scale Invariant Feature Transform) is a theory that can generate feature points that are invariant to location, scale, rotation or translation of the images. [23]. Figure 2 shows a SIFT feature descriptor extraction experiment:

As shown in Fig. 2, from left to right, there are six images. The first is the original image, the second is the SIFT feature extraction result of the original picture, and the remaining four show the experiment results corresponding to the situations in which the original one is enlarged, narrowed, rotated 20 degrees and rotated 130 degrees. Compared with pictures 2, 3 and 4, when the scale of the image is reduced or enlarged, the number and the location of feature points are the same as those of the original picture, which indicates that the SIFT feature has scale invariance. Compared with pictures 2, 5 and 6, we can see that the number of effective feature points and the matching locations are slightly different from those of the original image, which means that the anti-rotating interference ability of SIFT is slightly lower.

4.3 Harris corner features with SIFT descriptors (HSIFT features for short)

In general, the Harris corner detection algorithm has the characteristics of fast feature extraction and high stability, and has rotation invariance, but does not have scale invariance. The SIFT algorithm has scale invariance, but its feature extraction is more complex and computations are large, so its search speed is slower,

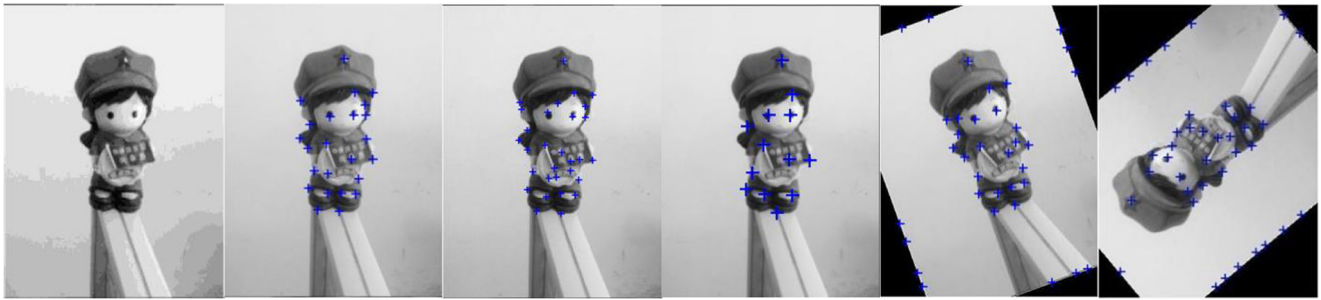


Fig. 1 Harris corner detection experiment

and its real-time performance is not suitable for complex target tracking scenarios. Considering the advantages and disadvantages of the two feature processing methods, this paper proposes the HSIFT feature extraction method, which performs SIFT operation (scale invariant feature transform) on the features obtained from Harris corner detection.

Specifically, we detect the feature points extracted from the Harris corner detection, then assign the main direction according to the SIFT algorithm, generate the feature vector descriptors, and establish the HSIFT operator model. This can not only greatly reduce the time required to extract feature points, but also enhance the robustness of the feature descriptors to image rotation, scale change, illumination condition change and so on.

Meanwhile, the proposed HSIFT feature is fused with colour feature by fuzzy control method to increase the feature recognition ability of the tracker. The similarity measure after the fusion of HSIFT features and colour features is defined as:

$$\rho'(y) = \alpha\rho_{HS}(y) + (1-\alpha)\rho_C(y) \quad (26)$$

where $\rho_{HS}(y)$ and $\rho_C(y)$ respectively represent the HSIFT and colour feature similarity. α is the weight parameter and has been assumed to be a fixed value in much of the literature, but it does not match the actual situation. Considering that the colour features and HSIFT features have different focuses in various kinds of practical situations, this paper uses a fuzzy control method to select the parameters. In this paper, the fuzzy logic rules are chosen to give higher weight values to the reliable features and give smaller weights to the unreliable

ones, as shown in Table 1.

Fuzzy logic consists of fuzzification, fuzzy rule base, fuzzy inference, and defuzzification. The control rule of parameter α in Table 1 is a strategy based on the fuzzy logic rule to make the fusion feature more realistic. In this paper, the fuzzy logic is designed based on fuzzy rule. Fuzzy logic inputs are the similarity of HSIFT and colour feature in the current frame. The feature similarity weight α is calculated by Eq.(26). The fuzzy sets of $\rho_{HS}(y)$ and $\rho_C(y)$ are $\{0.2, 0.4, 0.6, 0.8, 1\}$ and α is $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. According to the fuzzy rules shown in Table 1, we give smaller weight to unreliable features and larger weight to reliable features. When the similarity value of HSIFT and colour are equal, α takes an intermediate value of 0.5. In fact, the similarity value of HSIFT and color features is not constant. The weights based on logic rules designed in this paper can be adjusted in real time according to similarity value, which makes the fusion feature more effective and realistic.

The fuzzy weight selection strategy: When the similarity measure value of colour features and HSIFT features are equal, set α 0.5. Otherwise, the row and column positions of A in Table 1 are determined by $\rho_C(y)$ and $\rho_{HS}(y)$, respectively. Specifically, the row position of α takes the row of the upper limit of the interval where $\rho_C(y)$ is located; similarly, the column position of α takes the column of the upper limit of the interval where $\rho_{HS}(y)$ is located. For example, set $\rho_C(y)$ as 0.31 and $\rho_{HS}(y)$ as 0.72. Then 0.31 is in the interval (0.2,0.4) while 0.72 is in the interval (0.7,0.8). So the row position of α should be 0.4 while the column position of α should be 0.8, and the value of fuzzy weight α should be 0.7. (Fig. 3)

The flow chart of our tracker is as follows:



Fig. 2 SIFT feature descriptor extraction experiment

Table 1 Fuzzy control of α

ρ_{HS}	0.2	0.4	0.6	0.8	1.0
α					
ρ_C					
0.2	0.5	0.6	0.7	0.8	0.9
0.4	0.4	0.5	0.6	0.7	0.8
0.6	0.3	0.4	0.5	0.6	0.7
0.8	0.2	0.3	0.4	0.5	0.6
1.0	0.1	0.2	0.3	0.4	0.5

5 The adaptive adjustment model of the tracking frame

In the traditional target tracking algorithm, the size of the target tracking frame is determined by the initial size of the target, and is fixed throughout the tracking process. Actually, the size of the target area in the video is not invariable. If the target size is larger than the initial size, it will have the effect that the tracking frame can only extract some of the target features, resulting in reduced target location precision and an increased rate of tracking failure. If the target is smaller than the initial size, the tracking frame will include excessive background images and useless features, which will again result in reduced target location precision and an increased rate of tracking failure. In summary, it is necessary to realize the adaptive adjustment of the tracking box.

