Infrared target tracking based on proximal robust principal component analysis method

Chao Ma¹,² · Minjie Wan¹,² · Yunkai Xu¹,² · Kan Ren¹,² · Weixian Qian¹,² · Qian Chen¹,² · Guohua Gu¹,²

Accepted: 3 April 2021 / Published online: 22 June 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract
Infrared target tracking plays an important role in both civil and military fields. The main challenges in designing a robust and high-precision tracker for infrared sequences include overlap, occlusion, and appearance change. To this end, this paper proposes an infrared target tracker based on the proximal robust principal component analysis method. Firstly, the observation matrix is decomposed into a sparse occlusion matrix and a low-rank target matrix, and the constraint optimization is carried out with an approaching proximal norm which is better than \( l_1 \)-norm. Then, the Alternating Direction Method of Multipliers (ADMM) is employed to solve this convex optimization problem by estimating the variables alternately. Finally, the framework of the particle filter with model update strategy is exploited to locate the target. Through a series of experiments on real infrared target sequences, the effectiveness and robustness of our algorithm are proved.

Keywords Infrared target tracking · Proximal robust principal component analysis · ADMM · Particle filter framework · Template update

1 Introduction

Target tracking has been a very important part in many areas, such as traffic monitoring, detecting of approaching targets, video surveillance and activity control. Compared to the target tracking using visible camera, which has been deeply investigated in the past several decades, infrared target tracking is a more suitable and effective method to work in the lightless environment. Besides, infrared system is more robust in various environments, whether in dark or light condition. Thus, infrared target tracking has been applied more and more widely in military and civil applications [25]. Accordingly, we hold the opinion that it is necessary to design a high-precision infrared target tracking algorithm in different occasions.

Although the infrared system can work continuously, regardless of day or night, the image information obtained from infrared imaging devices is not as sufficient as that from visible devices. This is due to the imaging principle and technical bottleneck [6]. For example, infrared image usually lack color and texture information. Also there are sensor noises and occlusion clutters in infrared images. Thus, the loss of image details and low signal-to-noise ratio limit the development of infrared target tracking. In addition, partial occlusion and total occlusion are also challenging problems in the process of tracking. In brief, it is vital to establish a precise and credible model for infrared target tracking.

Although the existing target tracking methods have made great progress, the interferences of target overlap, pose change; and environmental occlusion still need to be further studied and overcome [19]. Xun et al. [18] introduced sparse representation into the field of target tracking for the first time, and transformed the tracking into solving an \( l_1 \) minimization problem. This method can be adapted to overcome the interferences such as pose change, severe occlusion and illumination change. Since then, generative models based on low rank and sparse representation have been proposed. Its basic principle is that the target matrix
composed of multiple similar groups of target images possess low-rank property. Besides, the occlusion matrix representing partial occlusion is sparse since its entries only occupy a small number of the image.

To overcome the problems of target overlapping, pose change and occlusion, this paper proposes a tracker based on proximal robust principal component analysis-based tracker. Firstly, the observation matrix is decomposed into a sparse occlusion matrix and a low-rank target matrix, and the constraint optimization is carried out with an approaching proximal norm. In order to solve this convex optimization problem, ADMM [7] algorithm is introduced to solve the variables. Finally, this paper uses the Bayesian state inference and template update strategy under the particle filter framework to achieve long-term tracking. Through a series of experiments on real infrared target sequences, the effectiveness and robustness of our algorithm are proved.

In conclusion, the main contributions of this paper can be summarized as follows:

1. An appearance model based on sparse representation and low-rank is proposed to solve the tracking problem by transforming into a convex optimization problem;
2. Proximal $p$-norm which declines to zero rapidly when the entry of the sparse matrix approaches zero is utilized to constrain occlusion matrix;
3. A dynamic template update strategy is designed so that the appearance change can be well addressed.

The rest of the article is divided into five sections. In Section 2, related works are presented in detail. Section 3, mainly introduces the related work of particle filter. In Section 4, we introduce the establishment of the model in detail. Section 5, demonstrates the reliability of the proposed algorithm through quantitative and qualitative experiments. The last section is a summary of the article.

2 Related work

Based on the review of the existing target tracking algorithms, we divide them into two categories: generative approaches and discriminant approaches. The generative model first extracts the feature of targets and learn the feature model. Through matching the whole image, the target which is the most similar area can be found. In recent years, target detection algorithms based on Bayesian model have been extensively studied and achieved satisfactory performances. The kernel-based tracker is widely investigated, including Mean-shift (MS) tracker [10], which is one of the mainstream target tracking methods. This method uses the steepest descent algorithm to solve iteratively in the direction of gradient descent, the core of which is to iterate over the solution until the optimal value is reached. This process is also to find the candidate which is the most similar to the template. However, MS algorithm cannot track target with changeable size adaptively. To solve this problem, Allen [3] proposed a Cam-shift tracker, which is an improved strategy of MS. It adds scale adaptive mechanism on the original MS to make it adaptively change the size of the tracking box according to the size of the target. Isard et al. [14] firstly investigated a condensation algorithm which applied particle filter framework to target tracking problem for the first time. Nevertheless, due to the lack of observation information at the current moment, the tracking performance turns to be poor in complex scenes. To work out the problems in target tracking, such as poor tracking accuracy, large computation, and poor real-time performance, researchers have proposed many improved particle filter-based tracking algorithms. Guo et al. [12] used the particle filter algorithm to predict the initial target position, in which whether the real-time model of the target is changed or not is decided according to the similarity between the real target model and the predicted target. In order to overcome the difficulty caused by the change of appearance, incremental visual tracking (IVT) [22] were proposed, which uses incremental algorithm based on principal component analysis.

