International Journal of Automation and Computing

17(5), October 2020, 652-666
DOI: 10.1007/s11633-020-1235-2

Robust Object Tracking via Information Theoretic Measures

Wei-Ning Wang[1,2] Qi Li[1,2,3] Liang Wang[1,2]

[1] Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100190, China

[3] Artificial Intelligence Research, Chinese Academy of Sciences, Qingdao 266300, China

Abstract: Object tracking is a very important topic in the field of computer vision. Many sophisticated appearance models have been proposed. Among them, the trackers based on holistic appearance information provide a compact notion of the tracked object and thus are robust to appearance variations under a small amount of noise. However, in practice, the tracked objects are often corrupted by complex noises (e.g., partial occlusions, illumination variations) so that the original appearance-based trackers become less effective. This paper presents a correntropy-based robust holistic tracking algorithm to deal with various noises. Then, a half-quadratic algorithm is carefully employed to minimize the correntropy-based objective function. Based on the proposed information theoretic algorithm, we design a simple and effective template update scheme for object tracking. Experimental results on publicly available videos demonstrate that the proposed tracker outperforms other popular tracking algorithms.

Keywords: Object tracking, information theoretic measures, correntropy, template update, robust to complex noises.

1 Introduction

Object tracking is a very important topic in computer vision. It aims to estimate the spatial state of a moving target in a video sequence[1–9]. With an object track in the first frame identified, the tracking problem is usually formulated as automatically tracking the trajectory of the object over the subsequent frames. It has been widely applied in many real world problems, such as vehicle navigation and video surveillance. However, accurate tracking of general objects under complex scenarios is still difficult due to partial occlusions, illumination variations, abrupt object motions, cluttered backgrounds, etc. Tremendous efforts in object tracking have been made to tackle these problems in recent years[10–15].

Deep learning based methods have shown superior performance over traditional methods on object detection, object segmentation, object recognition[16–20], etc. They have also been widely used for tracking[17, 21–23]. Although deep learning based tracking algorithms have achieved big breakthroughs in recent years, they still suffer from heavy computational cost and limited training data. In this paper, we mainly focus on traditional tracking algorithms. There are generally two major categories in traditional tracking techniques: generative and discriminative methods. Generative tracking methods usually use an appearance model to represent the tracked object and seek the most likely target candidates based on reconstruction errors. The essence of the general trackers is to search for target candidates that are the most similar to the object. Inspired by recent advances in sparse coding and compressive sensing, some popular generative trackers are proposed which include l_1 tracker (APGL1)[24], low rank sparse tracker (LRST)[25, 26], multi-task tracking (MTT)[14], incremental subspace learning (IVT)[27], consistent low rank sparse tracker (CLRST)[4] and structural sparse tracker (SST)[7]. APGL1 assumes that the tracked candidate can be represented by a sparse linear combination of both target templates and trivial templates. LRST resorts to the inherent low-rank structure of particle representations while MTT employs the sparsity-inducing mixed norm to enforce sparsity and learns the particle representation together. IVT tries to learn the principal component analysis (PCA)-based appearance model incrementally during the tracking process. CLRST can prune and select particles adaptively under the particle filter framework for tracking. SST exploits the relationship among particles via low rank sparse learning. Many extensions[6, 28, 29] have been proposed to address the time efficiency and template update scheme.

On the other hand, the discriminative tracking approaches cast the tracking problem as a binary classifica-

Research Article
Manuscript received February 10, 2020; accepted April 20, 2020; published online May 30, 2020
Recommended by Associate Editor Hui Yu

tion problem. It aims to find the target location that best separates the target from its background. Popular discriminative trackers include the multiple instance learning (MIL) tracker[30], structural output tracker (Struck)[5], online random forest tracker (ORT)[31], structural correlation filter (SCF)[32], etc. Similar to object detection, online multiple instance learning is employed in an MIL tracker. A structured output support vector machine is used for adaptive tracking in Struck. Hough forests are used in ORT for online visual tracking. SCF fuses part-based tracking strategy into a correlation filter and exploits circular shifts of all parts to preserve target object structure. An empirical comparison of different trackers refers to [33–35].

A tracking method usually consists of three parts: an observation model, a dynamic model and a search strategy[4, 27, 36]. The observation model is used to evaluate the likelihood of an observed image patch belonging to the object class, while the dynamic model describes the state of an object over time. The search strategy seeks the most likely states in the current frame. In this paper, we mainly address the partial occlusion and abrupt motion problems in tracking that is related to observation models. To deal with partial occlusions and abrupt motions, many sophisticated appearance models have been proposed through statistical analysis, model analysis and sparse representation. Among them, the trackers based on holistic appearance information provide a compact notion of the tracked object rather than treating the object as a set of independent pixels. Thus, it is more robust to appearance variations. Eigentracking[37] is one of the early works using a low dimensional subspace method for robust object tracking. One drawback of the tracker is that its template is from a large set of training images and will not be updated during the tracking process. Ross et al.[27] further developed an incremental visual learning (IVT) algorithm to handle pose variations, shape deformations and camera motions. IVT assumes the tracked target is generated from a low dimensional PCA subspace plus a Gaussian distributed error term. Compared with Eigentracking, it doesn't need the training phase and learns the eigenbasis online during the tracking process. IVT is effective in handling appearance variations caused by illumination and pose. However, it is not robust to some challenging scenarios due to the formulation based on construction error with Gaussian noise assumption and the update scheme without detecting outliers.

Some important issues regarding the holistic appearance-based tracker include how to measure reconstruction errors and how to update the template during the tracking process. The representation coefficients of a holistic appearance-based tracker are often obtained by least-squares solutions with the Gaussian distribution prior. Then l_2 norm is used to measure the reconstruction errors with the obtained representation coefficients. New observations are simply decided from new templates without detecting outliers. While in many real-world applications, the error term is much more complex and it is not appropriate to assume that the error term is the Gaussian distribution. l_1 regularization about the error term is introduced in [38] to deal with the Laplacian noise. It combines the IVT algorithm with recent sparse representation schemes for learning effective appearance models. Experimental results in [38] have shown that the l_1 norm better fits sparse noise than the l_2 norm. A least soft-threshold squares tracking is proposed in [39], which is used to handle both Gaussian and Laplacian noise. Both the robust particle representation and the robust template update are presented in [40] based on the Huber loss function. Then the Huber loss function is relaxed to a weighted least squares problem. One drawback of the algorithm in [40] is the computational cost is much higher compared with [38, 39]. One limitations of these tracking algorithms is that they are only robust to one or two type of various noises, e.g., Gaussian or Laplacian noise. To the best of our knowledge, there are no general frameworks developed to address the role of robust error functions in the holistic appearance-based tracking algorithms.

In this paper, we propose a robust holistic appearance-based tracker to address the challenging factors in object tracking. Our tracker employs a correntropy-based nonconvex loss function that has been introduced in information theoretic learning to handle the complex noises[41, 42]. Fig. 1 shows the flowchart of the proposed algorithm. The contributions of our algorithm are summarized as follows. First, a correntropy-based robust object tracking method is proposed to deal with complex noises caused by occlusions, abrupt motions, etc. A half-quadratic algorithm is employed to solve the general robust tracking problem. Second, a novel dynamic template update scheme is presented to capture the appearance variations of the tracked object and to ensure that outliers are properly identified. It is proven to be a simple and effective solution for updating the tracking template. Experimental results show that our algorithm outperforms other popular tracking algorithms on several benchmark video sequences.

The rest of this paper is organized as follows. In Section 2, we review some related work about online subspace learning. The concept of correntropy and its properties are introduced in Section 3. The details of our algorithm are presented in Section 4. Section 5 provides a series of experiments to systematically evaluate the effectiveness of the proposed methods, prior to the summary of this paper in Section 6.

2 Related work

Our work is motivated in part by the online subspace learning for robust object tracking[27, 43, 44]. The basic idea of these methods is to use an online update method for

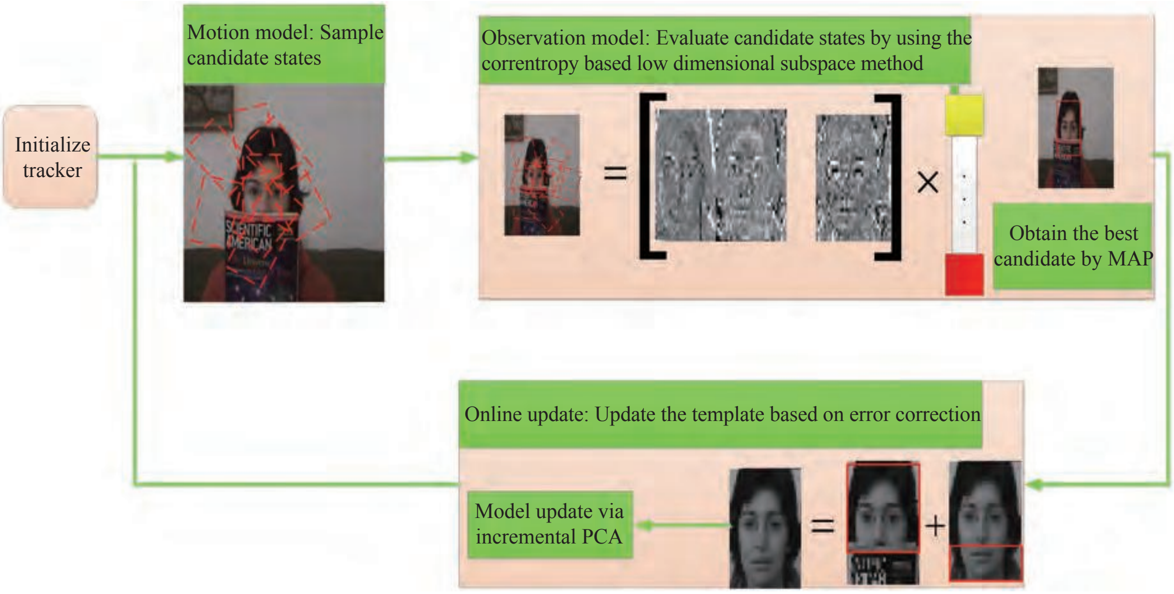

Initialize tracker
Motion model: Sample candidate states
Observation model: Evaluate candidate states by using the correntropy based low dimensional subspace method
=
×
Obtain the best candidate by MAP
Online update: Update the template based on error correction
Model update via incremental PCA
=
+

