Copyright ⓒ 2019 KSII

# Multi-feature local sparse representation for infrared pedestrian tracking

**Xin Wang[1,2*], Lingling Xu[1] and Chen Ning[3]**
[1]College of Computer and Information, Hohai University
Nanjing, Jiangsu 211100 - China
[e-mail: wang_xin@hhu.edu.cn]
[2]Jiangsu Key Laboratory of Image and Video Understanding for Social Safety,
Nanjing University of Science and Technology,
Nanjing, Jiangsu 211100 - China
[3]School of Physics and Technology, Nanjing Normal University
Nanjing, Jiangsu 210023 - China
*Corresponding author: Xin Wang

*Received December 19, 2017; revised June 10, 2018; revised September 16, 2018; accepted October 9, 2018; published March 31 2019*

## Abstract

Robust tracking of infrared (IR) pedestrian targets with various backgrounds, e.g. appearance changes, illumination variations, and background disturbances, is a great challenge in the infrared image processing field. In the paper, we address a new tracking method for IR pedestrian targets via multi-feature local sparse representation (SR), which consists of three important modules. In the first module, a multi-feature local SR model is constructed. Considering the characterization of infrared pedestrian targets, the gray and edge features are first extracted from all target templates, and then fused into the model learning process. In the second module, an effective tracker is proposed via the learned model. To improve the computational efficiency, a sliding window mechanism with multiple scales is first used to scan the current frame to sample the target candidates. Then, the candidates are recognized via sparse reconstruction residual analysis. In the third module, an adaptive dictionary update approach is designed to further improve the tracking performance. The results demonstrate that our method outperforms several classical methods for infrared pedestrian tracking.

***Keywords:*** Infrared, pedestrian tracking, sparse representation, multiple features

This work was supported in part by the Fundamental Research Funds for the Central Universities (Grant No. 2019B15314, 30918014107), in part by the National Natural Science Foundation of China (Grant No. 61603124), in part by the Jiangsu Government Study Scholarship, in part by the Six Talents Peak Project of Jiangsu Province (Grant No. XYDXX-007), and in part by the 333 High-Level Talent Training Program of Jiangsu Province.

**http://doi.org/10.3837/tiis.2019.03.020** **ISSN : 1976-7277**

## 1. Introduction

Infrared (IR) pedestrian tracking is a vital problem in infrared image analysis, and is important for a great number of practical applications, e.g. human motion analysis, video surveillance and monitoring. However, the infrared pedestrian image sequences usually have complex backgrounds, making the tracking task much difficult [1].

Decades of study on this issue have generated a series of approaches [2-13]. Thereinto, particle filter (PF) has gotten particular attention for the capability of solving non-linear and non-Gaussian questions [2-4]. Also, Gaussian mixture model (GMM) has been exploited for extracting foreground candidates from background [5]. Mean shift-based tracking technique has been put forward as an expeditous technique [6-10]. In [11], spatial-temporal filters have been designed to track infrared target. The dense structural learning has been proposed to train a classifier with dense samples through Fourier techniques for infrared object tracking [12]. In [13], generative and discriminative ideas have been adopted.

Currently, sparse representation (SR) based tracking methods have gained substantial interest [14-16]. Its main idea is that, for current frame, object candidates are sparsely represented and that having the lowest reconstruction error is thought to be the real target [17-19]. Many works have shown the effectiveness of such methods, but there still exist two critical problems. (1) Targets to be tracked are always thought to be holistic entities by SR. Consequently, when they face the difficulties of appearances changes, illumination variations, etc., they cannot guarantee the tracking performance and tend to fail. (2) At present, most SR based methods only rely on widely used feature, i.e., the gray feature for infrared videos, since gray is thought to be the most salient feature for infrared targets. Nevertheless, it may fail while encountering interferences with similar gray values.

In this paper, we solve the above challenges by proposing a novel infrared pedestrian tracking method. This method involves three important contributions. (1) Unlike most existing SR approaches, the addressed algorithm is to utilize local sparse representation to model the target locally. Compared with holistic description, local representation is more robust to variations. (2) Instead of only using gray cue, our method also extracts the edge feature for infrared pedestrians to enhance the robustness of the target model. (3) For robust tracking, researchers have proposed various approaches with regard to target model update, most of which update the model via the current frame tracking results. However, if the results are contaminated, the updated model will be inaccurate and some errors may be introduced in the tracking process. When the errors are accumulated to a certain extent, serious drifting may occur. To prevent the drifting problem, an adaptive dictionary update approach is designed, which judge whether the present target is dirtied before target feature set renovation. The current target feature set is only updated when the result is not dirtied. This scheme is very helpful for improving tracking robustness.