Discriminant approaches consider both the background and the target, and extract the target model by comparing the differences between the two parts, so that the position of the current frame can be predicted. Zhang et al. [15] applied the theory of compressed sensing (CS) to the target tracking and proposed a fast compression tracking (FCT) [34] method. Babenko et al. [4] proposed an online learning tracker, which conducted online training for the classifier, and put positive samples and negative samples into two different parts. Furthermore, a weight multiple instance learning (WMIL) tracker was developed [33] by adding sample weights into the learning process. Zhang et al.[35] proposed a simple and effective algorithm for online feature selection (ODFS). The objective function is optimized iteratively along the steepest ascending gradient of the positive sample and the steepest descending gradient of the negative sample to maximize the output of the weak classifier. A new online visual tracking method based on probability continuous outlier model (PCOM) was presented by Wang et al. [27], and achieved very good performance in both accuracy and speed.

As deep learning emerges in more and more areas, researchers have started to employ it to target tracking [26]. At present, most of the target tracking using deep learning belong to the discriminant approaches. Different from the significant role of deep learning in fields like object detection and recognition, the development of deep learning in the field of target tracking is not smooth,
Infrared target tracking based on proximal robust principal component analysis method

mainly due to the lack of training data. One of the core strengths of deep learning is the effective training of a large number of samples, whereas, target tracking only uses the first frame for sample training before starting. Wang et al. [28] proposed a deep learning tracker (DLT), which applied the methods of offline pre-training and online fine-tuning to neural network training, largely solving the problem of lack of samples. Zhang et al. [32] proposed a convolutional network tracker (CNT). Compared with the traditional method based on deep learning, CNT does not require a large sum of data for pre-training. In recent years, Siamese Network structure-based methods have attracted much attention in the area of object tracking, among which SiamFC [5] is the most representative one. These kinds of trackers combine the Siamese network with related operations and can always achieve satisfactory tracking result, but it is difficult to discriminate the negative samples in the real scenes with complex background. Since then, a number of similar tracking algorithms have emerged [11, 16, 36], and their performance has been continuously improved.

3 Particle filter framework

Particle filter is one of the most widely used methods based on random sampling in the field of target tracking. The tracking framework of particle filter can be regarded as a Bayesian inference problem, which contains two main steps: state update and state prediction. In the tracking process, the target state at time \( t \) can be represented by the a 6-dimensional affine component \( X_t = (h_t, w_t, s_t, r_t, \theta_t, \lambda_t)^T \) which is related to the initial time, where \( h_t \) and \( w_t \) represent the spatial position at time \( t \) , \( s_t, r_t, \theta_t \) and \( \lambda_t \) represent the scaling scale, rotation angle, aspect ratio and tilt angle. Assuming that \( Y_{1:t-1} = \{ Y_1, Y_2, ..., Y_{t-1} \} \) is the target observation from the initial time to the current time \( t \) , we can use the iterative equation shown below to predict the prior probability distribution of the current target state \( X_t \) :

\[
p(X_t|Y_{1:t-1}) = \int p(X_t|Y_{t-1})p(X_{t-1}|Y_{1:t-1})dX_{t-1} \tag{1}
\]

where, \( p(X_t|Y_{1:t-1}) \) is the transfer probability of the target state, and obeys the Gaussian distribution:

\[
p(X_t|X_{1:t-1}) \sim N(X_{t-1}; \Sigma^2) \tag{2}
\]

where, \( p(X_t|Y_{1:t-1}) \) is the diagonal covariance matrix, and its element is the variance of the affine component \( X_t = (h_t, w_t, s_t, r_t, \theta_t, \lambda_t)^T \) corresponding to the element parameter.

Since the observed \( Y_t \) is known at time \( t \), the posterior probability density can be given by the Bayes rule:

\[
p(X_t|Y_{1:t}) = \frac{p(Y_t|X_t)p(X_t|Y_{1:t})}{p(Y_t|Y_{1:t-1})} \tag{3}
\]

where, \( p(X_t|Y_{1:t}) \) is the likelihood probability to measure the similarity between the observed quantity \( Y_t \) and the actual target. Considering \( p(X_t|Y_{1:t}) \) is a constant, we can get

\[
p(X_t|Y_{1:t}) \propto p(Y_t|X_t)p(X_t|Y_{1:t-1}) \tag{4}
\]

At the current time \( t \), the target state can be gotten by solving the maximum likelihood probability \( p(X_t|Y_{1:t}) \) as (5).

\[
X_t^* = \arg \max_{X_t} p(X_t|Y_t) \tag{5}
\]

4 Model establishment

In this section, we mainly introduce the infrared target tracking model based on the approach to proximal principal component analysis. The main symbol meanings used in the model are explained in Section 4.1; the appearance model is established in Section 4.2; and the calculation algorithm is given in Section 4.3. In view of the change of target shape and attitude, a template update strategy is proposed in Section 4.4. In Section 4.5, the algorithms in this section are summarized.

4.1 Definitions

Before the start of tracking, target region of the image is manually selected, and perturbations are added to the target region to form \( i \) regions of object image. \( i + 1 \) target regions are vectorized, and then the template matrix \( F = \{ f_1, f_2, ..., f_i \}^{j \times i} \) is formed, where \( f_i \) represents the \( i \)-th column vector of the template matrix and \( j = w \times h \).

The candidate target \( m_k \in \mathbb{R}^j \times 1 \), which is normalized from \( w \times h \), and the template matrix \( \mathbf{F} \) together form the observation matrix \( \mathbf{M} = \{ \mathbf{F}, m_k \} = \{ f_1, f_2, ..., f_i, m_k \} \in \mathbb{R}^{j \times (i+1)} \). \( \mathbf{L} \) is the target matrix and has the same dimension as \( \mathbf{M} \). In other words, each column of \( \mathbf{L} \) represents an estimate of the corresponding target pixel in \( \mathbf{M} \). In this paper, \( \mathbf{S} = \{ s_1, s_2, ..., s_i, s_{i+1} \} \) is used as the occlusion matrix, where each column \( s_i \) represents the estimation of the occlusion pixel in \( \mathbf{M}_i \).