Fig. 1 The flowchart of our algorithm. It consists of three main parts. First, some candidate states around previous tracking results are sampled using Brownian model. Then an observation model based on low dimensional subspace method is adopted to obtain the best candidate. Finally, a novel online update scheme is presented to update the tracking template based on the information theoretic measures

learning and updating a PCA subspace. Hence, the success of these methods mainly benefits from the power of subspace representations as appearance models and the adaptability of on-line update strategies. Given an image patch y predicted by X, the online subspace learning methods assume y is generated from a low dimensional subspace with Gaussian distribution[27]:

$$p\left(y|X\right)=N\left(y;\mu,UU^{\rm T}+\varepsilon I\right) \tag{1}$$

where p denotes the probability that y predicted by X, U represents a matrix of column basis vector, I is an identity matrix, μ is the mean value, and εI corresponds to the additive Gaussian noise term in the observation process. The assumption is reasonable when the error term is Gaussian distributed with small variations. However, we are not sure about the error term in real world applications. Many solutions have been proposed to deal with different types of noises. For example, Wang et al.[38, 39] models the error term using the Gaussian-Laplacian distributed error term to alleviate the partial occlusion problem.

Our work also has some relationship to the sparsity-based object tracking. Sparse representation has been successfully applied in computer vision. Inspired by [45], sparsity-based trackers[6, 14, 24, 25, 46] have attracted much attention in recent years. The basic assumption is that the observation vector can be represented by a sparse linear combination of the trivial templates:

$$y=T\alpha+\varepsilon \tag{2}$$

where y denotes the observation vector, T is the target template, α is the corresponding coefficient, and ε is the noise term. α and ε are usually regarded as sparse terms. The sparsity-based trackers have achieved much progress in object tracking. However, they are computationally expensive, and how to efficiently handle the noise term is still an open problem.

3 The concept of correntropy

Similar to [45, 47], we use a linear regression model for a series of observations:

$$y=Wx+e \tag{3}$$

where $y\in\mathbf{R}^{d\times1}$ is an observation matrix, $W=[w_1,\cdots,w_d]\in\mathbf{R}^{d\times k}$ denotes the input data matrix, $x=[x_1,\cdots,x_k]\in\mathbf{R}^{k\times1}$ means the unknown representation coefficient, $e=[e_1,\cdots,e_d]\in\mathbf{R}^{d\times1}$ can be seen as an error term. The representation coefficient can be solved by maximizing the posteriori probability with the uniform prior: $\hat{x}=\arg\max\limits_{x}p\left(y|x\right)=\arg\max\limits_{x}p\left(e\right)$, which is also called the maximum likelihood estimation. A natural assumption about the error term $e_1,\cdots,e_d$ is that it follows independent and identically zero mean Gaussian distribution. Then the likelihood function of the estimator is: $p\left(e\right)=\prod_{i=1}^{d}p\left(e_i\right)$, where $e_i\in N\left(0,\sigma_N^2\right)$. Maximizing the log likelihood function of the model is equivalent to minimizing the following objective function:

$$-\sum\nolimits_{i=1}^{d}\log\left[\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}\left(y_i-W^{\rm T}x_i\right)^2\right)\right]= \sum\nolimits_{i=1}^{d}\left(y_i-W^{\rm T}x_i\right)^2+\frac{d}{2}\log\left(2\pi\sigma^2\right). \tag{4}$$

Equation (4) can be solved by minimizing the following least squares problem:

$$\min_{x} \frac{1}{2} \|y - Wx\|_2^2 . \tag{5}$$

In [38, 39], the error vector is modeled as an additive combination of two independent components: Gaussian and Laplacian noise vectors. Maximizing the likelihood function finally turns into the following optimization problem:

$$\min_{x,s} \frac{1}{2} \|y - Wx - s\|_2^2 + \lambda\|s\|_1 \tag{6}$$

where s corresponds to the Laplacian noise term.

In many real world applications, data samples usually suffer from the unpredictable errors caused by the noise and outliers. Wang et al.[40] uses the Huber loss function to model the error term. The assumption of the error term is Gaussian-Laplacian distributed. Recently, the concept of correntropy was widely used[41, 48] to deal with non-Gaussian noise and impulsive noise. He et al.[42] proposed a novel correntropy-based face recognition method, which is robust to occlusion, clutter and illumination changes. The methods proposed by Chen and Principe[48] are quite effective in dealing with Gaussian and non-Gaussian noise. Correntropy is the probability of how similar two random variables are in a joint space, which is controlled by the kernel bandwidth. It is defined as a local similarity measure between two variables A and B:

$$V_\sigma (A, B) = E [k_\sigma (A - B)] \tag{7}$$

where $k_\sigma (\cdot)$ is a kernel function that satisfies mercer theory, $E (\cdot)$ is the expectation operator. Fig. 2 shows the comparison of different loss functions. As shown in Fig. 2, a new metric has been introduced by correntropy. It is similar to the l_2 norm distance when the data samples become close. Then when the data samples get further, it performs similar to l_1 norm distance. Finally when the data samples are far away, it approaches the l_0 norm. This geometric meaning interprets the robustness of correntropy for outlier rejection. It is symmetric, positive and bounded with a theoretic foundation. The kernel bandwidth provides an effective way to eliminate the detrimental effect of outliers. It is different from the use of a threshold in previous methods[41]. Based on (7), Liu et al.[41] further extended the correntropy for a general similarity measurement between two random vectors: the correntropy induced metric (CIM). For two vectors $A = (a_1, \cdots , a_d)^{\mathrm{T}}$ and $B = (b_1, \cdots , b_d)^{\mathrm{T}}$, CIM is defined as

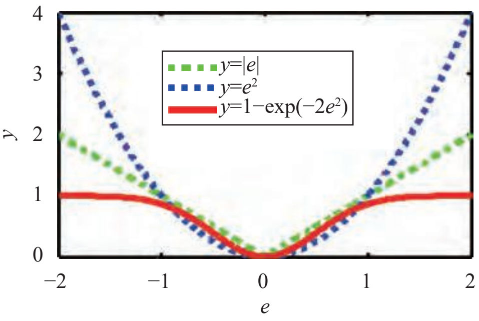


Fig. 2 Comparison of different loss functions. Compared with l_2 or l_1 loss functions, correntropy-based loss function is more robust to various noises.

$$CIM (A, B) = \left(g (0) - \frac{1}{d} \sum\nolimits_{i=1}^{d} g (e_i)\right)^{\frac{1}{2}} = \left(g (0) - \frac{1}{d} \sum\nolimits_{i=1}^{d} g (a_i - b_i)\right)^{\frac{1}{2}} \tag{8}$$

where the error term is defined as $e_i = a_i - b_i$. Compared with mean square and Huber-based loss functions, correntropy is more robust to outliers inherited from the advantage of the robust local metric.

4 Proposed model and algorithm

4.1 Markov model for object tracking

Object tracking is regarded as an inference task using a hidden Markov model in this paper. Given a set of observed images $y^{1:t} = \{y^1, \cdots , y^t\}$, object tracking aims to estimate the value of hidden state variable x^t, which corresponds to the affine parameters of the target at time t based on the observations to the previous time step. Suppose that x_i^t represents the i-th candidate sample of the state x^t, then the most probable hidden state variable can be obtained by maximizing a posteriori estimation:

$$\hat{x}^t = \arg\max_{x_i^t} p \left(x_i^t|y^{1:t}\right) . \tag{9}$$

Utilizing Bayes theorem, we have

$$p \left(x^t|y^{1:t}\right) \propto p \left(y^t|x^t\right) \int p \left(x^t|x^{t-1}\right) p \left(x^{t-1}|y^{1:t-1}\right) \mathrm{d}x^{t-1} \tag{10}$$

where $p \left(x^t|x^{t-1}\right)$ is the dynamic model. It represents the state transition probability between two frames. $p \left(y^t|x^t\right)$ is the observation model, which estimates the likelihood of an image y^t belonging to the state variable x^t. The dynamic model is formulated by Brownian model: Each parameter in x^t is modeled by a Gaussian distribution around the previous state variable x^{t-1}, i.e., $p \left(x^t|x^{t-1}\right) = N \left(x^t; x^{t-1}, \Sigma\right)$, where Σ is a diagonal covariance matrix that indicates the variance of state transition parameters. The tracking process is mainly governed by the observation model and the dynamic model. While the dynamic model is usually fixed, the key step for a robust object tracking algorithm is determined by the observation model.

4.2 Correntropy-based observation model

Similar to [27, 38], the tracked target object can be represented by a robust PCA subspace:

$$\phi\left(y-Uz-\mu\right) \tag{11}$$

where $y \in \mathbf{R}^{d\times 1}$ is a target tracking object, $U \in \mathbf{R}^{d\times k}$ means a PCA basis matrix, $z \in \mathbf{R}^{k\times 1}$ denotes the corresponding coefficients with respect to the PCA basis matrix, $\mu \in \mathbf{R}^{d\times 1}$ represents the average vector, $\phi\left(\cdot\right)$ is a robust loss function. In this paper, we use a correntropy-based function $\phi\left(x\right)=1-\exp\left(-x^2/\sigma^2\right)$. The correntropy-based methods treat each individual pixel differently and puts emphasis on those pixels corresponding to the same class as target tracking object y. That means if there are noises and outliers in the target tracking object y, they will have small contributions to the correntropy. Hence various noises can be handled uniformly under this framework.

Equation (11) is a nonconvex formulation and doesn't have a closed form solution. According to the conjugate function theory and half-quadratic theory[49], we can use the additive form of half-quadratic algorithm to solve this problem. Lemma 1 can be used for optimizing $\phi\left(\cdot\right)$ in a half-quadratic way.