5.1 Scale adaptive adjustment model of rectangular tracking box

When the target size changes, the average distance D between the target’s central position and particles in the target area will change with the target size and D is positively correlated with the target size, which means D increases with the increase of the target size and D decreases with the decrease of the target size. Accordingly, the size of the tracking box can be adaptively adjusted with D .

$$\begin{cases} d = \frac{1}{M} \sum_{i=1}^M \sqrt{(x-x_i)^2 + (y-y_i)^2} \\ d_x = \frac{1}{M} \sum_{i=1}^M \sqrt{(x-x_i)^2} \\ d_y = \frac{1}{M} \sum_{i=1}^M \sqrt{(y-y_i)^2} \end{cases} \quad w^i > w_T \quad (27)$$

As shown in Eq. (27), set the central position of the target area as (x, y) , then the position of the i -th particle is (x_i, y_i) , and the normalized weight of the i -th particle is $w^i (i = 1, 2, \dots, M)$. M is the number of particles whose weight is larger than the threshold w_T (w_T is set to 0.6 times the average particle

weight in this paper). Based on the particle space location, the adaptive adjustment model of the tracking frame is:

$$\begin{cases} h_{x,t} = h_{x,t-1} \cdot (\alpha \cdot d_t / d_{t-1} + \beta \cdot d_{x,t} / d_{x,t-1}) \\ h_{y,t} = h_{y,t-1} \cdot (\alpha \cdot d_t / d_{t-1} + \beta \cdot d_{y,t} / d_{y,t-1}) \end{cases} \quad (28)$$

where α and β are weighting coefficients. In the paper, α and β are 0.25 and 0.75 respectively. According to the size of the former frame and the d ratio of the former frame and the present frame, the size of the present tracking frame, h , can be obtained by Eq. (28).

5.2 Adaptive adjustment model of scale and direction in elliptical tracking box

The adaptive adjustment model of the rectangular tracking box mainly considers the change of the target in horizontal and vertical directions. But it cannot adapt to the change of angle of the target’s direction of movement. When the target rotation angle is large, the rectangular tracking box will be mixed with much redundant background interference. Therefore, the accurate tracking of rotation deformation should also be fully considered. In this paper, the covariance descriptors are applied to the proposed HSSGAPF tracking algorithm and are applied to the face and head tracking fields.

(1) Elliptical tracking box

Set a and b to be half the axis length of the elliptical tracking frame. The standard elliptic parametric equation is expressed as follows:

$$\begin{cases} x = a \cos \varphi \\ y = b \sin \varphi \end{cases} \quad \varphi \in [0, 2\pi) \quad (29)$$

After rotating the θ angle of the coordinate system, the nonstandard elliptic parametric equation is obtained as follows:

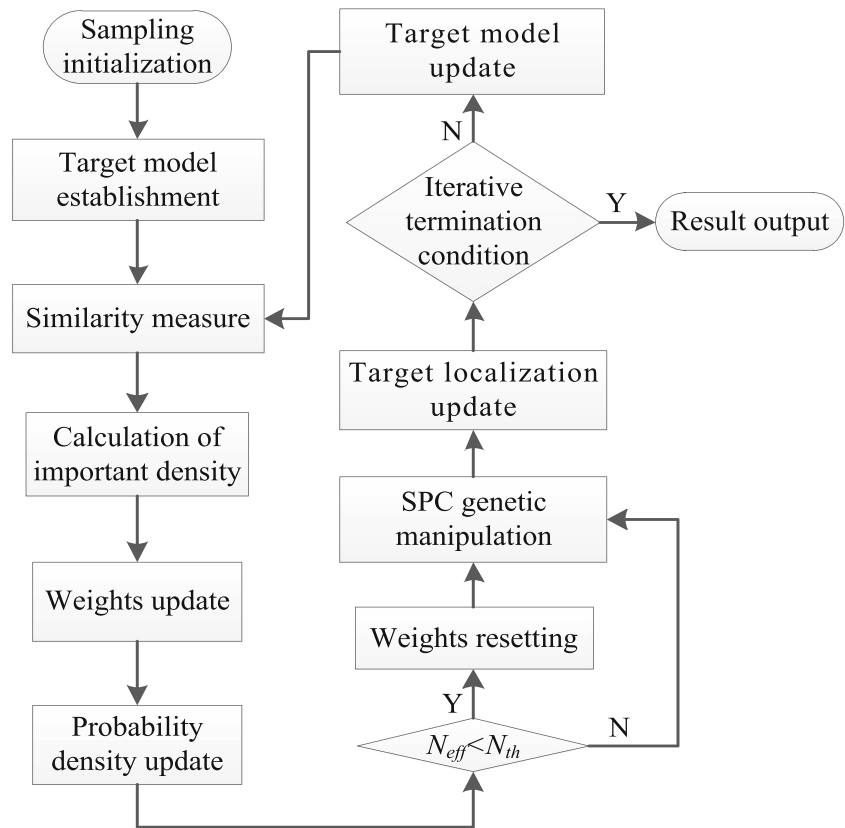
$$\begin{cases} x' = x \cos \theta - y \sin \theta \\ y' = y \cos \theta + x \sin \theta \end{cases} \quad (30)$$

With the combination of Eq. (29) and Eq. (30), the non-standard elliptic parametric equation obtained after the coordinate system rotates through the θ angle can be expressed as:

$$\begin{cases} x' = a \cos \varphi \cos \theta - b \sin \varphi \sin \theta \\ y' = b \sin \varphi \cos \theta + a \cos \varphi \sin \theta \end{cases} \quad (31)$$

For any point in the nonstandard ellipse, it can be regarded as an ellipse similar to the ellipse in the same centre, so the

Fig. 3 Flow chart of our tracker



parameter equation of the whole elliptic region contained in the nonstandard ellipse can be expressed as:

$$\begin{cases} x_s = (a\cos\varphi\cos\theta - b\sin\varphi\sin\theta)r_s \\ y_s = (b\sin\varphi\cos\theta + a\cos\varphi\sin\theta)r_s \\ r_s \in [0, 1], \varphi \in [0, 2\pi) \end{cases} \quad (32)$$

(2) Covariance descriptor

The size of the target image I is set to $W \times H$. Set f as the vector of fusion feature that combines the d types of the image I : $f(x, y) = \varphi(I, d, x, y)$, where $f()$ is a mapping function of colour, texture, edge and other features and d is the number of features. Set R as the image area of the nonstandard elliptic tracking box. Let (x_0, y_0) be the central coordinate of R . Let a and b be the length of the long and short half axes of the ellipse region R , respectively. The expression of the covariance descriptor C_R of R is as follows:

$$\begin{cases} C_R = \frac{\sum [f(x, y) - \mu_R][f(x, y) - \mu_R]^T}{W \times H} \\ x = x_0 + (a\cos\varphi\cos\theta - b\sin\varphi\sin\theta)r_s \\ y = y_0 + (b\sin\varphi\cos\theta + a\cos\varphi\sin\theta)r_s \end{cases} \quad (33)$$

where μ_R represents the mean value of the feature vectors.