The rest is arranged as below. SR theory is reviewed in Section 2. Our technique is introduced in Section 3. Section 4 gives the experimental results. Section 5 draws the conclusion.

## 2. Sparse Representation

The aim of SR is to seek sparsely representations for signals [20-24]. Given signals $Y=[y_1,y_2,\cdots,y_N]\in R^{n\times N}$, a reconstructive dictionary $D=[d_1,d_2,\cdots d_K]\in R^{n\times K}$ ($K>n$) is learnt as below in SR [25]:

$$\langle D,X\rangle=\arg\min_{D,X}\|Y-DX\|_2^2 \quad s.t. \quad \forall i,\|x_i\|_0\le T \tag{1}$$

where $X=[x_1,x_2,\cdots x_N]\in R^{K\times N}$ represent sparse codes, while $\|Y-DX\|_2^2$ is sparse error. $T$ is a constraint factor. $\|\cdot\|_0$ denotes $L_0$ norm [26].

Suppose $D$ is fixed, the sparse representation $x_i$ of $y_i$ can be calculated as [27-29]:

$$x_i=x^*(y_i,D)\equiv\arg\min_x\|y_i-Dx\|_2^2 \quad s.t. \quad \|x\|_0\le T \tag{2}$$

Then, $X$ and $D$ are updated as $\tilde{X}_i$ and $\tilde{D}_i$ consistently, and thus:

$$E_i=Y-\tilde{D}_i\tilde{X}_i \tag{3}$$

## 3. Presented Method

The overall framework of our technique is shown in **Fig. 1**. The first step is to develop a multi-feature local SR model for target to be tracked. Second, a tracker is developed. Furthermore, an adaptive dictionary update approach is designed for further robustness.