Lemma 1. Suppose that $\phi\left(\cdot\right)$ is a potential loss function that satisfies certain conditions, then there exists a dual potential function for a fixed x: $\psi\left(\cdot\right)$, such that $\phi\left(x\right)=\inf\limits_{s\in\mathbf{R}}\left\{\frac{1}{2}\left(x\sqrt{c}-\frac{p}{\sqrt{c}}\right)^2+\psi\left(p\right)\right\}$, where p is an auxiliary variable. It is determined by a minimizer function $\delta\left(\cdot\right)$ with respect to $\phi\left(\cdot\right)$.

Some of the functions $\phi\left(\cdot\right)$ and their minimizer functions $\delta\left(\cdot\right)$ are listed in Table 1. The correntropy-based observation model is defined as

$$1-\exp\left(-\left(y-Uz-\mu\right)^2/\sigma^2\right). \tag{12}$$

According to Lemma 1, the augmented cost-function of (12) reads as

$$J\left(z,p\right)=\arg\min_{z,p}\frac{1}{2}\left\|y-Uz-\mu-p\right\|_2^2+\sum\nolimits_{i=1}^{d}\psi\left(p_i\right) \tag{13}$$

where auxiliary variable p is determined by the minimization function $\delta\left(\cdot\right)$ with respect to $\phi\left(\cdot\right)$. Because the auxiliary variables are determined by their minimizer functions, when the auxiliary variables are fixed, the analytic forms of $\psi\left(\cdot\right)$ in (13) can be removed.

Based on the half-quadratic optimization theory, (13) can be alternately minimized as follows:

$$\begin{aligned}p^{t+1}&=\left(y-Uz^t-\mu\right)\left\{1-\exp\left(-\frac{\left(y-Uz^t-\mu\right)^2}{\sigma^2}\right)\right\}\\ z^{t+1}&=\arg\min_{z}\frac{1}{2}\left\|y-Uz-\mu-p^{t+1}\right\|_2^2+\sum\nolimits_{i=1}^{d}\psi\left(p_i^{t+1}\right)\end{aligned} \tag{14}$$

where t is the iteration number. Note that the basis matrix U comes from the PCA subspace and thus it is orthogonal. Equation (14) can be further optimized as follows,

$$\begin{aligned}p^{t+1}&=\left(y-Uz^t-\mu\right)\left\{1-\exp\left(-\frac{\left(y-Uz^t-\mu\right)^2}{\sigma^2}\right)\right\}\\ z^{t+1}&=U^{\rm T}\left(y-\mu-p^{t+1}\right).\end{aligned} \tag{15}$$

After obtaining the optimal solution $\hat{z}$ and $\hat{p}$, the distance between the tracked target and the linear representative PCA subspace can be calculated as

$$d\left(y;z,p\right)=\frac{1}{2}\left\|y-U\hat{z}-\mu-\hat{p}\right\|_2^2+\lambda\|\hat{p}\|_1. \tag{16}$$

For every observed target candidate, we calculate their distance according to (18). The observation likelihood can be represented by

$$p\left(y|x\right)=\exp\left(-\gamma d\right) \tag{17}$$

where γ is a constant. Finally, we choose one observed target object with the maximal observation likelihood.

According to the properties of the half-quadratic algorithm, for a fixed z^t, $J\left(z^t,p^{t+1}\right)\leq J\left(z^t,p^t\right)$. And for a fixed p^{t+1}, $J\left(z^{t+1},p^{t+1}\right)\leq J\left(z^t,p^{t+1}\right)$. Thus, $J\left(z^{t+1},p^{t+1}\right)\leq J\left(z^t,p^{t+1}\right)\leq J\left(z^t,p^t\right)$. The cost function is non-incre-

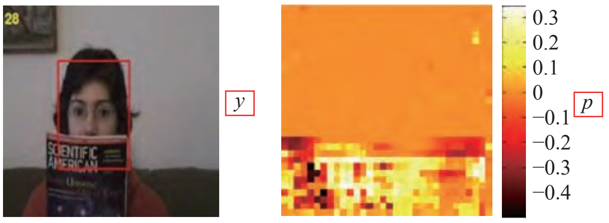
Table 1 Loss functions and their minimizer functions

Functions $\phi\left(x\right)$	$1-\exp\left(-\frac{x^2}{\sigma^2}\right)$	$\begin{cases}x^2/2, & \text{if } \lvert x\rvert\leq\lambda\\ \lambda\lvert x\rvert-\frac{\lambda^2}{2}, & \text{if } \lvert x\rvert>\lambda\end{cases}$	$\log\left(\cosh\left(\alpha x\right)\right)$
Minimizer functions $\delta\left(x\right)$	$x-x\exp\left(-\frac{x^2}{\sigma^2}\right)$	$\begin{cases}0, & \text{if } \lvert x\rvert\leq\lambda\\ x-\lambda\text{sgn}\left(x\right), & \text{if } \lvert x\rvert>\lambda\end{cases}$	$x-\alpha\tanh\left(\alpha x\right)$

asing at each step. Since the objective function is bounded, it should be decreased until converges.

4.3 Model update

Updating the observation model for handling the appearance change of a target object is essential for object tracking. If some imprecise samples are used, the tracking model degrades and thereby causes tracking drift. Correntropy-based loss functions can efficiently predict the outliers and occlusions, therefore ensuring the template is clearer and cleaner. After obtaining the best candidate state of each frame, the observation vector is extracted as $y = [y_1, y_2, \cdots, y_d]$ and the corresponding auxiliary variable is represented as $p = [p_1, p_2, \cdots, p_d]$. Based on half-quadratic analysis, p can be seen as an error correction term. Each auxiliary variable p corresponds to an image. An element of this image indicates that pixel is oc-



(a) Occluded image (b) Corresponding auxiliary variable p

Fig. 3 An illustrative example of auxiliary variable p

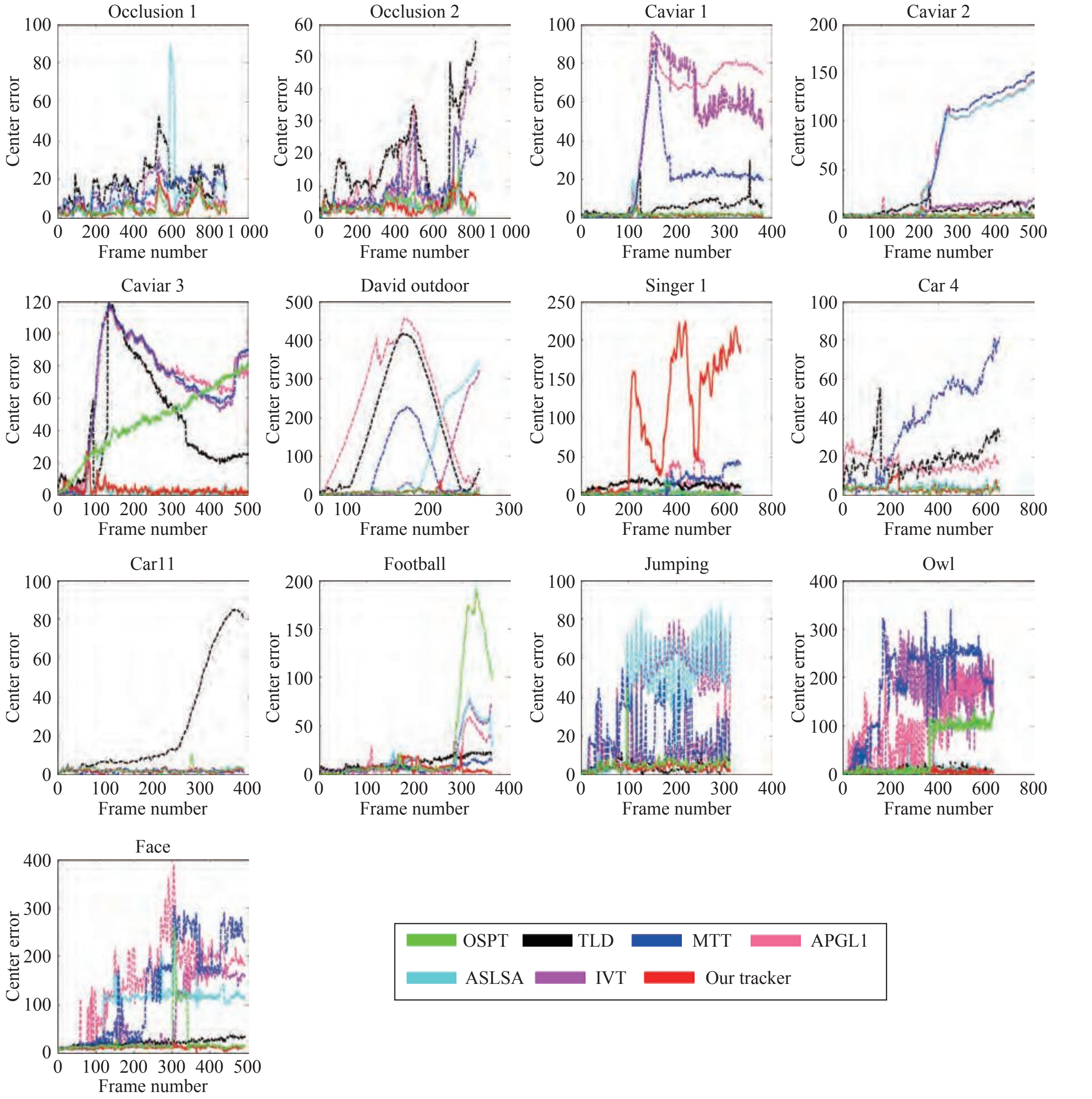


Fig. 4 Center location errors of different methods on thirteen video sequences. The smaller location error is, the better a tracker is.

cluded or not. Fig. 3 shows an example. As shown in Fig. 3, the occluded region can be accurately estimated by the methods based on correntropy. Compared with other loss functions, those of correntropy are smoother, especially in non-occluded regions.