(3) The distance measure of covariance descriptors

The covariance descriptor is a vector on the Riemann manifold, and the similarity between the elements is measured by the distance from the Riemann manifold. The distance $d(C_{R1}, C_{R2})$ of the two covariance descriptors on the Riemann manifold can be obtained from the $d \times d$ covariance matrix C_{R1} and C_{R2} :

$$d(C_{R1}, C_{R2}) = \sqrt{\sum_{i=1}^d \ln^2 \lambda_i(C_{R1}, C_{R2})} \quad (34)$$

where $\lambda_i(C_{R1}, C_{R2})$ is the generalized eigenvalue which satisfies the condition of $C_{R1}x = \lambda C_{R2}x$.

In the particle filter tracking algorithm based on the covariance descriptor, the expression of the particle's observation probability function (fitness function) is as follows:

$$p(y_i | x_i^j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{d(C_{R1}, C_{R2})}{2\sigma^2}\right) \quad (35)$$

where σ is the scale coordinate of the DOG (difference-of-Gaussian) space.

(4) The elliptical Gaussian difference operator

The standard Gaussian kernel function:

$$\begin{cases} D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \\ L(x, y, k\sigma) = G(x, y, k\sigma) \otimes I(x, y) \\ G(x, y, \sigma) = 1/2\pi\sigma^2 \cdot e^{-(x^2+y^2)/2\sigma^2} \end{cases} \quad (36)$$

For the elliptical tracking box, the standard Gaussian kernel function $G(x, y, \sigma)$ in Eq. (36) should be replaced by the elliptical Gaussian kernel function $G(x, y, \theta, \sigma_a, \sigma_b)$.

$$G(x, y, \theta, \sigma_a, \sigma_b) = \frac{\exp\left(-\frac{1}{2}[x \ y] \Sigma^{-1}(\theta, \sigma_a, \sigma_b) \begin{bmatrix} x \\ y \end{bmatrix}\right)}{2\pi\sigma_a\sigma_b} \quad (37)$$

The matrix $\Sigma^{-1}(\theta, \sigma_a, \sigma_b)$ is obtained with the direction angle θ of the long axis of the ellipse, the long half axis scale σ_a , and the short half axis scale σ_b :

$$\Sigma^{-1}(\theta, \sigma_a, \sigma_b) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad (38)$$

Similarly, the standard differential Gaussian function $D(x, y, \sigma)$ in Eq. (36) should be replaced by the elliptical differential Gaussian function $D(x, y, \theta, \sigma_a, \sigma_b)$:

$$D(x, y, \theta, \sigma_a, \sigma_b) = DoG(x, y, \theta, \sigma_a, \sigma_b) \otimes I(x, y) \quad (39)$$

$$DoG(x, y, \theta, \sigma_a, \sigma_b) = [G(x, y, \theta, k\sigma_a, k\sigma_b) - G(x, y, \theta, \sigma_a, \sigma_b)] \quad (40)$$

where $DoG(x, y, \theta, \sigma_a, \sigma_b)$ is the regular elliptical Gaussian difference operator.

(5) Adaptive adjustment of the scale and direction of the elliptical tracking box

The adaptive adjustment of the elliptical tracking box consists of four parts: the calculation of the centre position of the ellipse, the calculation of the feature scale, the calculation of the scale of the long half axis and the short half axis of the elliptical tracking box, as well as the calculation of the direction angle of the elliptical long axis.

a) Computing centre position:

The central position of the elliptical centre of the t time (frame T) $(x_{t,c}, y_{t,c})$ is determined by the zero order moment and the first moment of the target region image $I(x, y)$:

$$x_{c,t} = \frac{M_{10}}{M_{00}}, y_{c,t} = \frac{M_{01}}{M_{00}} \quad (41)$$

The expressions of zero order moment M_{00} , first order M_{01} and M_{10} are as follows:

$$\begin{cases} M_{00} = \sum_i^S w_i, \\ M_{10} = \sum_i^S x_i w_i, \\ M_{01} = \sum_i^S y_i w_i \end{cases} \quad (42)$$

where w_i is the weight of the i-th particle.

b) Calculating the feature scale of ellipse:

The relationship between the feature scale σ_t , the long half axis $\sigma_{t,a}$ and the short half axis $\sigma_{t,b}$ at time t are as follows:

$$\begin{cases} \sigma_t^2 = \sigma_{t,a}\sigma_{t,b} \\ \sigma_{t,a} = \sqrt{2}\sigma_t \\ \sigma_{t,b} = \sigma_t/\sqrt{2} \end{cases} \quad (43)$$

The value range of a is $[\sigma_{t, \min}, \sigma_{t, \max}]$, $\sigma_{t, \min}$ and $\sigma_{t, \max}$ are calculated as follows:

$$\begin{cases} \sigma_{\min} = \sqrt{M_{00}}/2 \\ \sigma_{\max} = \sqrt{r_{t,1}^2 + r_{t,2}^2} \end{cases} \quad (44)$$

where $r_{t,1}$ and $r_{t,2}$ are half of the width and height of the kernel function window at time t. In the range of a values $[\sigma_{t, \min}, \sigma_{t, \max}]$, the scale is increased from the minimum scale $\sigma_{t, \min}$ to the maximum scale $\sigma_{t, \max}$, and the interval of each increment is 1.

$D(x, y, \sigma)$ is carried out at an extreme point (x_0, y_0, σ) for Taylor expansion:

$$D(x, y, \sigma) = D(x_0, y_0, \sigma) + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \quad (45)$$

Calculate the value of the Gaussian difference function $D_t(x, y, \sigma)$ according to Formula (45) before each scale increase. The optimal feature scale a for the T frame corresponds to the scale that maximizes the value of $D_t(x, y, \sigma)$:

$$\sigma_t = \arg \max_{\sigma_t} D_t(x, y, \sigma_t) \quad (46)$$

c) Calculating the long and short half axis of the ellipse tracking box:

Obtain the feature scale σ_t by (45) and (46). Then calculate the long and short half axis of the ellipse tracking box with Eq. (43).

d) Calculation of the inclination angle of the long half axis of the ellipse tracking box:

The value range of the inclination angle θ_t at time t is determined by the inclination angle of the $(t-1)$ moment θ_{t-1} : $[\theta_{t-1} - 10^\circ, \theta_{t-1} + 10^\circ]$. Let the inclination angle increase from $(\theta_{t-1} - 10^\circ)$ to $(\theta_{t-1} + 10^\circ)$ with the interval of 1° . The Gaussian difference function $D_t(x, y, \theta_t, \sigma_{t,a}, \sigma_{t,b})$ is calculated in terms of Formulas (39) and (40) before each increment. The optimal inclination angle of the long ellipse axis of the ellipse in frame T corresponds to the inclination angle that maximizes the value of $D_t(x, y, \theta_t, \sigma_{t,a}, \sigma_{t,b})$, and the expression of θ_t is as follows:

$$\theta_t = \arg \max_{\theta_t} D_t(x, y, \theta_t, \sigma_{t,a}, \sigma_{t,b}) \quad (47)$$

5.3 Experimental analysis between the adaptive tracking models

In order to compare the actual tracking effect of the elliptical adaptive tracking model and rectangular adaptive tracking model, we choose the following four cases of a face tracking experiment: face rotation in the vertical plane, face rotation in the horizontal plane, complete occlusion and similar interference. The contrast is shown in Figs. 4, 5, 6 and 7.

Tracking effect analysis and performance evaluation: From Figs. 4 to 7, we can see that both the elliptical adaptive tracking model and the rectangular adaptive tracking model can achieve stable tracking of face targets in various situations, such as occlusion, similarity interference, scaling and deformation of targets, rotation deformation and so on. However, compared to the rectangular adaptive tracking model, the elliptical adaptive tracking model can not only adapt to the scale changes of the target and achieve adaptive adjustment of the tracking box scale, but can also adapt to the rotation of the target and adaptively adjust the inclination angle of the tracking box. Therefore, it can better capture and describe the target scale change and rotation deformation. The elliptical adaptive tracking model enables the tracking box to capture the real area of the target more accurately and reduces the interference of the invalid feature in the candidate region. So it is better to use the elliptical adaptive tracking model when rotational deformation occurs.



Fig. 4 Tracking effect of human face rotation in the vertical plane

6 Experimental analysis

For experimental validation, we use the tracking benchmark dataset (OTB2015) [32]. We compare our tracker with eight state-of-the-art trackers including the C-COT [24], SRDCFdecon [25], MUSTer [26], BACF [27] and LMCF [28], Staple [29], SAMF [30], DSST [31]. To better evaluate and analyze the strength and weakness of the tracking approaches, we evaluate the trackers with 11 attributes based on different challenging factors including low resolution, in-plane rotation, out-of-plane rotation, scale variation, occlusion, deformation, motion blur, fast motion and out-of-view.

6.1 Qualitative comparisons

In qualitative comparisons, eight challenging sequences are selected to evaluate our tracker intuitively. The results are in Fig. 8, and nine different colors indicate different trackers. The trackers are qualitatively compared in the following aspects:

1. Illumination variation: Taking the video of “Car2” for an example, when the illumination changes dramatically, all trackers can successfully track the target. However, only our tracking box stay fit with the target all the time.
2. Fast motion: For the target in “Biker”, the motion of the target is very fast. C-COT, Ours, SRDCFdecon, BACF, MUSTer and Staple can always track the target correctly, but the others lose the target easily.
3. Motion blur: Taking the video of “Ironman” for an example, the motion of the target causes the blur of the target region. The SAMF, DSST, SAMF and Staple fail to track the head of Ironman when serious motion blur happens, while Ours, C-COT, MUSTer and SRDCFdecon can track the target from beginning to the end.
4. Deformation: For the target in “Dancer”, target deformation occurs. C-COT, Ours, SRDCFdecon and Staple lose less part of the target than the others. Meanwhile only our tracker can always track most part of the target.
5. Background clutters: In “Football1”, the tracking drift arises in DSST, SAMF, LMCF, Struck and MUSTer when a background similar to the target appears in the tracking region, such as #59 and #70.

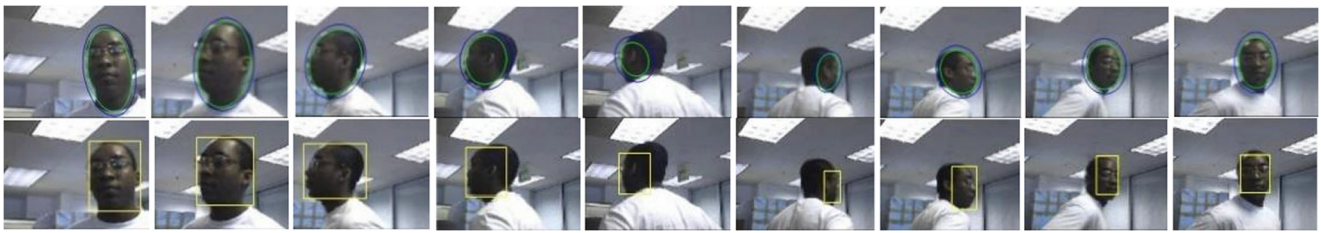


Fig. 5 Tracking effect of human face rotation in the horizontal plane

6. Low resolution: For targets of low resolution, such as “Freeman4”, the feature of the target is too small to extract. Due to the Harris-SIFT feature extracting strategy, our tracker captures more available features and performs more robust than the others.
7. Scale variation: In “Singer1”, due to the application of SIFT feature, our tracker can adapt better to the scale variation of the target, while serious tracking drifting appears in MUSTer, SAMF, BACF and DSST.
8. Rotation: It is divided into in-plane and out-of-plane rotation. Both rotations are in “Skater” in which C-COT and ours track the target consistently and the sizes of the bounding boxes match the target well.
9. Occlusion: In “Skating2”, it indicates the full or partial occlusion of the target by the other target. Only ours, C-COT and SRDCFdecon can track the target immediately and accurately from beginning to the end.

6.2 Quantitative comparisons

To evaluate our tracker comprehensively and reliably, we choose success rate and precision rate to carry out quantitative analysis.

1. Success rate: Given a threshold t_0 , the tracker is considered to succeed if and only if the overlap rate α is greater than t_0 . The success rate is defined as the percentage of the successful frames, and the larger value indicates a better performance of the tracker.
2. Precision: Precision shows the ratio of frames whose CLE (center location error) is within a given threshold, and the larger value indicates a better performance of the tracker.

In our experiments, we quantitatively compare the trackers from two aspects: the overall performance and the attribute-based performance for OTB2015 [32].

6.2.1 Overall performance for 51 sequences

For evaluating our tracker’s overall performance for OTB2015 [32], we plot the success plots and the precision plots of the above 9 trackers. The success plot shows the success rates at a varied overlap threshold t_0 in the interval $[0,1]$, and the precision plot shows the precisions at a varied CLE threshold from 0 to 50 pixels. The overall performance comparisons of the trackers are shown in Fig. 9:

The above experimental data illustrate that our tracker outperforms these state-of-art trackers and achieves satisfactory tracking results in different challenging scenarios.