4.2 Appearance model

The observation matrix \( \mathbf{M} \) is broken down into sparse occlusion matrix \( \mathbf{S} \) and low-rank target matrix \( \mathbf{L} \). Then the target appearance model of the algorithm in this paper can be expressed as

\[
\mathbf{M} = \mathbf{L} + \mathbf{S} \tag{6}
\]

© Springer
For the purpose of reconstructing target observation matrix $\mathbf{M}$, the function is used to calculate the minimum reconstruction error, as follows:

$$\min_{\mathbf{L},\mathbf{S}} \frac{1}{2} ||\mathbf{M} - \mathbf{L} - \mathbf{S}||_F^2$$  \hspace{1cm} (7)

For a general infrared image sequence, the objects in adjacent frames have little difference in shape and gray scale, i.e., they have strong linear correlation. So these similar vectorized image patches can constitute a low-rank matrix [9]. Therefore, we can get the following constraint applied to the target matrix $\mathbf{L}$:

$$\text{rank}(\mathbf{L}) \leq \omega$$  \hspace{1cm} (8)

where, $\text{rank}(\bullet)$ is the rank operator and is a small constant. Since the nuclear norm can replace the rank operator [21], principal component analysis (RPCA) [13], we set $p < \omega$, $\lambda$ is the weight factor. According to robust principal component analysis (RPCA) [13], we set $\lambda = (1/10)\sqrt{\max(j, i + 1)}$ in this paper.

In the practical infrared target tracking process, the change of illumination, partial occlusion or total occlusion, may lead to unpredictable positioning error. According to the prior hypothesis information, the occlusion area only occupies a small part of the selected image block. So the occlusion matrix is sparse and the proximal $p$-norm can be incorporated into the model [29].

Next, the proximal $p$-norm of the matrix is defined as follows:

$$G_{\mu,p}(\mathbf{S}) = \sum_{m,n} g_{\mu,p}(s_{m,n})$$  \hspace{1cm} (9)

where, $\mu$ is the parameter and $g_{\mu,p}$ is implicitly determined by

$$|s|^2/2 + \mu g_{\mu,p}(s) = (|s|^2/2 - \mu h_{\mu,p})^+(s)$$  \hspace{1cm} (10)

where, the operator $^+$ is the Legendre-Fenchel transformation to the function [17], and $h_{\mu,p}$ is defined as

$$h_{\mu,p}(t) = \begin{cases} -t^2/2\mu & \text{if } |t| \leq \mu^{1/p} \\ -t^p/p - \delta & \text{if } |t| \geq \mu^{1/p}, p \neq 0 \\ \ln(-t) - \ln\mu/2 + \frac{1}{2} & \text{if } |t| \geq \mu^{1/p}, p = 0 \end{cases}$$  \hspace{1cm} (11)

where, $\delta = (1/p - 1/2)\mu^{2/(2-p)}$. When $p = 0$, $h_{\mu,0}(t)$ can be defined as the limit as $p$ approaches 0.

As shown in the Fig. 1, the function $g_{\mu,p}$ and $l_p$-norm have similar characteristics. When $p < 1$, $g_{\mu,p}$ rapidly drops to 0 at $x = 0$, showing that the curve has a sharp edge at $x = 0$. Then, the matrix element is 0, the penalty function value is 0. What is more, if a matrix element deviates from

![Fig. 1 Curves of $g_{\mu,p}$ with $\mu = 0.1$ and different $p$](image)

$x = 0$, the value of the penalty function increases rapidly. At last, the value of the function is nearly unchanged, when the value of the matrix element is far from 0. Therefore, the proximal $p$-norm can well constrain the sparsity of the matrix. Finally, the target apparent model of the algorithm in this paper can be expressed as

$$\mathcal{L}(\mathbf{L}, \mathbf{S}, \mathbf{I}; \mu) = G_{\mu,p}(\rho(\mathbf{S})) + \lambda G_{\mu,\lambda,p}(\mathbf{S}) + \frac{1}{2\mu} ||\mathbf{M} - \mathbf{L} - \mathbf{S}||_F^2$$  \hspace{1cm} (12)

where, the kernel norm of matrix $\mathbf{L}$ is changed to proximal $p$-norm of its singular value matrix $\rho(\mathbf{L})$. $\mathbf{I}$ is the Lagrange multiplier, $\mu$ is the scale factor, and $G_{\mu,p}$ is changed to $G_{\mu,\lambda,p}$ for the convenience of subsequent solutions.

For each frame of image, $N$ corresponding candidate targets can be obtained from modeling the state transition probability $p(\mathbf{X}_t|\mathbf{X}_{t-1})$ with Gaussian distribution. The reconstruction error matrix $\mathbf{e}_t$ of each candidate target can be calculated by

$$\mathbf{e}_t = \mathbf{M}_{1:t+1} - \mathbf{L}_{1:t+1} - \mathbf{S}_{1:t+1}$$  \hspace{1cm} (13)

where, $\mathbf{M}_{1:t+1}$, $\mathbf{L}_{1:t+1}$, $\mathbf{S}_{1:t+1}$ respectively represent the observation matrix, target matrix and occlusion matrix of the current candidate target $m_k$.

The likelihood probability $p(\mathbf{Y}_t|\mathbf{X}_t)$ can be calculated according to the reconstruction error:

$$p(\mathbf{Y}_t|\mathbf{X}_t) = \frac{1}{\sqrt{2\pi}} \prod_x \prod_y \exp \left( -\frac{(\mathbf{Y}_t - \mathbf{X}_t)^2}{2\sigma^2} \right)$$  \hspace{1cm} (14)

Equation (15) shows that $p(\mathbf{Y}_t|\mathbf{X}_t) \propto N(\mathbf{e}_t(x, y); 0, \sigma^2)$. Then, the reconstruction error matrix $\mathbf{e}_t$ follows the Gaussian distribution with the mean value of zero and the variance of $\sigma^2$.

**4.3 Solution**

Equation (13) is a multivariable optimization problem, which can be calculated by ADMM algorithm [24]. The ADMM algorithm solves one of the variables alternately and constantly updates the Lagrange multiplier, and the convergence of the Lagrange algorithm can be referred to
Infrared target tracking based on proximal robust principal component analysis method

3) Update of Lagrange multiplier \( \mu_t \)

The above equation can be decomposed into several univariate programming problems.