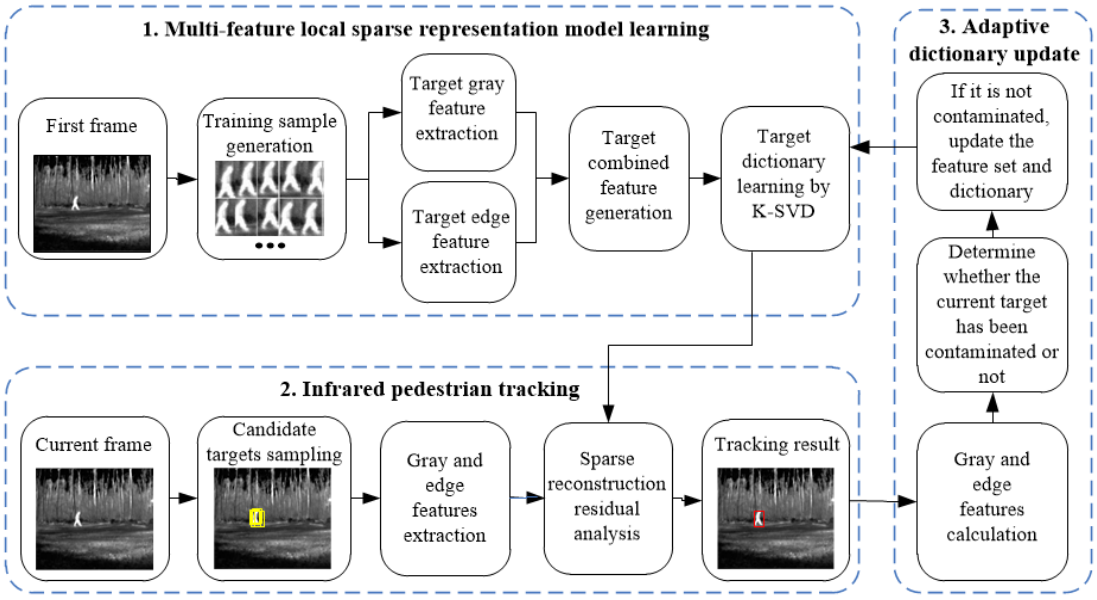


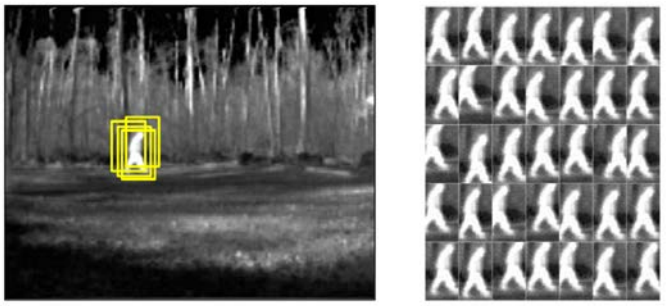

**Fig. 1.** Framework of our method.

### 3.1 Multi-feature Local Sparse Appearance Model

#### 3.1.1 Training Samples Construction

First, we present to sample a number of templates for object to be tracked by using a patch-based scheme.

As shown in **Fig. 2**, a sliding window with the size of $m \times n$ is used to sample $N$ target templates $T = [t_1, t_2, \cdots, t_N]$ from the first frame $I$ of an IR video.



(a) First frame (b) Some target templates

**Fig. 2.** Training samples construction.

#### 3.1.2 Target Gray Feature Extraction

Gray feature is the most widely selected characteristic for infrared image sequences [6], since infrared targets often possess higher gray values than static background areas.

For a target template $t_i$, we utilize the gray histogram to describe the gray characteristics of it. Suppose pixel locations in target area are $\{x_i\}_{i=1,\cdots,M}$, gray histogram $p = \{p(u)\}_{u=1}^{L_G}$ of target can be calculated by:

$$p(u) = \sum_{i=1}^{M} \delta(b(x_i) - u), \quad u = 1, \cdots, L_G \tag{4}$$

where $b(x_i)$ is the gray mapping function of the pixel point $x_i$. $L_G$ is the gray mapping level. Normally, the gray histogram needs to be normalized as $\sum_{u=1}^{L_G} p(u) = 1$.

By observing the Eq. (4), we find that the traditional gray histogram lacks of spatial position information of the pixel. In fact, different pixels in the target area make different contributions to the description of the target gray. The traditional gray histogram will cause the pixels that are closer to the target center and have greater contribution to gray description are not very prominent. Therefore, we use a weighting function [4] to include the spatial distribution of pixels in the histogram. The weighting function is described as:

$$k(r) = \begin{cases} 1 - r^2, & r < 1 \\ 0, & r \geq 1 \end{cases} \tag{5}$$

where $r$ denotes the distance between the pixel and the center of target. By using such kernel function, we can obtain the modified gray histogram of the target template:

$$p(u) = C_1 \sum_{i=1}^{M} k\left( \left\| \frac{x_0 - x_i}{h} \right\|^2 \right) \delta(b(x_i) - u), \quad u = 1, \cdots, L_G \tag{6}$$

where $x_i$ denotes the position of a pixel in target area. $x_0$ denotes the central location of the target area. $h$ denotes the size of target area. $M$ represents the total number of pixels of target. $C_1$ is the normalization constant.

Thus, the probability density at each gray level of the target template can be computed by Eq. (6). And the corresponding gray feature vector of the target template can be contained. Then, the gray feature vectors of all target templates are quantified respectively, and a gray feature set with spatial location information can be formed:

$$P_G = [p_1, p_2, \cdots, p_N] \tag{7}$$

where $p_j \in R^{L_G \times 1}$ $(j = 1, 2, \cdots, N)$ denotes the gray feature vector of the $j$th target template.

### 3.1.3 Target Edge Feature Extraction

Although the gray feature, which is not sensitive to pedestrian translation, postural changes and partially occlusion, is an effective method for infrared target modeling, it has strong dependence on illumination and is easily affected by the background disturbances with its similar gray, which may bring about unsatisfactory results. Hence, edge feature is also utilized here to model the object structure.

For a template $t_i$, we design the edge direction histogram to describe its edge characteristics. Suppose the gray value of target is $I(x)$. $G(x)$ and $\alpha(x)$ denote the edge strength and direction. $\alpha(x) \in [0, 360^\circ]$ is used to define the edge direction histogram $q = \{q(v)\}_{v=1}^{L_E}$ of the target template:

$$q(v) = \sum_{i=1}^{M} \delta(b^*(x_i) - v), v = 1, 2, \cdots, L_E \tag{8}$$

where $b^*(x_i)$ denotes the edge direction mapping function. $L_E$ denotes the edge direction mapping level [30],.