6.2.2 Attribute-based performance for 51 sequences

To further analyse the performance of our tracker under different tracking conditions, we evaluate the above nine trackers on 11 attributes, which are defined in [32]. The success plots and precision plots on different attributes are shown in Figs. 10 and 11, respectively.

Among the 11 attributes, no matter in the success plot or in the precision plot, our tracker ranks first in the six attributes, the top two in the eight attributes and the top three in all attributes. It can be obtained from these attribute-based data that our tracker achieves a good performance in most attributions. But our tracker cannot perform as well as C-COT and SRDCFdecon on several attributes, such as occlusion in the success plots and out-of-plane rotation in the precision plots. These problems may be the next research areas for improving our tracker.



Fig. 6 Tracking effect of human face under severe occlusion



Fig. 7 Tracking effect of human face with similar target interference

6.3 Tracking speed of tracker

We implemented our tracker in MATLAB on a machine equipped with Intel Core i-7-6700 @3.40GHz and 64 GB RAM. On our

experimental platform, our tracker achieves a practical tracking speed of an average of 83 frame per second (FPS). Table 2 shows the tracking speed of the above 9 trackers. All the data are published by the authors in their papers. The “-” indicates that the

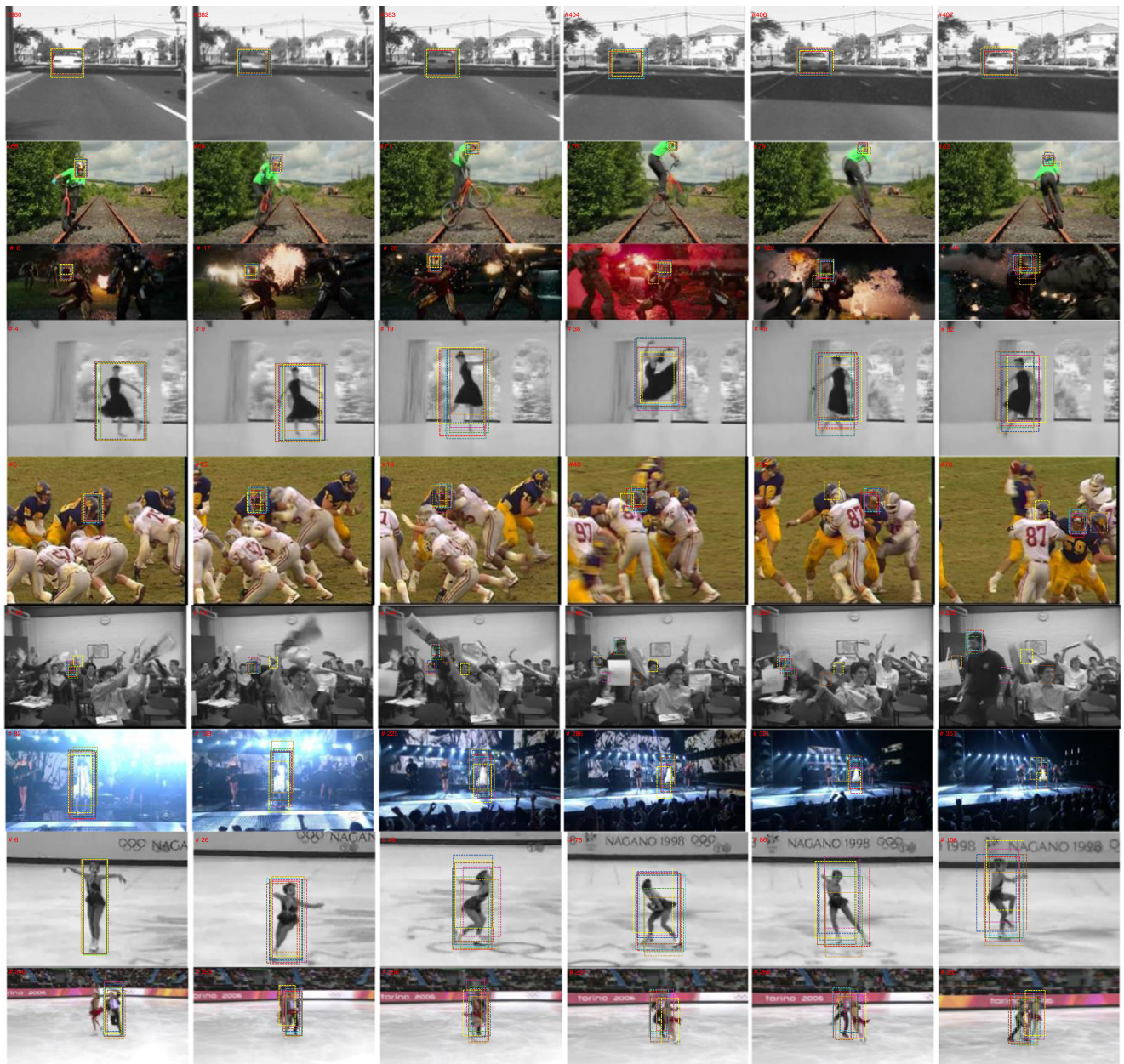


Fig. 8 Qualitative comparisons of 9 trackers (denoted in different colors) on nine challenging sequences (from top to bottom are Car2, Biker, Ironman, Dancer, Football11, Freeman4, Singer1, Skater and Skating2)