Considering the above properties of correntropy, we use the following template update strategy:

$$y_{rec} = \begin{cases} y_i, & \text{if } p_i \le \text{mean}(p) \\ u_i, & \text{if } p_i > \text{mean}(p) \end{cases} \tag{18}$$

where y_{rec} represents the reconstructed template, and $\mu = [\mu_1, \mu_2, \cdots, \mu_d]$ is the mean template. We use the mean value of p as the threshold to determine the pixel in the template is occluded or not. If $p_i > \text{mean}(p)$, it indicates the pixel is not occluded. Then the pixel is used as part of the reconstructed template. Otherwise, we replace the occluded pixels by the corresponding average observation u. Wang et al.[38] simply used hand-tuned values to determine the template is occluded or not. There are mainly two benefits for our template updating strategy. On the one side, we have a theoretical foundation for updating the template based on the half-quadratic analysis. The error correction term p can be used as an evaluation of the occlusion. On the other side, a fixed threshold is not flexible enough for handing the complex video sequences. Compared with [38], the proposed template update scheme takes the mean error term into account, thereby making it more robust to various noise.

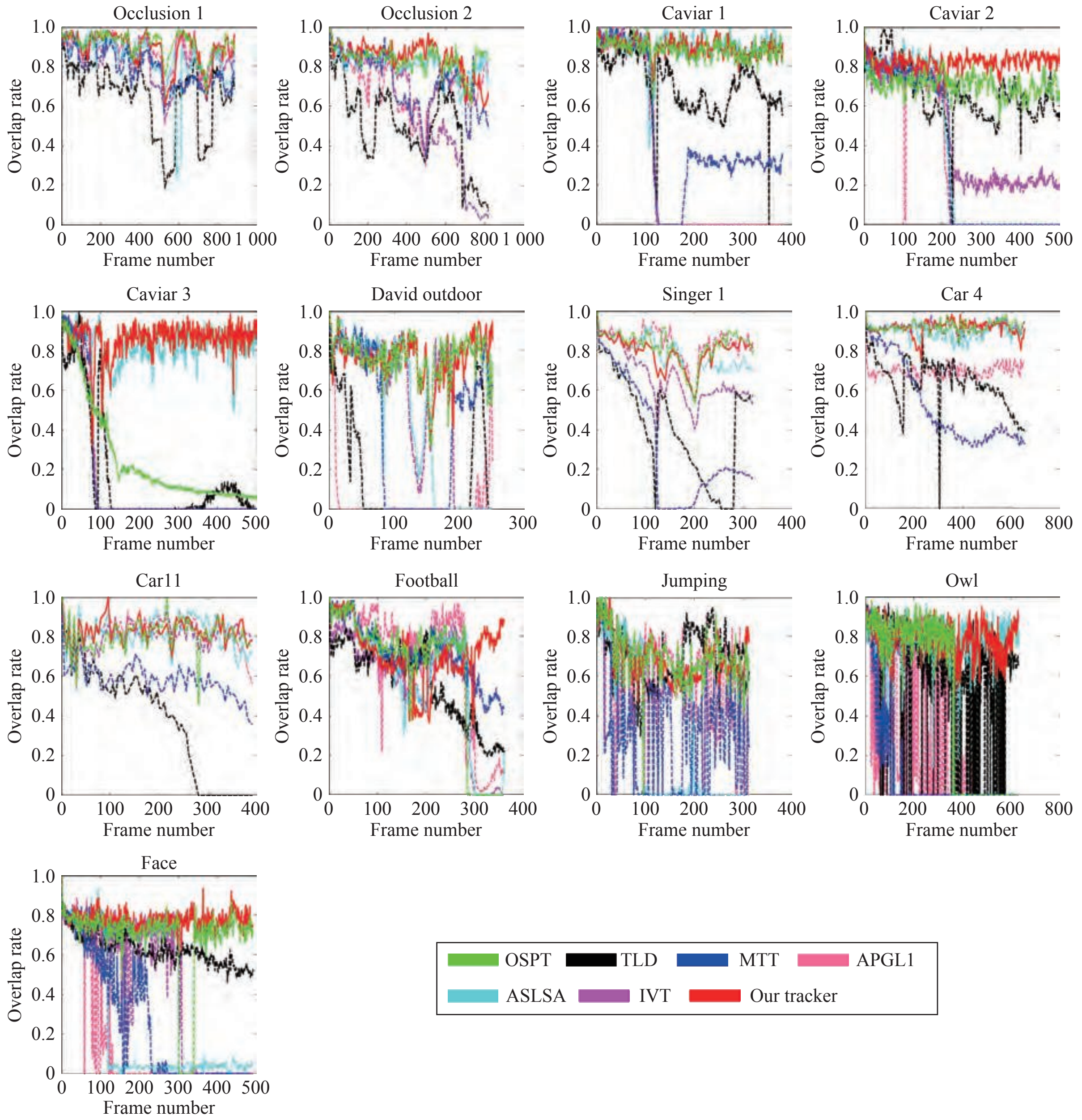


Fig. 5 Overlap rates of different methods on thirteen video clips. The higher an overlap rate is, the better a tracker is.

5 Experiments

In order to evaluate our method and other methods thoroughly, we have selected 13 representative publicly available video sequences with different challenging properties from the benchmark dataset[34] and the Caviar dataset. These datasets are captured in different scenarios and contain challenging appearance variations due to various noises. These challenging video sequences suffer from partial occlusions, illumination variations, pose variations, background clutters and motion blurs. In this section, we have also compared our method with different methods on OTB-13 datasets[34], which contains 50 fully-annotated sequences. Note that some of the videos in OTB-13 are the same as those in the Caviar dataset.

5.1 Experimental settings

For the affine parameters in the particle filter, we set them according to previous tracking papers instead of performing an exhaustive grid search. This can verify the generalization performance of our method. The location of the tracked target in the first frame is labeled for all video sequences. Several state-of-the-art methods are compared, including the frag-track (FragT)[50], incremental subspace learning (IVT)[27], multiple instance learning (MIL)[30], visual tracking decomposition (VTD)[51], local sparse appearance tracking (LSAT)[52], tracking-learning-detection (TLD)[53], accelerated proximal Gradient L_1 (APGL1)[29], multi-task tracking (MTT)[14], sparsity-based collaborative model (SCM)[46], adaptive structure local sparse appearance (ASLSA)[54] and object tracking with sparse prototypes (OSPT)[38].

The kernel size σ in (15) is important, which controls the robust properties of correntropy. Outliers and noise can be effectively eliminated by the kernel size. It is set as the mean reconstruction error. The values of auxiliary variables are also determined by the Gaussian kernel function. The regularization parameter σ is set to 0.3. The target image observation is resized to a resolution of