Similar to the modified gray histogram, the edge direction histogram is improved by using the kernel function in Eq. (5), so that anti-noise performance of it can be improved. The modified edge direction histogram is described by:

$$q(v) = C_2 \sum_{i=1}^{M} k\left( \left\| \frac{x_0 - x_i}{h} \right\|^2 \right) \delta(b^*(x_i) - v), v = 1, \cdots, L_E \tag{9}$$

where $C_2$ is a normalization constant.

Consequently, we can extract the edge feature vector of each target template. Then, the edge feature vectors of all target templates are quantified respectively, and an edge feature set with spatial location information can be formed:

$$Q_E = [q_1, q_2, \cdots, q_N] \tag{10}$$

where $q_j \in R^{L_E \times 1}$ $(j = 1, 2, \cdots, N)$ denotes the edge feature vectors of the $j$th target template.

### 3.1.4 Target Combined Feature Generation

From the Eq.(7) and Eq.(10), we can find that, the $j$ th column vectors of these two matrices represent the gray feature and edge feature of a target model, respectively. Subsequently, the gray feature and edge feature are vertical connected, so that the gray feature and edge feature of the same target model can be represented in the one column vector. Ultimately, a combined feature set can be formed:

$$featset = [feature_1, \cdots, feature_N] = \begin{bmatrix} P_G \\ Q_E \end{bmatrix} = \begin{bmatrix} p_1, p_2, \cdots, p_N \\ q_1, q_2, \cdots, q_N \end{bmatrix} \tag{11}$$

where $featset \in R^{L \times N}$, $L=L_G + L_E$ denotes the combined feature set of all target templates, $feature_j \in R^{L \times 1}$ denotes the combined gray and edge feature vector of the $j$ th target template. It is worth pointing out that the combination scheme is simple but effective. In the one aspect, the computational load of simple arraying in the form of vertical rows is very light. In the other aspect, since after obtaining the fusion results, the following step is to use these results to learn a reconstructive dictionary. Ultimately, the learned dictionary can well represent the infrared pedestrian objects.

### 3.1.5 Target Dictionary Learning

In this paper, we utilize the simple and efficient K-singular value decomposition approach to learn the target dictionary.

The corresponding objective function is shown in Eq. (1). It uses the iterative approach to update the sparse coding and dictionary. When the dictionary $D$ is fixed, we use the OMP algorithm to calculate the sparse coefficients $X$ of the feature set $featset$ under the dictionary. When the sparse coefficient is fixed, we use the SVD method to update the dictionary $D$ by column. The processes are iterated until the number of iterations reaches the preset value. Finally, the learned reconstructive dictionary $D \in R^{L \times S}$ can be obtained.

Since the dictionary contains both of the gray feature and edge feature of the target templates, it can be used effectively to overcome the difficulties, such as posture changes, background noise, illumination and partial occlusion, when it models a tracked target.

### 3.2 Infrared Pedestrian Tracking

Based on the multi-feature local sparse appearance model learned above, we propose an infrared pedestrian tracker subsequently. The robust tracking of the target is to search the image regions which have the highest similarity in each frame. Sparse reconstruction residual analysis is applied to measuring the similarity.

(1) Sample a series of candidates for current frame.

- First, suppose that $R_{fr}$ denotes the target region located in the last frame at position $O^*$. $R_{se}$ is a region around the location, As shown in Fig. 3, red point denotes the $O^*$, red box indicates $R_{fr}$, and yellow box indicates $R_{se}$. It is worth pointing out that for the first frame, the target to be tracked is labeled manually and its location is recorded as $O^*$.
- Second, sample a number of candidates from the search neighborhood $R_{se}$. To handle the target scale variation problem, a multi-scale window scheme is applied in this process. The scales are set as $\beta \in [0.8,1.2]$, in steps of 0.1, of the previous target size.

- Finally, put $h$ candidate targets that are obtained by the multi-scale window scheme into $F$:

$$F = [f_1, f_2, \cdots, f_h] \tag{12}$$

where $f_g (1 \le g \le h)$ denotes the $g$ th candidate target.