1) Estimation of the occlusion matrix \( S \). Equation (13) is simplified to

\[
\min_{S} G_{\lambda\mu_t,p}(S) + \frac{1}{2\lambda\mu_t} ||M - L_t - S + I_t||_F^2
\]

The analytic solution of (17) is given through the projection operator [7]:

\[
S_{t+1} = \hat{S}_P(M - L_t + I_t; \lambda\mu_t)
\]

2) Estimation of the target matrix \( L \). Similarly, (13) is simplified to:

\[
\min_{L} G_{\mu_t,p}(\sigma(L)) + \frac{1}{2\mu_t} ||M - L - S_{t+1} + I_t||_F^2
\]

The analytical solution of (19) is given by SVD decomposition method:

\[
L_{t+1} = US_P(\Sigma, \mu_t)V^T
\]

where, \( \Sigma \) is the singular value matrix of \( M - S_{t+1} + I_t \), namely \( M - S_{t+1} + I_t = USV^T \).

3) Update of Lagrange multiplier \( I \) and scale factor \( \mu \).

According to (13), the decay factor \( \rho \) is multiplied in each iteration, and \( \mu \) is geometrically attenuated. When the convergence condition \( ||M - L - S_t||_F/||M|| < 1e-5 \) is satisfied, the iteration terminates and the detection result is obtained. The convergence and stability of the algorithm can be referred to [9] and [17].

4.4 Template update strategy

In the tracking process, both the tracking environment (such as occlusion, illumination, etc.) and the appearance of the target (such as scale, shape, etc.) may change, resulting in some templates being not applicable. Therefore, in case of significant deformation or occlusion mutation of the target, the elements in the template matrix \( T \) must be updated in time. However, the templates should not be updated too frequently. Otherwise, the cumulative tracking error will be exacerbated. On the other hand, it also cannot be too slow in order to avoid the drifting from ground truth. It is worth mentioning that the template matrix cannot be updated when the target is occluded [8, 23].

We use the reconstruction error \( \varepsilon_k \) of the current frame tracking results \( y \) to calculate the importance weight \( w_k \), and give the weights to each element of the template matrix \( F_{1,i} \). Specifically, the iteration strategy as \( w_k = w_k \cdot \exp(-||\varepsilon_k(,k)||_F) \) is adopted, where \( w_k \) is initialized to \( w_k = 1/i \) in the first frame. In addition, the sum of the \( (i+1) \)-th column vector \( \sum(S_{t+1}) \) in the current occlusion matrix \( S \) is applied to measure the occlusion. If \( \sum(S_{t+1}) \) is above the threshold \( \xi^* \), \( \xi^* \in [0,1] \), we take the opinion that there are obvious occlusion in the current frame, thus the template matrix \( T \) cannot be updated. In conclusion, the template update steps are summarized in Algorithm 1.

In this paper, cosine similarity, the included angle between vectors, is applied to express the similarity between two vectors. Obviously, the greater the cosine similarity, the greater the difference between the current frame tracking result and the elements in the template matrix. According to the template update strategy, we replace the template of the minimum weight in with the current tracking result when the difference between the tracking result and the template matrix template is larger than a certain value. Of course, there is also the premise that occlusion is not too serious. In addition, the weight of the newly added template is established to the median value within all the templates, to prevent it from occupying the absolute dominance and affecting the tracking ability.

Algorithm 1 Template update strategy.

Input: the template matrix \( T \), the pre-defined thresholds \( \psi^* \) and \( \xi^* \).
1: Initialize: \( w_0 = 1/k \) \( (k = 1, 2, ..., i) \), where \( i \) is the number of templates.
2: for \( t = 2, 3, \) do
3:     Calculate the current tracking result \( y_k \), the corresponding target matrix \( L \), the occlusion matrix \( S \), and then calculated the reconstruction error \( \varepsilon_k \) based on (14).
4:     Update the template weight as \( w_k = w_k \cdot \exp(-||\varepsilon_k(,k)||_F) \).
5:     Calculate the cosine similarity as \( \psi_k = \arccos(m, \xi) \).
6:     Select the minimum value \( \psi_{min} \) in cosine similarity.
7:     if \( \psi_{min} > \psi^* \) and \( \sum(s_{n+1}) > (\xi^* \cdot m) \) then
8:         Replace the least weighted template in \( F \) with \( \tilde{m} \), i.e. \( \tilde{m} = \hat{m} \), where \( c = \arg \min w_k \).
9:         The new template is assigned a new weight \( w_k = \text{median}(w) \), where median(\( ) \) represents the median value.
10:    end if
11:   Normalize the template weight \( w_k = w_k/\sum(w) \).
12:   Adjust \( w \) to insure that \( \max(w) = 0.3 \).
13:   Output: updated template matrix \( F \) and corresponding weight \( w_k \).
14: end for
4.5 Summary of the proposed tracker

In this section, we summarize the proposed infrared target tracking algorithm based on proximal robust principal component analysis method.

Algorithm 2  The target tracking with complete steps.

Input: the target box selected manually at the first frame.
1: Initialize: target state vector \( X_1 \), template matrix \( F = \{ f_1, f_2, ..., f_i \} \in \mathbb{R}^{j \times i} \).
2: \textbf{for} \( t = 2 \) to the last frame \textbf{do}
3: \textbf{for} \( k = 1 \) to the last particle \textbf{do}
4: \quad Collect candidate sample \( m_k \in \mathbb{R}^{j \times 1} \), forming observation matrix \( M = \{ f_1, f_2, ..., f_i, m_k \} \in \mathbb{R}^{j \times (i+1)} \).
5: \quad Initialize: \( \lambda = \left( \frac{1}{10} \right) \sqrt{\max(j, i+1)} \).
6: \quad \textbf{loop}
7: \quad \textbf{L-step:} \( \mathbf{L}_{t+1} = \arg \min \mathcal{L}(\mathbf{L}, \mathbf{S}_{t+1}, \mathbf{L}_t; \mu_t) \).
8: \quad \textbf{S-step:} \( \mathbf{S}_{t+1} = \arg \min \mathcal{L}(\mathbf{L}_t, \mathbf{S}_t, \mathbf{L}_t; \mu_t) \).
9: \quad \textbf{T-step:} \( \mathbf{I}_{t+1} = \mathbf{I}_t + \frac{1}{\mu_t} (M - \mathbf{L}_{t+1} - \mathbf{S}_{t+1}) \).
10: \quad \textbf{end loop}
11: \textbf{end for}
12: Utilize the likelihood probability \( P(Y_t|X_t) \) to resample.
13: \textbf{Output:} \( X_t^* = \arg \max \ p(X_t|Y_t) \).
14: \textbf{end for}

5 Experimental results

In this section, 8 groups of typical infrared sequences are utilized to test the proposed infrared target tracking algorithm. Then, the algorithm is compared quantitatively and qualitatively with other typical tracking algorithms.