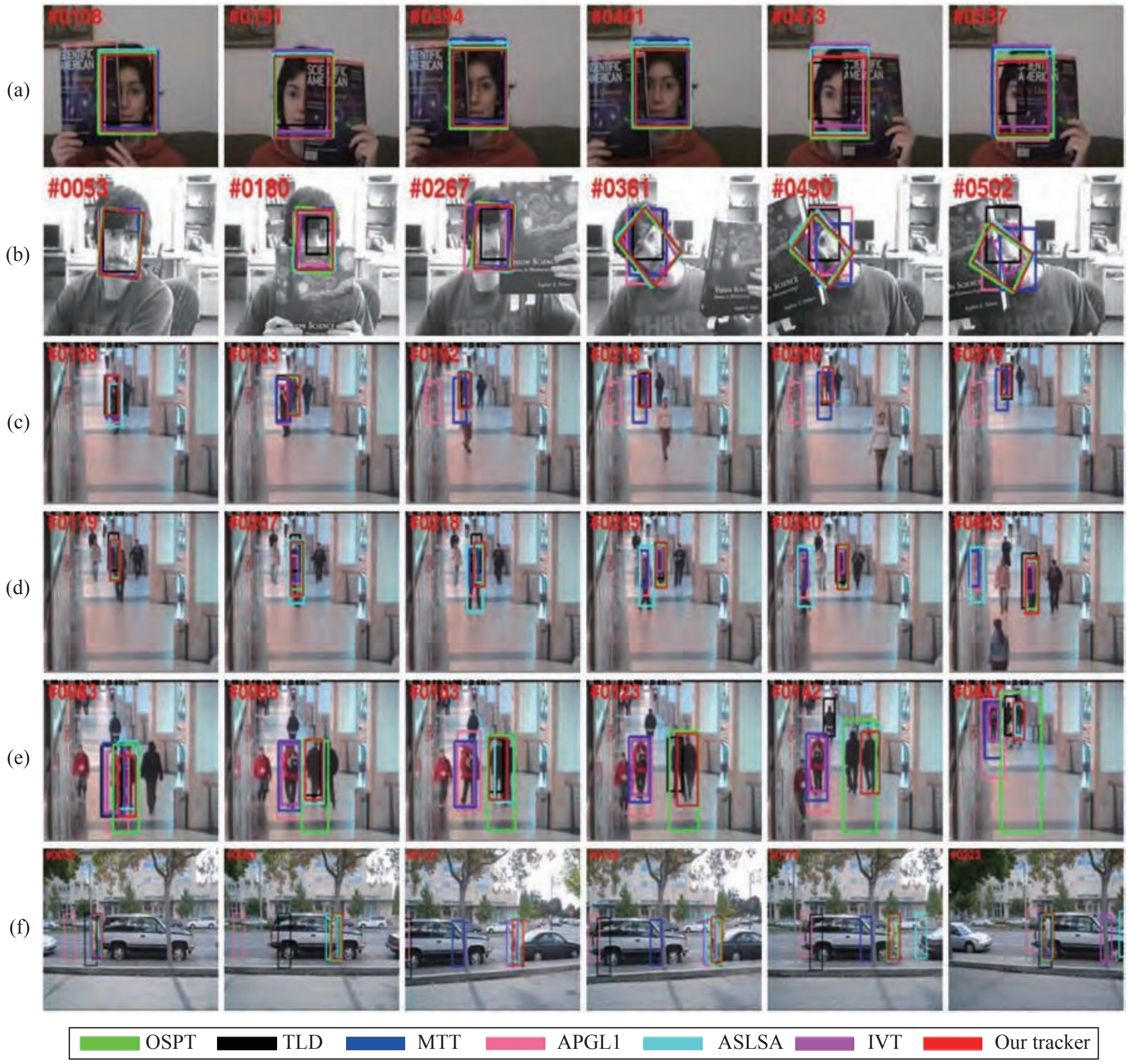


Fig. 6 Qualitative results on some typical frames with partial occlusions

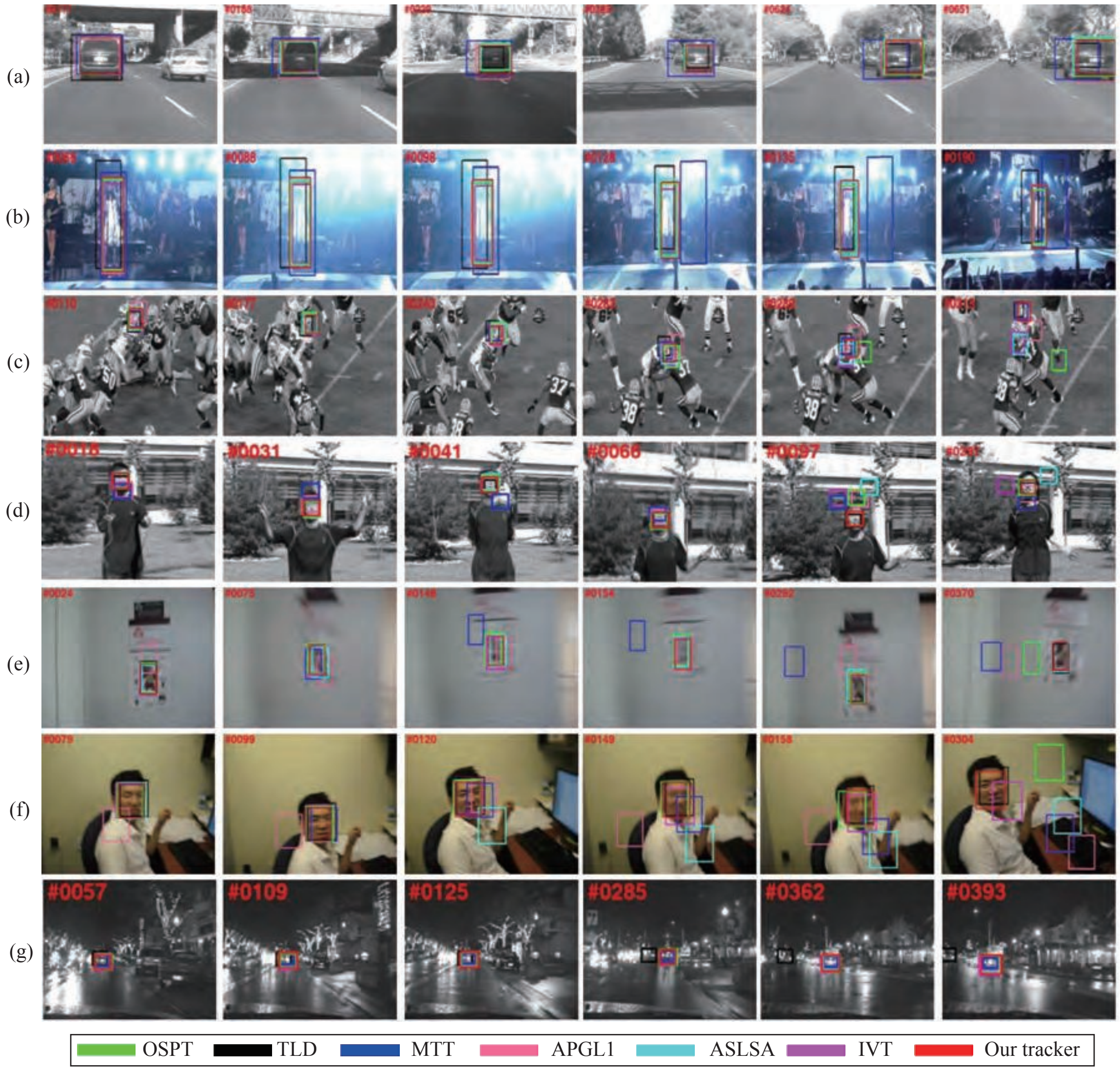


Fig. 7 Qualitative results on some typical frames with illumination variations, background clutters and abrupt motions

32 × 32. Sixteen eigenvectors are used for PCA representation for the template update, and are incrementally updated every 5 frames. For the particle filter, 600 particles are adopted.

5.2 Quantitative evaluation via standard criteria

Three widely used criteria are employed to evaluate the performance of different trackers: the center location error, the overlap rate and the successful tracking rate. The center location error denotes the relative average errors between the predicted and the ground truth center locations. A smaller center location error indicates a more accurate result. The overlap rate is defined as

$$score = \frac{area\left(BB_T \cap BB_G\right)}{area\left(BB_T \cup BB_G\right)} \tag{19}$$

where BB_T is the bounding box of each frame predicted by the trackers, BB_G is the ground truth bounding box. An object is regarded as being successfully tracked when this score is bigger than 0.5. A larger overlap rate indicates a more accurate result. The successful tracking rate is defined as

$$sr = \frac{N_s}{N_t} \times 100\% \tag{20}$$

where N_s and N_t denote the number of successfully tracked frames and the total number of tracked frames.

Table 2 tabulates the average center location errors of different tracking methods. Table 3 summarizes the average overlap rates of different tracking methods. Table 4 further presents the successful tracking rate of different tracking methods. As shown in Tables 2–4, our method outperforms other popular tracking methods in terms of the average center location error, the average overlap rate and the average successful tracking rate. Figs. 4 and 5 further plot the center location error plots and the over-

Table 2 Average center location errors (number of pixes) of different methods. We highlight the best three results using red, blue, and green colors respectively.

Sequence	IVT	FragT	TLD	MIL	VTD	APGL1	MTT	LSAT	SCM	ASLSA	OSPT	Ours
Caviar1	45.2	5.7	5.6	48.5	3.9	50.1	20.9	1.8	0.9	1.4	1.7	1.4
Caviar2	8.6	5.6	8.5	70.3	4.7	63.1	65.4	45.6	2.5	62.3	2.2	2.3
Caviar3	66.0	116.1	44.4	100.2	58.2	68.6	67.5	55.3	2.2	2.2	45.7	3.3
Occlusion1	9.2	5.6	17.6	32.3	11.1	6.8	14.1	5.3	3.2	10.8	4.7	5.2
Occlusion2	10.2	15.5	18.6	14.1	10.4	6.3	9.2	58.6	4.8	3.7	4.0	3.7
DavidOutdoor	53.0	90.5	173.0	38.4	61.9	233.4	65.5	101.7	64.1	87.5	5.8	6.5
Singer1	8.5	22.0	32.7	15.2	4.1	3.1	41.2	14.5	3.7	5.3	4.7	4.6
Car4	2.9	179.8	18.8	60.1	12.3	16.4	37.2	3.3	3.5	4.3	3.0	3.3
Car11	2.1	63.9	25.1	43.5	27.1	1.7	1.8	4.1	1.8	2.0	2.2	1.8
Football	18.2	16.7	11.8	16.0	4.1	12.4	6.5	14.1	10.4	18.0	33.7	6.4
Jumping	36.8	58.4	3.6	9.9	63.0	8.8	19.2	55.2	3.9	39.1	5.0	3.8
Owl	141.4	148.0	8.2	148.9	86.8	104.2	184.3	110.7	7.3	7.6	47.4	6.0
Face	69.7	48.8	22.3	134.7	141.4	148.9	127.2	16.5	125.1	95.1	24.1	11.7
Average	38.8	59.7	30.0	56.3	37.6	55.7	50.8	37.4	18. 0	26.1	14.2	4.6