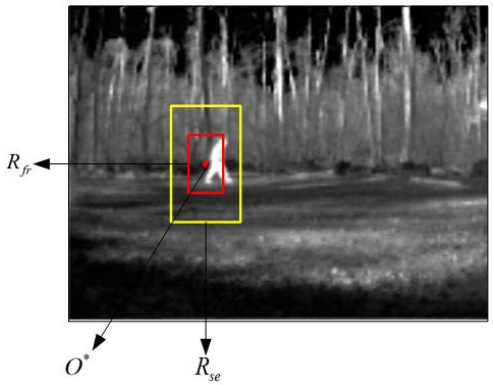



**Fig. 3.** Candidates search region.

(2) Extract gray and edge features of each candidate target.

- First, normalize the size of each candidate target to the same size $\mu \times \mu$, so as to get the unified feature dimension.
- Second, extract the gray feature vector and edge feature vector of each candidate target in the set $F$. Each of the two characteristic vectors is quantified and vertically connected to form $fea_g = \begin{bmatrix} p_g \\ q_g \end{bmatrix}$, where $1 \le g \le h$. $p_g$ and $q_g$ are gray and edge cues of the $g$ th candidate. $fea_g$ is the combined feature vector.

(3) Recognize the candidates using sparse reconstruction residual analysis.

- First, calculate sparse coding coefficients $X_g$ for $fea_g$ under dictionary $D$.
- Second, calculate the reconstruction error of each candidate target by:

$$\varepsilon_g = f_g - D_g X_g \tag{13}$$

where $\varepsilon_g$ denotes the reconstruction residual of the $g$ th candidate target.

- Finally, compare the $h$ reconstruction errors to screen out the minimum reconstruction error $\varepsilon_m$:

$$\varepsilon_m = \min[\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_h] \tag{14}$$

The candidate target corresponding to the minimum reconstruction error $\varepsilon_m$ is then identified as the tracked target in the current frame image.

### 3.3 Adaptive Dictionary Update

In most tracking situations, target to be tracked may not remain the same. It may undergo illumination or appearance changes during the tracking process. Therefore, it is essential to update the dictionary while tracking, which will help the tracker work steadily. In fact, for

target tracking, researchers have proposed various approaches with regard to target model update, most of which update the model by using the current frame's tracking result [31]. However, if the result is contaminated, the updated model will be inaccurate and some errors may be introduced in the tracking process. When the errors are accumulated to a certain extent, serious drifting may occur.

To prevent the drifting problem, an adaptive dictionary update approach is designed. If the current target is not dirtied, the current target feature set is updated with the tracking result of the current frame; otherwise, it is not updated.

(1) Calculate the gray and edge features of the tracked target in the current frame.

- First, extract the tracked target for the current frame.
- Second, calculate the gray feature vector and edge feature vector of the tracked target, which are denoted by $p_{cur}=[p_{cur}(u)]_{u=1}^{L_G}$ and $q_{cur}=[q_{cur}(v)]_{v=1}^{L_E}$, respectively

(2) Judge whether the current target is dirtied.

- First, in Sections 3.1.2 and 3.1.3, the gray feature set $P_G=[p_1,p_2,\cdots,p_N]\in R^{L_G\times N}$ and the edge feature set $Q_E=[q_1,q_2,\cdots,q_N]\in R^{L_E\times N}$ of $N$ target templates have been obtained, where $p_j=[p_j(u)]_{u=1}^{L_G}$ $(j=1,2,\cdots,N)$ denotes the gray feature vector of $j$th target template, and $q_j=[q_j(v)]_{v=1}^{L_E}$ $(j=1,2,\cdots,N)$ denotes the edge feature vector of $j$th target template.
- Second, compute the Bhattacharyya coefficients [6] $\rho_{gray,j}$ between $p_{cur}$ and $p_j$, and the Bhattacharyya coefficients $\rho_{edge,j}$ between $q_{cur}$ and $q_j$:

$$\rho_{gray,j}=\rho_{gray,j}[p_{cur},p_j]=\sum_{u=1}^{L_G}\sqrt{p_{cur}(u)p_j(u)} \tag{15}$$

$$\rho_{edge,j}=\rho_{edge,j}[q_{cur},q_j]=\sum_{v=1}^{L_E}\sqrt{q_{cur}(v)q_j(v)} \tag{16}$$

where $j=1,2,\cdots,N$. Note that the Bhattacharyya coefficients $\rho_{gray,j}$ is related to the gray feature, while the Bhattacharyya coefficients $\rho_{edge,j}$ is related to the edge feature.