5.1 Datasets

The interference in 8 test sequences include pose variation, scale change, occlusion and so on. Some sequences can be downloaded from [1, 2]. Details of each sequence are shown in Table 1, and the target of the first frame is manually selected (marked with red rectangle), as shown in Fig. 2.

5.2 Parameter setting

Table 2 lists all the parameter values involved in this algorithm, as well as their corresponding meanings and default values.

5.3 Results

In our proposed tracking algorithm, the number of templates is set as 10, and the number of particles is set as 500. Five results of each sequence images are shown in Fig. 3. In Fig. 3a, the target does not change significantly in appearance, nor does the template. The difficulty of this sequence is that the target is partially or completely obscured by trees around 245-th frame. This problem is easy to solve because of the occlusion matrix in our algorithm. The size and posture of the rhino in Seq.2 are constantly changing, and there is also partial occlusion. Owing to the occlusion matrix and template update strategy, it has good tracking result. In Seq.3, the difficulty lies in the encounter with another person when the man is moving, which causes the remarkable overlap. Considering the low-rank property of the target matrix and the sparsity of the occlusion matrix, this kind of tracking performance is well. In addition, Seq.4 and Seq.5 are similar to Seq.2, with scale change and pose change. Seq.6 is similar to Seq.3 in which there is target overlap. Seq.7 and Seq.8 have a similar difficulty that the size of target is too small, only about 15-20 pixels.

5.4 Qualitative comparisons

This section presents the algorithm compared with 13 kinds of conventional algorithms, these algorithms including FCT [34], CNT [32], DLT [35], IVT [22], Meanshift (MS) [10], ODFS [35], L1APG [18], PCOM [27], WMIL [33], SiamFC [5], SiamRPN [16], DaSiamRPN [36], and SiamCAR.
Table 2 List of parameter values

| Parameter | Meaning                  | Default Value |
|-----------|--------------------------|---------------|
| $n$       | template number          | 10            |
| $p$       | particle number          | 500           |
| $\xi^*$   | Occlusion discrimination threshold | 0.1    |
| $\psi^*$  | cosine similarity threshold | 30     |

The parameter values of all comparison algorithms are selected as their default parameters, and the particle numbers of DLT, CNT, IVT and $L_1$APG based on particle filter framework are uniformly set to 500. Figure 3 shows part of the tracking results. FCT, ODFS and WMIL use haar-like features to represent the local information of the target. As shown in Fig. 4c, f, the algorithms were not able to track the object.
Fig. 4 Tracking results of the 14 tracking algorithms over datasets: a man; b rhino; c man; d horse; e man; f man; g drone; h drone
Infrared target tracking based on proximal robust principal component analysis method
of interest well when there was target aliasing in Seq.3 and Seq.6.

MS and FCT can only achieve good tracking effect when the change of target’s appearance is not drastic and there is not too much shielding. Because these two algorithms only take pixel gray level as the feature of the object appearance modeling. It is obvious that they cannot deal with relatively complex tracking scenes.

PCOM can utilize PCA subspace to estimate the change of target’s appearance. As the figure shows that the algorithm achieved satisfactory tracking effect in Seq.3 and Seq.4. If the adjacent occlusion pixel is misjudged as the main component while the target pixel is identified as a singular value, the algorithm will drift seriously. Because the size of infrared small target is too dim, it is difficult to extract its features. As shown in the Fig. 4g, h, the tracking effect is poor.

The L1APG uses the fragment template to represent the interferences such as occlusion and noise, so it performs stably in most test sequences (such as Seq.1, Seq.2, Seq.3, Seq.6, Seq.7, and Seq.8). Due to the random distribution of non-zero elements in the fragment template, the optimization equation based on $l_1$-norm may converge to the wrong value, resulting in tracking failure in some scenarios, such as Seq.4 and Seq.5.

DLT and CNT are convolutional network-based tracking algorithms which need to train the dataset before tracking. Compared with other comparison algorithms, the tracking results of these two algorithms are more accurate in general. However, DLT drifted on Seq.3, Seq.5, Seq.7, and Seq.8, indicating that it was not stable for overlap, appearance variation and small target size. CNT performed poorly in Seq.5 and Seq.6, because of size changing and target overlapping.

SiamFC, SiamRPN, DaSiamRPN, and SiamCAR combied Siamese Network with correlation operations, and obtained excellent results on the whole. However, it is difficult to suppress the difficult samples in the background. Therefore, when the targets severely deform, or overlap, it is easy to make a mistake in the tracking, such as Seq.4 and Seq.6. Generally deep learning methods require multiple convolution, and it is difficult to fully extract features of small targets, resulting in decreased accuracy. SiamRPN and SiamRPN are not missing the target of Seq.7 and Seq.8, but the accuracy is still to be improved.

On the contrary, the performance of the proposed tracker is stable, and it can track the target accurately in all the selected sequences. Experimental results shown above demonstrate that it is robust to occlusion, deformation, scale change and target overlap when compared to other traditional methods.

### 5.5 Quantitative comparisons

In this section, we take two metrics to quantify the tracking results normalized. There are normalized center position error and average overlap score (AOS) \[30, 31\]. The normalized center position error $\epsilon_0$ is defined as:

$$\epsilon_0 = \frac{\sqrt{(x_t - x_0)^2 + (y_t - y_0)^2}}{L_0} \quad (21)$$

where, $(x_t, y_t)$ is the central coordinate of the tracking result, $(x_0, y_0)$ is the central coordinate of the ground truth, and $L_0$ is the diagonal length of the truth rectangular box.