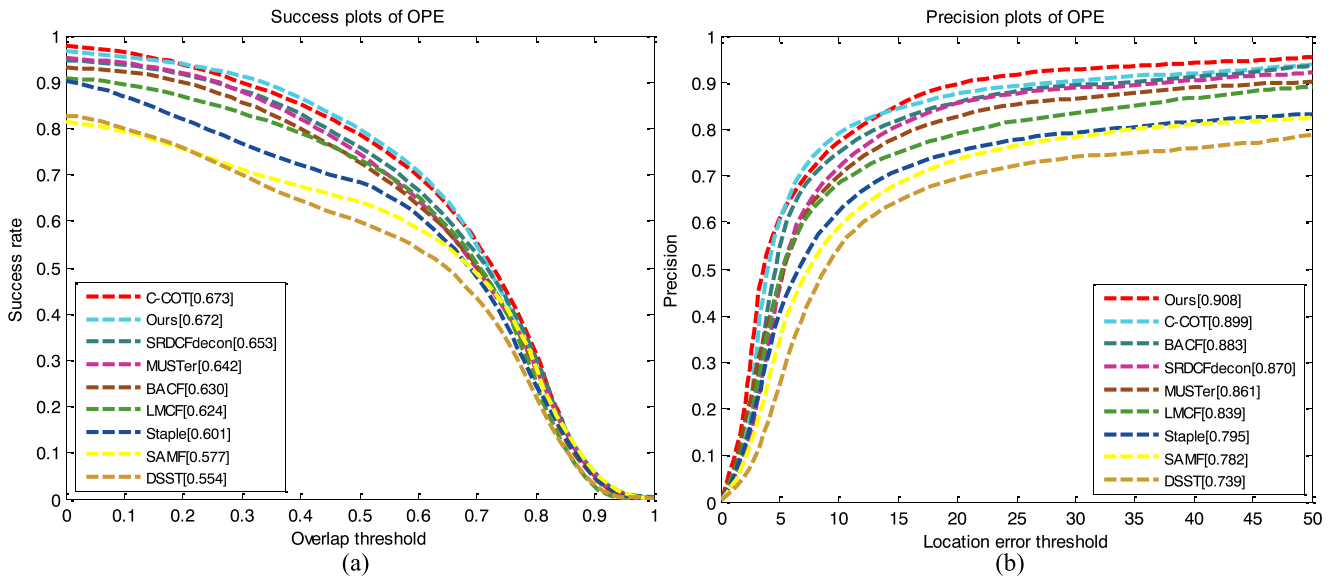


Fig. 9 The success plots and precision plots of OPE for the trackers: (a) success plots and (b) precision plots

author does not give the tracking speed explicitly. From Table 2, we can see that LMCf has the highest tracking speed, and our tracker achieves a faster speed than the other 5 trackers.

The reasons why the proposed method is faster than other methods are analyzed as follows:

- (1) Firstly, the particle filter algorithm has the advantages of low hardware dependence and high computational efficiency, and it is recognized that the particle filter framework is suitable for solving non-linear problems such as target tracking.