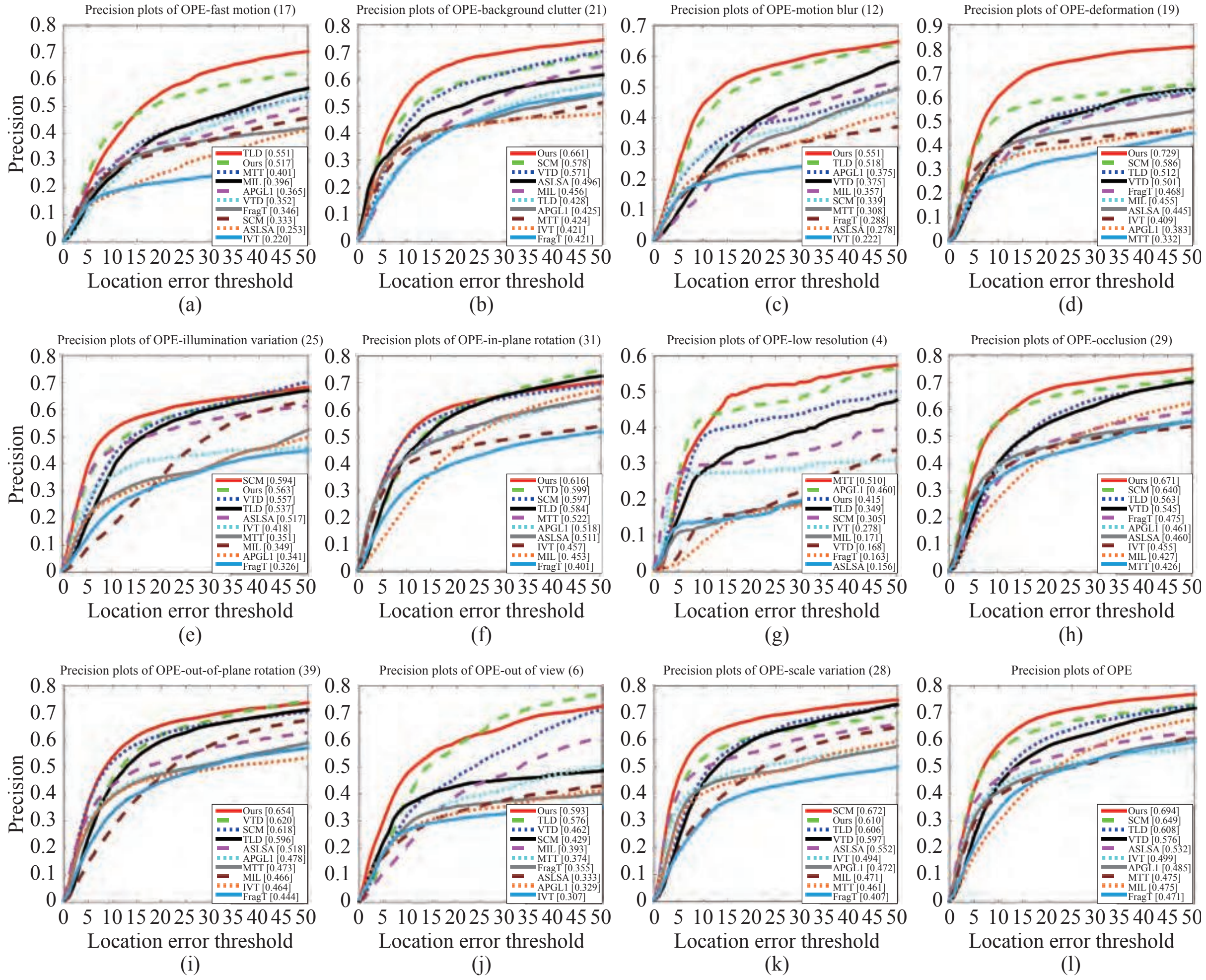


Fig. 8 Precision plots of different methods based on attributes of image sequences on the OTB-13 dataset. (a)–(k): precision plots on 11 tracking challenges of fast motion, background clutter, motion blur, deformation, illumination variation, in-plane rotation, low resolution, occlusion, out-of-plane rotation, out of view and scale variation. (l): overall precision plots of OPE. The legend contains the precision score of each method.

Table 3 Average overlap rates of different methods. The best three results are shown in red, blue, and green colors respectively.

Sequence	IVT	FragT	TLD	MIL	VTD	APGL1	MTT	LSAT	SCM	ASLSA	OSPT	Ours
Caviar1	0.28	0.68	0.70	0.25	0.83	0.28	0.45	0.85	0.91	0.90	0.89	0.90
Caviar2	0.45	0.56	0.66	0.26	0.67	0.32	0.33	0.28	0.81	0.35	0.71	0.82
Caviar3	0.14	0.13	0.16	0.13	0.15	0.13	0.14	0.58	0.87	0.82	0.25	0.86
DavidOutdoor	0.52	0.39	0.16	0.41	0.42	0.05	0.42	0.36	0.46	0.45	0.77	0.76
Occlusion1	0.85	0.90	0.65	0.59	0.77	0.87	0.79	0.90	0.93	0.83	0.91	0.89
Occlusion2	0.59	0.60	0.49	0.61	0.59	0.70	0.72	0.33	0.82	0.81	0.84	0.84
Singer1	0.66	0.34	0.41	0.34	0.79	0.83	0.32	0.52	0.85	0.78	0.82	0.79
Car4	0.92	0.22	0.64	0.34	0.73	0.70	0.53	0.91	0.89	0.89	0.92	0.91
Car11	0.81	0.09	0.38	0.17	0.43	0.83	0.58	0.49	0.79	0.81	0.81	0.83
Football	0.55	0.57	0.56	0.55	0.81	0.68	0.71	0.63	0.69	0.57	0.62	0.72
Jumping	0.28	0.14	0.69	0.53	0.08	0.59	0.30	0.09	0.73	0.24	0.69	0.69
Owl	0.22	0.09	0.60	0.09	0.12	0.17	0.09	0.13	0.79	0.78	0.48	0.79
Face	0.44	0.39	0.62	0.15	0.24	0.14	0.26	0.69	0.36	0.21	0.68	0.77
Average	0.52	0.39	0.52	0.34	0.51	0.48	0.43	0.52	0.76	0.65	0.72	0.81

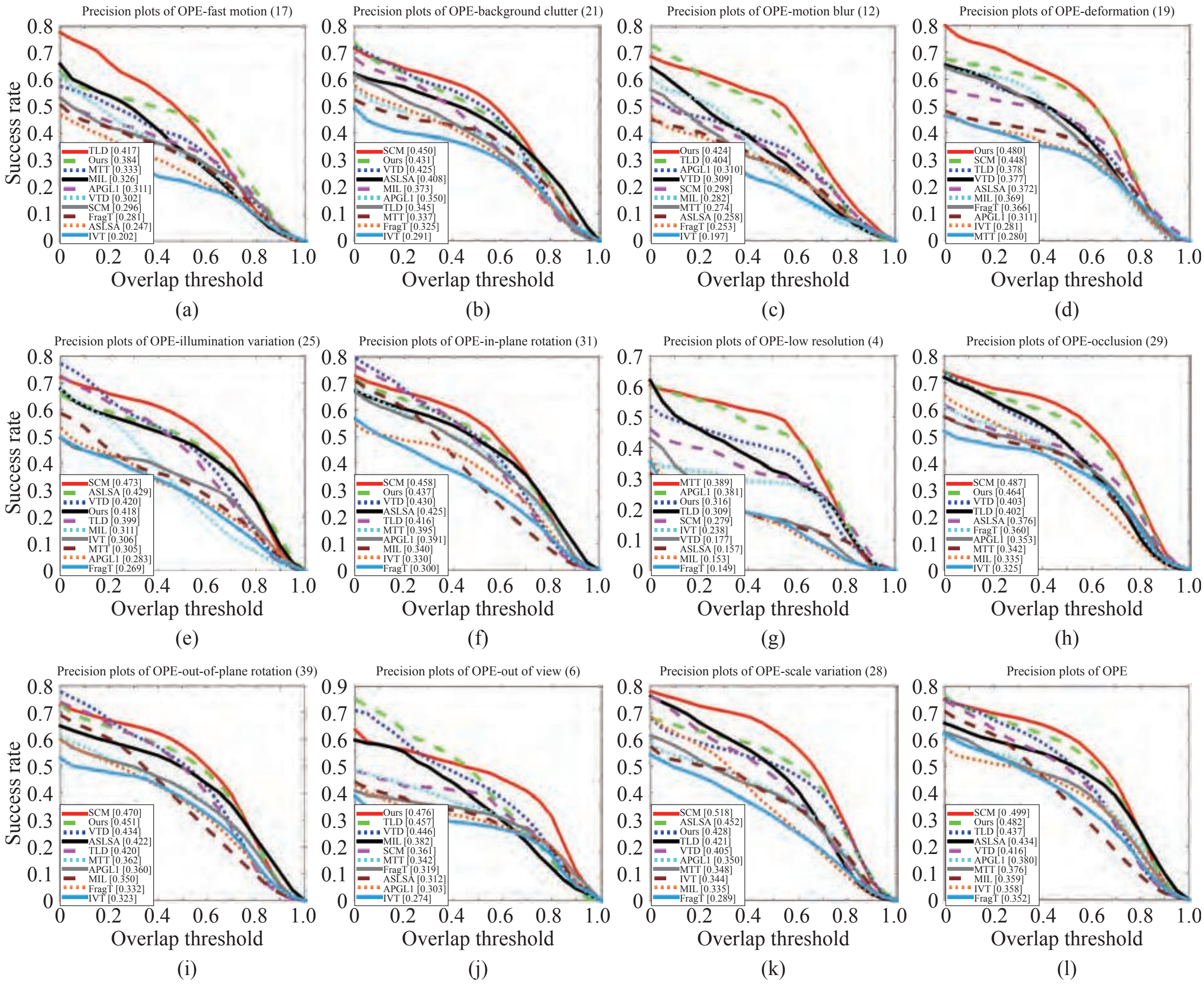


Fig. 9 Success plots of different methods based on attributes of image sequences on the OTB-13 dataset. (a)–(k): success plots on 11 tracking challenges of fast motion, background clutter, motion blur, deformation, illumination variation, in-plane rotation, low resolution, occlusion, out-of-plane rotation, out of view and scale variation. (l): overall success plots of OPE. The legend contains the AUC score of each method.

Table 4 Successful tracking rates of different methods. The best three results are shown in red, blue, and green colors respectively.

Sequence	IVT	FragT	TLD	MIL	VTD	APGL1	MTT	LSAT	SCM	ASLSA	OSPT	Ours
Caviar1	0.30	0.96	0.96	0.28	0.97	0.30	0.30	0.99	1	0.99	1	1
Caviar2	0.43	0.58	0.94	0.34	0.76	0.40	0.43	0.37	1	0.41	1	1
Caviar3	0.16	0.16	0.17	0.16	0.14	0.15	0.16	0.69	0.99	0.99	0.17	0.99
Occlusion1	1	1	0.78	0.76	0.97	1	1	1	1	0.97	1	1
Occlusion2	0.56	0.65	0.52	0.72	0.68	0.93	0.93	0.40	1	1	1	1
DavidOutdoor	0.62	0.46	0.18	0.35	0.53	0.05	0.56	0.48	0.58	0.51	0.97	0.97
Singer1	0.94	0.25	0.46	0.25	0.95	1	0.35	0.51	1	1	1	1
Car4	1	0.27	0.86	0.27	1	1	0.38	0.99	1	1	1	1
Car11	1	0.09	0.47	0.08	0.53	1	0.87	0.42	0.99	1	0.99	1
Football	0.72	0.75	0.61	0.70	0.99	0.77	0.88	0.77	0.83	0.69	0.77	0.89
Jumping	0.37	0.14	0.92	0.46	0.10	0.72	0.22	0.10	0.98	0.30	0.96	0.99
Owl	0.28	0.06	0.82	0.08	0.07	0.16	0.07	0.12	0.96	0.98	0.57	1
Face	0.58	0.33	0.97	0.15	0.23	0.17	0.26	1	0.46	0.24	0.92	1
Average	0.61	0.44	0.67	0.35	0.61	0.59	0.49	0.60	0.91	0.78	0.87	0.99

lap rate plots of different video sequences. These results further verify that our method achieves better tracking results than its competitors.