The average normalized center position error $\bar{\epsilon}_0$ is listed in Table 3. At the same time, the precision plot introduced in Fig. 5 intuitively describes the tracking precision, which is defined as the percentage of image numbers in which the center position error is smaller than a specific value.

As shown in Table 3 and Fig. 5, the algorithm in this paper has the smallest overall center error in the six test sequences and is more ahead of other algorithms. The IVT and MS tracking results deviate from the truth value due to the severe occlusion in Seq.1. In Seq.2, all algorithms performed well and no targets were lost.

|           | FCT  | CNT  | DLT  | IVT  | MS   | ODFS | WMIL | L1APG | PCOM | SiamRPN | DaSiamRPN | SiamFC | SiamCAR | Ours |
|-----------|------|------|------|------|------|------|------|-------|------|---------|-----------|--------|---------|------|
| Seq.1     | 0.0749 | 0.0324 | 0.1436 | 1.7798 | 1.6545 | 0.1415 | 0.1531 | 0.0638 | 0.1069 | 4.0380  | 0.0932    | 0.0476 | 0.0431 | 0.0564|
| Seq.2     | 0.0757 | 0.2371 | 0.1338 | 0.1785 | 0.1506 | 0.2605 | 0.2434 | 0.0637 | 0.1792 | 0.0604  | 0.1153    | 0.0465 | 0.0284 | 0.1099|
| Seq.3     | 0.1314 | 0.0302 | 0.0655 | 0.0207 | 0.1252 | 0.2480 | 0.1770 | 0.0210 | 0.0629 | 0.0622  | 0.0865    | 0.0386 | 0.2536 | 0.0236|
| Seq.4     | 0.1058 | 0.0599 | 0.0558 | 0.0341 | 0.2152 | 0.0885 | 0.1427 | 0.5369 | 0.0846 | 0.1205  | 0.1783    | 0.0695 | 0.0822 | 0.0772|
| Seq.5     | 0.3295 | 0.8714 | 0.7897 | 0.8657 | 0.6845 | 0.1957 | 0.1806 | 0.1381 | 0.8103 | 0.0655  | 0.0843    | 0.0742 | 0.0483 | 0.0432|
| Seq.6     | 2.0589 | 2.0386 | 0.0579 | 1.9819 | 0.1045 | 2.0852 | 0.2206 | 0.0401 | 2.1962 | 0.1064  | 0.1704    | 2.0129 | 2.0069 | 0.0286|
| Seq.7     | 17.115 | 0.0391 | 15.192 | 0.0809 | 0.6722 | 17.365 | 16.820 | 0.1162 | 12.848 | 0.0770  | 0.0923    | 0.0670 | 0.0580 | 0.0318|
| Seq.8     | 35.669 | 0.1072 | 15.450 | 2.6135 | 0.3837 | 35.796 | 35.796 | 0.1176 | 2.9677 | 0.1282  | 0.1725    | 0.2330 | 0.0893 | 0.0914|
| Ave.      | 6.9450 | 0.5452 | 0.2077 | 0.8101 | 0.4890 | 0.5032 | 0.1862 | 0.1439 | 0.5733 | 0.7422  | 0.1213    | 0.3815 | 0.4104 | 0.0564|
Fig. 5 Precision plots: a man; b rhino; c man; d horse; e man; f man; g drone; h drone

Infrared target tracking based on proximal robust principal component analysis method
In Seq.3, some tracking algorithms based on haar-like characteristics showed different degrees of deviation after target overlap. The reason why \( \text{L}_1 \)APG did not perform well in Seq.4 is that the posture of this group of sequences changed too much. In Seq.5, except the algorithm in this paper, all other algorithms lose targets to varying degrees, which is caused by scale changes and target overlap in this group of sequences at the same time, but this is what our algorithm is good at. In Seq.6, FCT, CNT, IVT, ODFS, PCOM, SiamFC, and SiamCAR all have target loss after target overlap. In the last 2 sets of sequences, PCOM, IVT, DLT, and ODFS have trouble tracking the targets, because they are not sensitive enough to small targets.

Unlike \( \varepsilon_0 \), AOS further introduces the consideration of target scale when evaluating algorithm tracking progress[20]. Assuming that the target boundary box obtained from the algorithm is \( R_p \), and the boundary box of the target truth box is \( R_q \), then AOS is defined as

\[
\text{AOS} = \frac{|R_p \cap R_q|}{|R_p \cup R_q|} \quad (22)
\]

Table 4 lists the average values of AOS on each sequence. At the same time, Fig. 6 shows the success plot of each algorithm on the test sequence. The abscissa is the AOS value between [0,1], and the ordinate is the percentage of image frames corresponding to AOS. As AOS increases from 0 to 1, the success rate decreases to 0.

The algorithm in this paper has achieved an excellent success rate in each test sequence and is obviously superior to other algorithms on the whole. In Seq.1, the CNT and DLT performed poorly in terms of success rate compared to excellent performance in accuracy, because such algorithms track local information about the target rather than the whole. ODFS and WMIL have a low success rate in Seq.2 due to the change of target pose and partial occlusion. All algorithms can track the target in Seq.3, but the success rate of DLT and MS is low. The reason for the poor performance of MS and \( \text{L}_1 \)APG in Seq.4 is that there is target pose change, and these two algorithms cannot deal with such problem. In Seq.5, scale change and target overlap occur simultaneously, and only the algorithm in this paper achieves excellent results. Other algorithms offset the target to different degrees. In Seq.6, FCT, CNT, IVT, ODFS, SiamFC, and SiamCAR tracked the wrong targets due to target overlap. In the last 2 sequences, PCOM, IVT, DLT, ODFS, and WMIL have unsatisfactory results since they failed to track the infrared small target. In our algorithm, the low-rank constraint is applied to the target matrix which is constituted by those similar vectorized image patches. In this way, the similarity between the reconstructed target and the target templates can be guaranteed. At the same time, the template is dynamically adjusted to reduce the cumulative error in multiple frames. Experimental results show that the proposed algorithm achieves excellent precision rate and success rate in those two sequences.