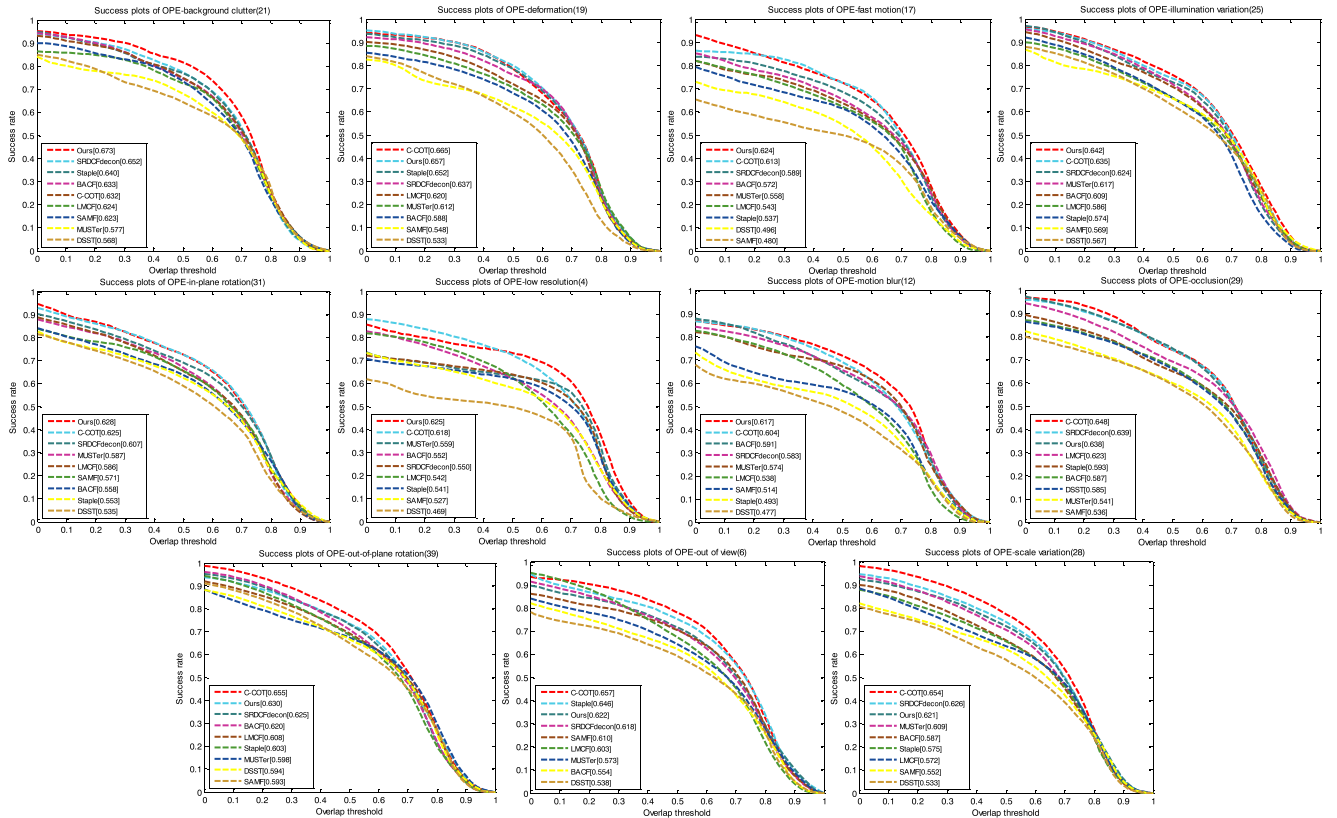


Fig. 10 The success plots of OPE for the trackers on different attributes

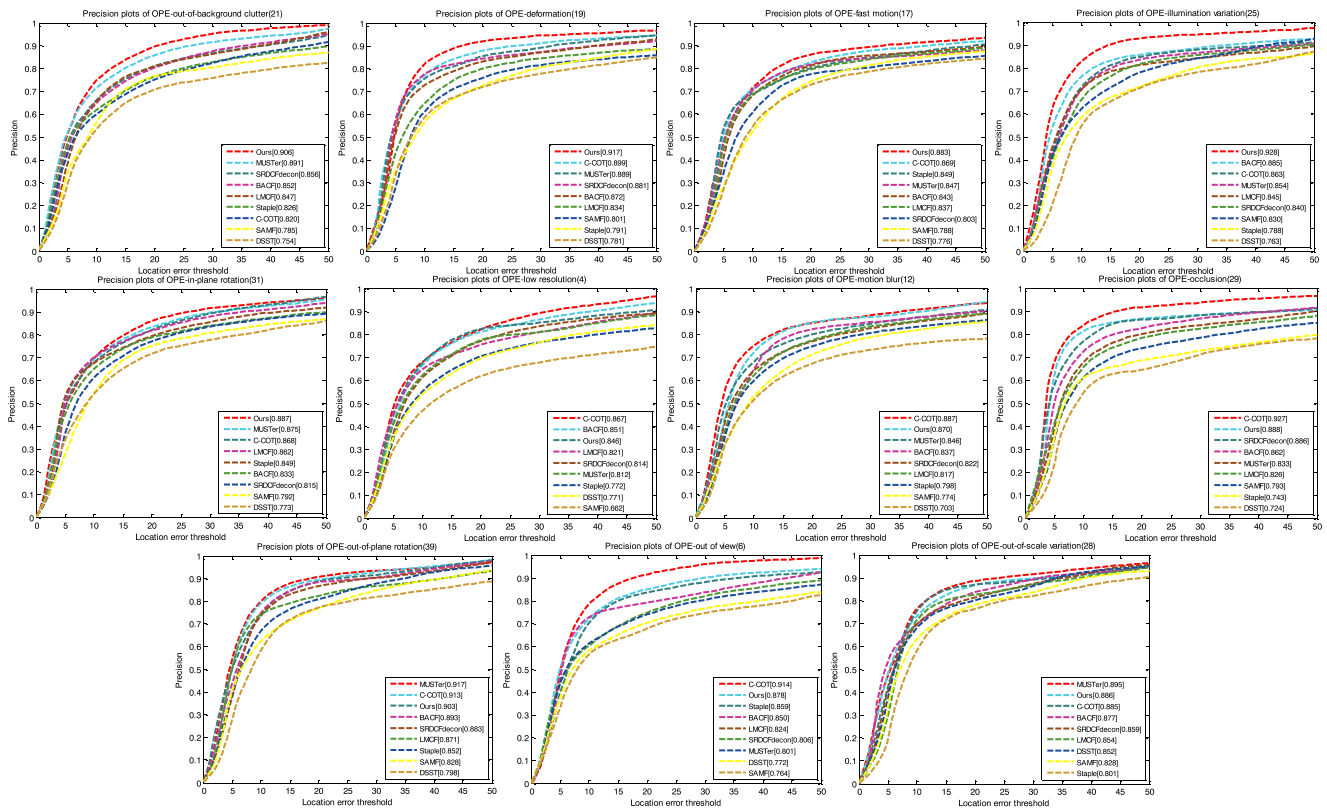


Fig. 11 The precision plots of OPE for the trackers on different attributes

- (2) Secondly, a lot of tracking methods, including particle filter, face the problem of sample impoverishment. This weakens the searching ability of the trackers a great extent, and affects the improvement of computing efficiency of them. While the proposed adaptive genetic algorithm supervised by population convergence (SGA for short) can greatly improve the sample diversity of our tracker.
- (3) Traditionally, the feature selected by traditional tracking algorithms is for specific tracking scenarios, or the calculation of fusion features is too complex. This issue greatly weakens the tracking speed, stability or adaptability of traditional tracking algorithms. In this paper, we propose the HSIFT feature extraction method, which performs SIFT operation (scale invariant feature transform) on the features obtained from Harris corner detection. It can select effective target features intelligently based on the degree of similarity between candidate models and target models. The HSIFT feature selection method

greatly improves the efficiency, accuracy and robustness of our tracker.

- (4) The adaptive adjustment model of tracking frame proposed in this paper effectively weakens the influence of invalid background and invalid feature interference on our tracker, and helps to improve its efficiency of feature extraction and target tracking.

6.4 Statistical comparison

In order to further prove the robustness and stability of the proposed algorithm, a statistical comparison of the above nine trackers is performed using the VOT2016 dataset [33] (To make a more comprehensive comparison, this section adds three other algorithms, ECO [1], TCNN [34] and SSAT [33], which are outstanding in VOT2016 dataset.). Specifically, on the basis of VOT2016 dataset, the trackers are statistically compared.

In Table 3 we compare our approach, in terms of average overlap (EAO for short), accuracy (Acc. for short),

Table 2 The tracking speed comparison for the 9 trackers

Tracker	C-COT	Ours	SRDCFdecon	MUSTer	BACF	LMCF	Staple	SAMF	DSST
FPS	–	83②	3	–	35	85①	80③	7	24

Table 3 Statistical comparison

Trackers	ECO	Ours	C-COT	TCNN	SSAT	SRDCF	MUSTer	BACF	LMCF	Staple	SAMF	DSST
EAO	0.374 ^①	0.296 ^⑤	0.331 ^②	0.325 ^③	0.321 ^④	0.247	0.258	0.236	0.212	0.295	0.186	0.181
Acc.	0.540 ^⑤	0.542 ^④	0.539	0.554 ^②	0.577 ^①	0.535	0.526	0.529	0.538	0.544 ^③	0.507	0.533
R.Fail.	0.202 ^①	0.253 ^③	0.238 ^②	0.268 ^④	0.291 ^⑤	0.419	0.438	0.513	0.453	0.378	0.587	0.704
EFO	4.530 ^⑤	12.55 ^②	0.507	1.049	0.475	1.990	2.312	3.152	5.021 ^④	11.140 ^③	4.099	12.747 ^①

robustness(R. for short, failure rate) and speed (in EFO units), with the top-ranked trackers in VOT2016 challenge. We can see that our tracker ranks fifth in EAO, fourth in ACC., third in R. and second in EFO.

Among the top five trackers in the challenge, there is a small difference in their EAO, R.Fail. and Acc. score. Specifically, the EAO score of our method is 26.4% less than the best EAO score, the accuracy score of our method is 6.5% less than the best accuracy score, and the R. Fail. of ours is 25.3% larger than the least failure rate. As for EFO scores, our tracker achieves a significant performance improvement over the other four trackers. The first-ranked performer, ECO, is based on correlation filter tracking framework and provides an EFO score of 4.530. Our approach achieves an almost 3-fold speed up in EFO compared to ECO.

In a word, compared with the state-of-the-art tracking trackers, our tracker has no obvious weaknesses and performs well in all four aspects of statistical comparison. On the basis of guaranteeing the accuracy and robustness of the algorithm, our tracker is particularly outstanding in EFO performance.

7 Conclusion

This paper presented a robust visual tracker based on the framework of Particle Filters (PF). We proposed an adaptive genetic algorithm supervised by the degree of population convergence to enrich the sample diversity and the accuracy of PF. Then the powerful features including Harris corner detection, SIFT and colour features are fused together to further boost the overall performance for PF tracker based on SGA. Moreover, we proposed two kinds of models for the adaptive adjustment of the tracking frame, which can better capture and describe the target scale change and rotation deformation and help reduce the interference of the invalid feature in the candidate region of our tracker. Extensive experimental results on the benchmark dataset (OTB2015) and VOT2016 dataset demonstrate the effectiveness and

robustness of our tracker against the state-of-the-art tracking methods.