From Tables 2, 3 and 4, we can also see that most of the tracking methods can handle the partial occlusion problems (e.g., Occlusion1 and Occlusion2). Furthermore, SCM, OSPT and our algorithm can handle the severe occlusion problems (e.g., Caviar2, Caviar3 and DavidOutdoor). Our method can also deal with the illumination variations problems (e.g., Singer1, Car4 and Car 11) due to the usage of robust error functions and the novel template updating strategy. While for the abrupt motion (e.g., Football, Jumping, Owl and Face), our method seems to be better than other tracking methods.

5.3 Qualitative evaluation for challenging factors

Figs. 6 and 7 show some tracking results of different tracking methods on 14 challenging video sequences. The challenging factors of these sequences include partial occlusions, scale changes, in-plane rotations, pose variations, illumination variations, abrupt motions and background clutters. As shown in Fig. 6, video sequences Occlusion1, Occlusion2, Caviar1, Caviar2, Caviar3, DavidOutdoor pose long-time partial occlusions, and scale changes. From Fig. 6, we can see that our method performs well even when the target undergoes severe partial occlusions. In contrast, most of the other tracking methods only work well on three or four sequences. When the tracked object presents large scale changes and partial occlusions (e.g., Occlusion2 #0431), APGL1, TLD, MTT and IVT fail to track the target. OSPT and ASLSA have small drift from the tracked object in Caviar2 #0088 and DavidOutdoor #0083. When our tracker deviates from the target in Caviar3 #0083, we still track the object successfully in the following sequences (e.g., #0087). Compared with IVT, our method is more robust for various outliers. This is partly because our method adopts the correntropy-based holistic appearance model to handle the complex noises. What is more, our method only has one parameter to determine: the kernel size of correntropy, which is different from its competitors.

Fig. 7 further shows the tracking results on the sequences (Car4, Singer1, Football) with significant illumination variations and background clutters. Even with severe occlusions and illumination variations, our method can handle the situation in Singer1 #0098 and Singer1 #0135. Although OSPT and IVT can track the target, their output positions are not very accurate. Other tracking methods cannot track the target successfully. MTT and TLD are less effective in these cases (e.g., Car4 #0651 and Singer1 #0088). IVT, OSPT and our method achieve good performance even with the appearance variations caused by light changes due to the usage of incremental PCA algorithm. Our tracker performs better than its competitors in these videos when there are drastic illumination variations and background clutters (e.g., Singer1 #0128 and Football #0177). Fig. 7 also shows the tracking results on the Jumping, Owl and Face sequences with abrupt motion. MTT, ASLSA and IVT fail to predict the true locations of the target objects when they undergo abrupt motions (e.g., Jumping #0097 and Owl #0024). The appearance changes caused by motion blur pose great challenges for accurately capturing the tracked targets. Experimental results on Fig. 7 further demonstrate that our method performs better than its competitors.

5.4 Experimental results on the OTB-13 dataset

In this subsection, we evaluate different algorithms on the OTB-13 dataset, which contains much more challenging factors than previous datasets. The proposed tracker is evaluated and compared with different trackers. The one-pass evaluation (OPE) protocol with precision and success plots is used in this dataset to evaluate different trackers. The precision metric computes the percentage of frames whose estimated center location is within the given threshold distance with the ground truth location. The success metric calculates the number of successful frames whose overlap ratio between the tracked and ground truth bounding boxes is larger than a given threshold. Similar with [34], for precision plots we use the results at error threshold of 20 pixels for ranking, while we use area under curve (AUC) scores to summarize the trackers for success plots.

Fig. 8 illustrates the precision plots of different methods on the OTB-13 dataset based on center location error. The sequence attributes of this dataset include 11 challenging factors in the tracking problem, e.g., fast motion, background clutter, motion blur, deformation, etc. We can analyze the performance of different trackers in different aspects with these attributes. As shown in Fig. 8, SCM performs better than other trackers. The overall success rate of SCM is 0.649, which performs better than TLD and VTD by 4.1% and 7.3% in terms of success rate. The overall success rate of our method is 0.694, which further beats SCM by 4.5%. Note that SCM performs best in dealing with challenging factors including illumination variation and scale variation. Our method performs best in dealing with challenging factors including background clutter, motion blur, deformation, in-plane rotation, occlusion, out-of-plane rotation and out of view. For the sequences with all of the attributes, our method performs best among all the other trackers.

Fig. 9 further shows the OPE success plots of different tracking methods. Overall, the proposed algorithm performs well against other methods. For example, the AUC score of TLD, ASLSA and VTD is 0.437, 0.434 and 0.416. The AUC score of our method is 0.482, which outperforms TLD, ASLSA, VTD by 4.5%, 4.8% and 6.6%, respectively. It is worth noticing that SCM performs slightly better than our algorithm in terms of the success plots. The AUC score of SCM is 0.499, which beats our method by about 1.7%. The possible reason is that under some circumstances, our method fails to follow the targets, while SCM can still track the objects with a small overlap ratio due to the combination of generative and discriminative modules.

6 Conclusions

In this paper, we have proposed a robust incremental object tracking algorithm. Correntropy has been introduced to deal with various non-Gaussian noises in video sequences. Furthermore, a novel template update scheme has been proposed to deal with the imprecise samples during the tracking process. The effectiveness of our algorithm is demonstrated by various video sequences with complex noises. Compared with the current incremental object tracking algorithms that assume a certain noise distribution, our tracker can perform better under much more complex noises in various tracking tasks.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Nos. 61702513, 61525306, 61633021), National Key Research and Development Program of China (No. 2016YFB1001000), Capital Science and Technology Leading Talent Training Project (No. Z181100006318030), CAS-AIR and Shandong Provincial Key Research and Development Program (Major Scientific and Technological Innovation Project) (No. 2019JZZY010119)

References

[1] S. Sun, N. Akhtar, H. S. Song, A. S. Mian, M. Shah. Deep affinity network for multiple object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to be published. DOI: 10.1109/TPAMI.2019.2929520.

[2] X. Y. Lan, M. Ye, S. P. Zhang, H. Y. Zhou, P. C. Yuen. Modality-correlation-aware sparse representation for RGB-infrared object tracking. *Pattern Recognition Letters*, vol. 130, pp. 12–20, 2020. DOI: 10.1016/j.patrec.2018.10.002.

[3] C. Ma, J. B. Huang, X. K. Yang, M. H. Yang. Adaptive correlation filters with long-term and short-term memory for object tracking. *International Journal of Computer Vision*, vol. 126, no. 8, pp. 771–796, 2018. DOI: 10.1007/s11263-018-1076-4.

[4] T. Z. Zhang, S. Liu, N. Ahuja, M. H. Yang, B. Ghanem. Robust visual tracking via consistent low-rank sparse learning. *International Journal of Computer Vision*, vol. 111, no. 2, pp. 171–190, 2014. DOI: 10.1007/s11263-014-0738-0.

[5] S. Hare, A. Saffari, P. H. S. Torr. Struck: Structured output tracking with kernels. In *Proceedings of IEEE International Conference on Computer Vision*, IEEE, Barcelona, Spain,pp. 263–270,2011.DOI:10.1109/ICCV.2011.6126251.

[6] X. Mei, H. B. Ling, Y. Wu, E. Blasch, L. Bai. Minimum error bounded efficient ℓ_1 tracker with occlusion detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, USA, pp. 1257–1264, 2011. DOI: 10.1109/CVPR.2011.5995421.

[7] T. Z. Zhang, S. Liu, C. S. Xu, S. C. Yan, B. Ghanem, N. Ahuja, M. H. Yang. Structural sparse tracking. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Boston, USA, pp. 150–158, 2015. DOI: 10.1109/CVPR.2015.7298610.

[8] Z. B. Kang, W. Zou, Z. Zhu, H. X. Ma. Smooth-optimal adaptive trajectory tracking using an uncalibrated fish-eye camera. *International Journal of Automation and Computing*, vol. 17, no. 2, pp. 267–278, 2020. DOI: 10.1007/s11633-019-1209-4.

[9] Q. Fu, X. Y. Chen, W. He. A survey on 3D visual tracking of multicopters. *International Journal of Automation and Computing*, vol. 16, no. 6, pp. 707–719, 2019. DOI: 10.1007/s11633-019-1199-2.

[10] S. Liu, G. C. Liu, H. Y. Zhou. A robust parallel object tracking method for illumination variations. *Mobile Networks and Applications*, vol. 24, no. 1, pp. 5–17, 2019. DOI: 10.1007/s11036-018-1134-8.

[11] Y. K. Qi, L. Qin, S. P. Zhang, Q. M. Huang, H. X. Yao. Robust visual tracking via scale-and-state-awareness. *Neurocomputing*, vol. 329, pp. 75–85, 2019. DOI: 10.1016/j.neucom.2018.10.035.

[12] D. Wang, H. C. Lu. Visual tracking via probability continuous outlier model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, USA, pp. 3478–3485, 2014. DOI: 10.1109/CVPR.2014.445.

[13] F. Yang, H. C. Lu, M. H. Yang. Robust superpixel tracking. *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1639–1651, 2014. DOI: 10.1109/TIP.2014.2300823.

[14] T. Z. Zhang, B. Ghanem, S. Liu, N. Ahuja. Robust visual tracking via multi-task sparse learning. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, USA, pp. 2042–2049, 2012. DOI: 10.1109/CVPR.2012.6247908.

[15] H. G. Ren, W. M. Liu, T. Shi, F. J. Li. Compressive tracking based on online Hough forest. *International Journal of Automation and Computing*, vol. 14, no. 4, pp. 396–406, 2017. DOI: 10.1007/s11633-017-1083-x.

[16] Z. Q. Zhao, P. Zheng, S. T. Xu, X. D. Wu. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019. DOI: 10.1109/TNNLS.2018.2876865.