Table 4 The average values of AOS

|       | FCT | CNT | DLT | IVT | MS | ODFS | WMIL | \( \text{L}_1 \)APG | PCOM | SiamRPN | DaSiamRPN | SiamFC | SiamCAR | Ours |
|-------|-----|-----|-----|-----|----|------|------|-----------------|------|---------|-----------|--------|---------|------|
| Seq.1 | 0.7026 | 0.5676 | 0.2120 | 0.0721 | 0.2427 | 0.6173 | 0.6008 | 0.7004 | 0.3228 | 0.2609 | 0.6056 | 0.7600 | 0.6778 | 0.7217 |
| Seq.2 | 0.7267 | 0.4792 | 0.4071 | 0.5772 | 0.6079 | 0.3799 | 0.3812 | 0.7635 | 0.5518 | 0.7513 | 0.6294 | 0.7774 | 0.8812 | 0.5311 |
| Seq.3 | 0.5942 | 0.8499 | 0.5181 | 0.8965 | 0.5184 | 0.5139 | 0.5445 | 0.8843 | 0.6902 | 0.7497 | 0.6362 | 0.8448 | 0.4979 | 0.8819 |
| Seq.4 | 0.5572 | 0.7738 | 0.5519 | 0.7076 | 0.4182 | 0.5570 | 0.5204 | 0.3817 | 0.5834 | 0.721 | 0.6509 | 0.7020 | 0.7826 | 0.7338 |
| Seq.5 | 0.3980 | 0.2197 | 0.1944 | 0.2179 | 0.3039 | 0.5259 | 0.4587 | 0.5145 | 0.2281 | 0.5079 | 0.5045 | 0.7071 | 0.6929 | 0.7664 |
| Seq.6 | 0.3245 | 0.4275 | 0.6437 | 0.2281 | 0.5814 | 0.3434 | 0.5299 | 0.3767 | 0.3221 | 0.5646 | 0.4812 | 0.4158 | 0.3769 | 0.8451 |
| Seq.7 | 0 | 0.8708 | 0.0044 | 0.6637 | 0.5724 | 0 | 0.7431 | 0.0970 | 0.5433 | 0.5282 | 0.8403 | 0.7174 | 0.8911 |
| Seq.8 | 0 | 0.6846 | 0.0176 | 0.1083 | 0.4819 | 0 | 0.7438 | 0.0246 | 0.3091 | 0.2276 | 0.5167 | 0.8079 | 0.7885 |
| Ave.  | 0.4129 | 0.6091 | 0.3186 | 0.4339 | 0.4658 | 0.3671 | 0.3794 | 0.6835 | 0.3525 | 0.5509 | 0.5329 | 0.6955 | 0.6793 | 0.7699 |

6 Conclusion

A target tracker based on approximate robust principal component analysis is proposed in this paper. Firstly, the observation matrix is decomposed into a sparse occlusion matrix and a low-rank target matrix, and the constraint optimization is carried out with an approximate norm that is closer than the approximation. In order to solve this convex optimization problem, ADMM algorithm is used to solve one variable alternately. For long-term tracking, this algorithm uses the Bayesian state inference and template update strategy under the particle filter framework. Through a series of experiments on real infrared target sequences, the effectiveness and robustness of the algorithm are verified, and it is better than other advanced algorithms.

Although the algorithm has achieved excellent results in testing infrared sequences, there are still some interferences, such as non-rigid motion and long-term occlusion, which...
Fig. 6 Success plots: a man; b rhino; c man; d horse; e man; f man; g drone; h drone
deserve further study. To cope with these disturbances, we plan to take other measures, such as improving appearance model and introducing motion estimation. At the same time, we hope to port MATLAB code to GPU or FPGA, to achieve real-time tracking.

Acknowledgements This work was supported jointly by the National Natural Science Foundation of China (62001234), Natural Science Foundation of Jiangsu Province(BK20200487), China Postdoctoral Science Foundation (2020M681597), Postdoctoral Research Funding Program of Jiangsu Province(2020Z051) and Shanghai Aerospace Science and Technology Innovation Foundation(SAST2020-071).