References

1. Danelljan M, Bhat G, Khan FS, et al (2017) ECO: Efficient Convolution Operators for Tracking. 2017 IEEE Conference on Computer Vision and Pattern Recognition
2. Fujita H, Cimr D (2019) Computer aided detection for fibrillations and flutters using deep convolutional neural network. *Inf Sci* 486: 231–239
3. Chongsheng Z, Changchang L, Xiangliang Z, George A (2017) An up-to-date comparison of state-of-the-art classification algorithms. *Expert Syst Appl* 82:128–150
4. Gonczarek A, Tomczak JM (2016) Articulated tracking with manifold regularized particle filter. Springer-Verlag New York, Inc.
5. An X, Kim J, Han Y (2014) Optimal colour-based mean shift algorithm for tracking objects. *IET Comput Vis* 8(3):235–244
6. Zhou Z, Zhou M, Shi X (2016) Target tracking based on foreground probability. *Multimed Tools Appl* 75(6):3145–3160
7. Zhang X, Peng J, Yu W, Lin KC (2012) Confidence-level-based new adaptive particle filter for nonlinear object tracking. *Int J Adv Robot Syst* 9(1)
8. Li T, Sun S, Sattar TP, Corchado JM (2013) Fight sample degeneracy and impoverishment in particle filters: a review of intelligent approaches. *Expert Syst Appl* 41(8):3944–3954
9. Lin SD, Lin JJ, Chuang CY (2015) Particle filter with occlusion handling for visual tracking. *Image Processing IET* 9(11):959–968
10. Chen P, Qian H, Wang W, Zhu M (2011) Contour tracking using Gaussian particle filter. *IET Image Process* 5(5):440–447
11. Rymut B, Kwolek B, Krzeszowski T (2013) GPU-accelerated human motion tracking using particle filter combined with PSO. In: *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer-Verlag, New York, pp 426–437
12. Yang J et al (2015) Fast Object Tracking Employing Labelled Particle Filter for Thermal Infrared Imager. *International Journal of Distributed Sensor Networks* 2015:2
13. Wei Q, Dai T, Ma T, Liu Y, Gu Y (2016) Crystal identification in dual-layer-offset doi-pet detectors using stratified peak tracking based on svd and mean-shift algorithm. *IEEE Trans Nucl Sci* 63(5):2502–2508
14. Wang X et al (2010) Annealed particle filter based on particle swarm optimization for articulated three-dimensional human motion tracking. *Opt Eng* 49(1)
15. Dai CH, Zhu YF, Chen WR (2006) Adaptive probabilities of crossover and mutation in genetic algorithms based on cloud model.

- Information Theory Workshop, 2006. ITW '06 Chengdu. IEEE 24: 710–713
16. Hong TP, Wang HS, Lin WY, Lee WY (2002) Evolution of appropriate crossover and mutation operators in a genetic process. *Appl Intell* 16(1):7–17
 17. Mallah R, Quintero A, Farooq B (2017) Distributed classification of urban congestion using VANET. *IEEE Trans Intell Transp Syst* 18(9):2435–2442
 18. Azab MM, Shedeed HA, Hussein AS (2014) New technique for online object tracking-by-detection in video. *IET Image Process* 8(12):794–803
 19. Tanzmeister G, Wollherr D (2017) Evidential grid-based tracking and mapping. *IEEE Trans Intell Transp Syst* 18(6):1454–1467
 20. Shmaliy YS (2012) Suboptimal FIR filtering of nonlinear models in additive white Gaussian noise. *IEEE Trans Signal Process* 60(10): 5519–5527
 21. Ning J, Zhang L, Zhang D, Wu C (2010) Robust mean-shift tracking with corrected background-weighted histogram. *IET Comput Vis* 6(1):62–69
 22. Sun J (2010) Object tracking using an adaptive Kalman filter combined with mean shift. *Opt Eng* 49(2):020503-1–020503-3
 23. Lowe DG (1999) Object Recognition from Local Scale-Invariant Features. *ICCV IEEE Computer Society* 2:1150
 24. Danelljan M, Robinson A., Khan FS, Felsberg M (2016) Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking. European Conference on Computer Vision. Springer, Cham. 472–488
 25. Danelljan M, Häger G, Khan FS, Felsberg M (2016) Adaptive Decontamination of the Training Set: A Unified Formulation for Discriminative Visual Tracking. *Computer Vision and Pattern Recognition. IEEE*:1430–1438
 26. Hong Z, Chen Z, Wang C et al (2015) Multi-Store Tracker (MUSTer): A cognitive psychology inspired approach to object tracking. *Computer Vision and Pattern Recognition. IEEE*:749–758
 27. Galoogahi HK, Fagg A, Lucey S (2017) Learning Background-Aware Correlation Filters for Visual Tracking. *IEEE Computer Society*:1144–1152
 28. Wang M, Liu Y, Huang Z (2017) Large Margin Object Tracking with Circulant Feature Maps. *IEEE Computer Society*:4800–4808
 29. Bertinetto L, Valmadre J, Golodetz S, Miksik O, Torr P (2015) Staple: complementary learners for real-time tracking. *Proc IEEE Conf Comput Vis Pattern Recognit* 38(2):1401–1409
 30. Li Y, Zhu J (2014) A Scale Adaptive Kernel Correlation Filter Tracker with Feature Integration. *ECCV Workshops*
 31. Danelljan M, Häger G, Khan FS, Felsberg M (2014) Accurate Scale Estimation for Robust Visual Tracking. *British Machine Vision Conference*:65.1–65.11
 32. Wu Y, Lim J, Yang MH (2013) Online object tracking: a benchmark. *Proc IEEE Computer Vision and Pattern Recognition* 9: 2411–2418
 33. Kristan M, Leonardis A, Matas J, Felsberg M, Pflugfelder R, Čehovin L et al (2016) The visual object tracking vot2016 challenge results. In: *ECCV workshop*
 34. Nam H, Baek M, Han B (2016) Modeling and propagating CNNs in a tree structure for visual tracking. *arXiv preprint arXiv: 1608.07242*

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Yuqi Xiao was born in Yiyang, Hunan Province, China in 1991. He received the B.E. and M.E. degrees in transportation engineering from Central South University, China, respectively, in 2013 and 2016. He is currently working toward the Ph.D. degree with the Key Laboratory of Rail Traffic Safety of Ministry of Education, Central South University. His research interest includes traffic efficiency, and safety applications.



Difu Pan was born in Xingning, Guangdong Province, China in 1957. He received the M.E. degree in railway traction electrification and automation from Southwest Jiao Tong University, China, in 1988. He is a Full Professor with the Department of Transportation Engineering, Central South University, China. He has co-authored over 60 technical publications, including journal and proceedings papers.