[17] Q. Wang, L. Zhang, L. Bertinetto, W. M. Hu, P. H. S. Torr. Fast online object tracking and segmentation: A unifying approach. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Long Beach, USA, pp. 1328–1338, 2019. DOI: 10.1109/CVPR.2019.00142.

[18] J. R. Xue, J. W. Fang, P. Zhang. A survey of scene understanding by event reasoning in autonomous driving. *International Journal of Automation and Computing*, vol. 15, no. 3, pp. 249–266, 2018. DOI: 10.1007/s11633-018-1126-y.

[19] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, ACM, Lake Tahoe, USA, pp. 1097–1105, 2012.

[20] J. Long, E. Shelhamer, T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Boston, USA, pp. 3431–3440, 2015. DOI: 10.1109/CVPR.2015.7298965.

[21] H. Fan, L. T. Lin, F. Yang, P. Chu, G. Deng, S. J. Yu, H. X. Bai, Y. Xu, C. Y. Liao, H. B. Ling. LaSOT: A high-quality benchmark for large-scale single object tracking. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Beach, USA, pp. 5374–5383, 2019. DOI: 10.1109/CVPR.2019.00552.

[22] K. H. Zhang, Q. S. Liu, Y. Wu, M. H. Yang. Robust visual tracking via convolutional networks without training. *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1779–1792, 2016. DOI: 10.1109/TIP.2016.2531283.

[23] T. Z. Zhang, C. S. Xu, M. H. Yang. Learning multi-task correlation particle filters for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 365–378, 2018. DOI: 10.1109/TPAMI.2018.2797062.

[24] X. Mei, H. B. Ling. Robust visual tracking using ℓ_1 minimization. In *Proceedings of the 12th IEEE International Conference on Computer Vision*, IEEE, Kyoto, Japan, pp. 1436–1443, 2009. DOI: 10.1109/ICCV.2009.5459292.

[25] T. Z. Zhang, B. Ghanem, S. Liu, N. Ahuja. Low-rank sparse learning for robust visual tracking. *In Proceedings of the 12th European Conference on Computer Vision*, *Springer*, *Florence*, *Italy*, pp. 470–484, 2012. DOI: 10.1007/978-3-642-33783-3_34.

[26] T. Z. Zhang, K. Jia, C. S. Xu, Y. Ma, N. Ahuja. Partial occlusion handling for visual tracking via robust part matching. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, USA, pp. 1258–1265, 2014. DOI: 10.1109/CVPR.2014.164.

[27] D. A. Ross, J. Lim, R. S. Lin, M. H. Yang. Incremental learning for robust visual tracking. *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125–141, 2008. DOI: 10.1007/s11263-007-0075-7.

[28] Y. Wu, H. B. Ling, J. Y. Yu, F. Li, X. Mei, E. K. Cheng. Blurred target tracking by blur-driven tracker. In *Proceedings of IEEE International Conference on Computer Vision*, IEEE, Barcelona, Spain, pp. 1100–1107, 2011. DOI: 10.1109/ICCV.2011.6126357.

[29] C. L. Bao, Y. Wu, H. B. Ling, H. Ji. Real time robust L1 tracker using accelerated proximal gradient approach. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, USA, pp. 1830–1837, 2012. DOI: 10.1109/CVPR.2012.6247881.

[30] B. Babenko, M. H. Yang, S. Belongie. Visual tracking with online multiple instance learning. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Miami, USA, pp. 983–990, 2009. DOI: 10.1109/CVPR.2009.5206737.

[31] J. Gall, A. Yao, N. Razavi, L. Van Gool, V. Lempitsky. Hough forests for object detection, tracking, and action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2188–2202, 2011. DOI: 10.1109/TPAMI.2011.70.

[32] S. Liu, T. Z. Zhang, X. C. Cao, C. S. Xu. Structural correlation filter for robust visual tracking. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, *IEEE*, *Las Vegas*, *USA*, pp. 4312–4320, 2016. DOI: 10.1109/CVPR.2016.467.

[33] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, M. Shah. Visual tracking: An experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442–1468, 2014. DOI: 10.1109/TPAMI.2013.230.

[34] Y. Wu, J. Lim, M. H. Yang. Online object tracking: A benchmark. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Portland, USA, pp. 2411–2418, 2013. DOI: 10.1109/CVPR.2013.312.

[35] Z. T. Li, W. Wei, T. Z. Zhang, M. Wang, S. J. Hou, X. Peng. Online multi-expert learning for visual tracking. *IEEE Transactions on Image Processing*, vol. 29, pp. 934–946, 2019. DOI: 10.1109/TIP.2019.2931082.

[36] T. Z. Zhang, S. Liu, C. S. Xu, B. Liu, M. H. Yang. Correlation particle filter for visual tracking. *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2676–2687, 2018. DOI: 10.1109/TIP.2017.2781304.

[37] M. J. Black, A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*, vol. 26, no. 1, pp. 63–84, 1998. DOI: 10.1023/A:1007939232436.

[38] D. Wang, H. C. Lu, M. H. Yang. Online object tracking

with sparse prototypes. *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 314–325, 2013. DOI: 10.1109/TIP.2012.2202677.

[39] D. Wang, H. C. Lu, M. H. Yang. Least soft-threshold squares tracking. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Portland, OR, USA, pp. 2371–2378, 2013. DOI: 10.1109/CVPR.2013.307.

[40] N. Y. Wang, J. D. Wang, D. Y. Yeung. Online robust non-negative dictionary learning for visual tracking. In *Proceedings of IEEE International Conference on Computer Vision*, IEEE, Sydney, NSW, Australia, pp. 657–664, 2013. DOI: 10.1109/ICCV.2013.87.

[41] W. F. Liu, P. P. Pokharel, J. C. Principe. Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007. DOI: 10.1109/TSP.2007.896065.

[42] R. He, W. S. Zheng, B. G. Hu. Maximum correntropy criterion for robust face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1561–1576, 2011. DOI: 10.1109/TPAMI.2010.220.

[43] W. M. Hu, X. Li, X. Q. Zhang, X. C. Shi, S. Maybank, Z. F. Zhang. Incremental tensor subspace learning and its applications to foreground segmentation and tracking. *International Journal of Computer Vision*, vol. 91, no. 3, pp. 303–327, 2011. DOI: 10.1007/s11263-010-0399-6.

[44] T. Wang, I. Y. H. Gu, P. F. Shi. Object tracking using incremental 2D-PCA learning and ml estimation. In *Proceedings of IEEE International Conference on Acoustics*, Speech and Signal Processing, IEEE, Honolulu, HI, USA, pp. I-933–I-936, 2007. DOI: 10.1109/ICASSP.2007.366062.

[45] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,vol. 31,no. 2,pp. 210–227,2009.DOI:10.1109/TPAMI.2008.79.

[46] W. Zhong, H. C. Lu, M. H. Yang. Robust object tracking via sparsity-based collaborative model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, RI, USA, pp. 1838–1845, 2012. DOI: 10.1109/CVPR.2012.6247882.

[47] R. He, W. S. Zheng, T. N. Tan, Z. N. Sun. Half-quadratic-based iterative minimization for robust sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 261–275, 2014. DOI: 10.1109/TPAMI.2013.102.

[48] B. D. Chen, J. C. Principe. Maximum correntropy estimation is a smoothed map estimation. *IEEE Signal Processing Letters*, vol. 19, no. 8, pp. 491–494, 2012. DOI: 10.1109/LSP.2012.2204435.

[49] M. Nikolova, M. K. Ng. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM Journal on Scientific Computing*, vol. 27, no. 3, pp. 937–966, 2005. DOI: 10.1137/030600862.

[50] A. Adam, E. Rivlin, I. Shimshoni. Robust fragments-based tracking using the integral histogram. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, New York, USA, pp. 798–805, 2006. DOI: 10.1109/CVPR.2006.256.

[51] J. Kwon, K. M. Lee. Visual tracking decomposition. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, San Francisco, USA, pp. 1269–1276, 2010. DOI: 10.1109/CVPR.2010.5539821.

[52] B. Y. Liu, J. Z. Huang, L. Yang, C. Kulikowsk. Robust tracking using local sparse appearance model and K-selection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, USA, pp. 1313–1320, 2011. DOI: 10.1109/CVPR.2011.5995730.

[53] Z. Kalal, K. Mikolajczyk, J. Matas. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2012. DOI: 10.1109/TPAMI.2011.239.

[54] X. Jia, H. C. Lu, M. H. Yang. Visual tracking via adaptive structural local sparse appearance model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, RI, USA, pp. 1822–1829, 2012. DOI: 10.1109/CVPR.2012.6247880.

Wei-Ning Wang received the B. Eng. degree in automation from North China Electric Power University, China in 2015. She is currently a Ph. D. degree candidate at National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), China.

Her research interests include computer vision, pattern recognition and video analysis.

E-mail: weining.wang@cripac.ia.ac.cn

ORCID iD: 0000-0001-7299-6431

Qi Li received the B. Eng. degree in automation from the China University of Petroleum, China in 2011 and the Ph. D. degree in pattern recognition and intelligent systems from CASIA, China in 2016. He is currently an associate professor with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China.

His research interests include face recognition, computer vision, and machine learning.

E-mail: qli@nlpr.ia.ac.cn (Corresponding author)

ORCID iD: 0000-0002-7905-2860

Liang Wang received both the B. Eng. and M. Eng. degrees from Anhui University, China in 1997 and 2000, respectively, and the Ph. D. degree from the Institute of Automation, Chinese Academy of Sciences (CASIA), China in 2004. From 2004 to 2010, he was a research assistant at Imperial College London, UK, and Monash University, Australia, a research fellow with the University of Melbourne, Australia, and a lecturer with the University of Bath, UK, respectively. Currently, he is a full professor of the Hundred Talents Program at the National Laboratory of Pattern Recognition, CASIA, China. He is currently an IEEE Fellow and IAPR Fellow.

His research interests include machine learning, pattern recognition, and computer vision.

E-mail: wangliang@nlpr.ia.ac.cn