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

1. Otcbvs benchmark dataset collection. available online:. http://vcipolstate.org/phvs/bench/.(Accessed on 9 Jan 2020)
2. Vot-tir2015 dataset. available online: http://www.votchallenge.net/vot2015/dataset.html
3. Allen JG, Xu RY, Jin JS et al (2004) Object tracking using camshift algorithm and multiple quantized feature spaces. In: ACM international conference proceeding series, vol 100. Citeseer, pp 3–7
4. Babenko B, Yang MH, Belongie S (2010) Robust object tracking with online multiple instance learning. IEEE Trans Pattern Anal Mach Intell 33(8):1619–1632
5. Bertinetto L, Valmadre J, Henriques JF, Vedaldi A, Torr PH (2016) Fully-convolutional siamese networks for object tracking. In: European conference on computer vision. Springer, pp 850–865
6. Borji A, Itti L (2012) Exploiting local and global patch rarities for saliency detection. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp 478–485
7. Boyd S, Parikh N, Chu E (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc
8. Cao Y, Wang G, Yan D, Zhao Z (2016) Two algorithms for the detection and tracking of moving vehicle targets in aerial infrared image sequences. Remote Sens 8(1):28
9. Chartrand R (2012) Nonconvex splitting for regularized low-rank+ sparse decomposition. IEEE Trans Signal Process 60(11):5810–5819
10. Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. IEEE Trans Pattern Anal Mach Intell 25(5):564–577
11. Guo D, Wang J, Cui Y, Wang Z, Chen S (2020) Siamicar: Siamese fully convolutional classification and regression for visual tracking. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 6269–6277
12. Guo H (2016) An efficient object tracking algorithm of sports video. In: 2015 5th International conference on computer sciences and automation engineering (ICCSAE 2015). Atlantis Press
13. Huang S, Ye J, Wang T, Jiang L, Wu X, Li Y (2015) Extracting refined low-rank features of robust pca for human action recognition. Arab J Sci Eng 40(5):1427–1441
14. Isard M, Blake A (1998) Condensationconditional density propagation for visual tracking. Int J Comput Vis 29(1):5–28
15. Kaibua Z, Lei Z, Ming-Hsuan Y (2012) Real-time compressive tracking. In: Proceedings of the 12th European conference on computer vision, pp 864–877
16. Li B, Yan J, Wu W, Zhu Z, Hu X (2018) High performance visual tracking with siamese region proposal network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8971–8980
17. Lin Z, Chen M, Ma Y (2010) The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv:1009.5055
18. Mei X, Ling H, Wu Y, Blasch E, Bai L (2011) Minimum error bounded efficient / 1 tracker with occlusion detection. In: CVPR 2011. IEEE, pp 1257–1264
19. Nam H, Baek M, Han B (2016) Modeling and propagating cnns in a tree structure for visual tracking. arXiv:1608.07242
20. Parate MR, Satpute VR, Bhurchandi KM (2018) Global-patch-hybrid template-based arbitrary object tracking with integral channel features. Appl Intell 48(2):300–314
21. Recht B, Fazel M, Parrilo PA (2010) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM review 52(3):471–501
22. Ross DA, Lim J, Lin RS, Yang MH (2008) Incremental learning for robust visual tracking. Int J Comput Vis 77(1-3):125–141
23. Tang H, Sun W, Yu H, Lin A, Xue M, Song Y (2019) A novel hybrid algorithm based on pso and foa for target searching in unknown environments. Appl Intell 49(7):2603–2622
24. Wan M, Gu G, Qian W, Ren K, Chen Q, Zhang H, Maldague X (2018) Total variation regularization term-based low-rank and sparse matrix representation model for infrared moving target tracking. Remote Sens 10(4):510
25. Wan M, Gu G, Qian W, Ren K, Maldague X, Chen Q (2019) Unmanned aerial vehicle video-based target tracking algorithm using sparse representation. IEEE Internet of Things Journal 6(6):9689–9706
26. Wan M, Ye X, Zhang X, Xu Y, Gu G, Chen Q (2021) Infrared small target tracking via gaussian curvature-based compressive convolution feature extraction. IEEE Geosci Remote Sens Lett: 1–5. https://doi.org/10.1109/LGRS.2021.3051183
27. Wang D, Lu H (2014) Visual tracking via probability continuous outlier model. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3478–3485
28. Wang N, Yeung DY (2013) Learning a deep compact image representation for visual tracking. In: Advances in neural information processing systems, pp 809–817
29. Wang P, Chen Q, Qian W, Ren K, Yao Z, Xu F (2017) Salient motion detection using proximal robust principal component analysis method. J Electronic Imaging 26(2):023004
30. Wang Y, Hu S, Wu S (2015) Visual tracking based on group sparsity learning. Mach Vis Appl 26(1):127–139
31. Wu Y, Lim J, Yang MH (2013) Online object tracking: A benchmark. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2411–2418
32. Zhang K, Liu Q, Wu Y, Yang MH (2016) Robust visual tracking via convolutional networks without training. IEEE Trans Image Process 25(4):1779–1792
33. Zhang K, Song H (2013) Real-time visual tracking via online weighted multiple instance learning. Pattern recognition 46(1):397–411
34. Zhang K, Zhang L, Liu Q, Zhang D, Yang MH (2014) Fast visual tracking via dense spatio-temporal context learning. In: European conference on computer vision. Springer, pp 127–141
35. Zhang K, Zhang L, Yang MH (2013) Real-time object tracking via online discriminative feature selection. IEEE Trans Image Process 22(12):4664–4677
36. Zhu Z, Wang Q, Li B, Wu W, Yan J, Hu W (2018) Distractor-aware siamese networks for visual object tracking. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 101–117

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

Chao Ma received his B.S. degree in electronic science and technology from Nanjing University of Science & Technology, Nanjing, China, in 2018. He is currently pursuing his Ph.D. degree in optical engineering from Nanjing University of Science and Technology. His main research interests include image processing, computer vision and infrared imaging.

Minjie Wan received his B.S. degree and Ph.D. degree in electronic science and technology from Nanjing University of Science & Technology, Nanjing, China, in 2014 and 2020, respectively. He was a visiting Ph.D. student with the Department of Electrical and Computing Engineering, Université Laval, Quebec, Canada, from 2017-2018. He is currently a post-doctoral with the School of Electronic and Optical Engineering, Nanjing University of Science & Technology, Nanjing, China. His current research interests include optical signal processing, infrared imaging and polarization imaging.

Yunkai Xu received his B.S. degree in electronic science and technology from Nanjing University of Science & Technology, Nanjing, China, in 2020. He is currently studying for the M.S. degree in optical engineering from Nanjing University of Science & Technology, Nanjing, China. His main research interests include computer vision and machine learning.

Qian Chen received the B.S. and M.S. degrees in optoelectronic technology from Nanjing University of Science & Technology, in 1987 and 1991, respectively, and the Ph.D. degree in optical engineering from Nanjing University of Science & Technology, Nanjing, China, in 1996. Since 1996, he has been a Professor with the School of Electronic and Optical Engineering, Nanjing University of Science & Technology, Nanjing, China. His current research interests include optical design and computer version.

Kan Ren received the B.S. degree in optical information science and technology from Nanjing University of Science and Technology, Nanjing, China, in 2006 and the Ph.D. degree in electronic engineering from University of Surrey, Guildford, UK in 2011. Since 2011, he has been an Associate Professor with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China. His current research interests include computer vision and machine learning.
Guohua Gu received the B.S. and M.S. degrees in optical instrument from Nanjing University of Science & Technology, Nanjing, China, in 1989 and 1996, respectively, and the Ph.D. degree in optical engineering from Nanjing University of Science & Technology, Nanjing, China, in 2001. Since 2007, he has been a Professor with the School of Electronic and Optical Engineering, Nanjing University of Science & Technology, Nanjing, China. His current research interests include optical design, computer vision and machine learning.