- Third, in Eq. (15) and Eq. (16), the larger $\rho_{gray,j}$ or $\rho_{edge,j}$ is, the more likely the current target is to be the $j$th target template. However, in different scenes, the discrimination ability of gray features and edge features may be different. So we propose to combine them together to determine their similarity:

$$\rho_{sum,j}=w_{gray,j}\rho_{gray,j}+w_{edge,j}\rho_{edge,j} \tag{17}$$

where $\rho_{sum,j}$ is the fused Bhattacharyya coefficient. $w_{gray,j}$ and $w_{edge,j}$ are the weights of $\rho_{gray,j}$ and $\rho_{edge,j}$, respectively:

$$w_{gray,j}=\frac{\rho_{gray,j}}{\rho_{gray,j}+\rho_{edge,j}} \tag{18}$$

$$w_{edge,j} = \frac{\rho_{edge,j}}{\rho_{gray,j} + \rho_{edge,j}} \tag{19}$$

After such processing, we can draw a more precise conclusion that the higher $\rho_{sum,j}$ is, the more likely the current target is to be the $j$ th target template.

- Fourth, according to the above steps, we can obtain $N$ Bhattacharyya coefficients:

$$Sim = [\rho_{sum,1}, \cdots, \rho_{sum,N}] \tag{20}$$

- Finally, seek the maximum value $\rho_{sum,ma} = \max(\rho_{sum,1}, \cdots, \rho_{sum,N})$ from $Sim$. Compare $\rho_{sum,ma}$ with a preset threshold $th \in [0,1]$. If $\rho_{sum,ma} < th$, it means that the current target is not similar to any template. In this case, the current result is thought to be dirtied and we do not use it for update.

(3) If the current result is not dirtied, the dictionary is updated.

- First, seek the minimum value $\rho_{sum,mi} = \min(\rho_{sum,1}, \cdots, \rho_{sum,N})$ from $Sim$.
- Second, replace the gray feature vector $p_{mi}$ and edge feature vector $q_{mi}$ of the $mi$ th target template by $p_{cur}$ and $q_{cur}$ to get the updated feature set.
- Finally, update the dictionary every $\gamma$ frames with the updated feature set.

## 4. Experimental Results

### 4.1 Experimental Setup

Experiments are done by MATLAB R2013b on an Intel Dual Core 2.3 GHz laptop with 4 GB RAM. The proposed multi-feature local sparse representation algorithm is tested and also compared with several classical tracking methods. The test infrared pedestrian sequences are gotten from the public OTCBVS database [32]. The size of each image is 120×160. This paper illustrates the experimental results of four representative infrared pedestrian video sequences that have various challenging factors in video tracking, including illumination change, occlusion, background disturbance and posture change. The specific information of the four image sequences is shown in **Table 1**.

**Table 1.** Infrared sequences information.

| Image sequences | Number of frames | Object size |
|---|---|---|
| Q1 | 200 | 17×9 |
| Q2 | 160 | 15×8 |
| Q3 | 180 | 18×9 |
| Q4 | 180 | 24×12 |

In our experiments, besides the qualitative evaluation, we also make the quantitative evaluation by using two criteria: tracking success rate as well as center location error [33, 34].

First, center location error for a frame $i$ is used to measure the distance between the centers of the ground truth and tracking result (i.e., $O_G$ and $O$):

$$CLE_i = d_i(O, O_G) \tag{21}$$

where $d_i(O, O_G)$ denotes the Euclidean distance between $O$ and $O_G$. For a whole image sequence, the center position error is calculated by:

$$CLE = \frac{1}{U}\sum_{i=1}^{U} CLE_i \tag{22}$$

where $U$ is the total number of frames in a video. From Eq. (22), we can see that a lower CLE means a higher tracking accuracy.

Second, the tracking success rate, which describes the percentage of frames precisely tracked in a sequence, is defined as:

$$TSR = \frac{U_{su}}{U} \tag{23}$$

where $U_{su}$ denotes the number of frames that are processed successfully. A larger *TSR* means a better performance. The following measure is utilized to judge whether the tracking is successful in a frame:

$$\frac{\Omega_T \cap \Omega_G}{\Omega_T \cup \Omega_G} \geq \eta \tag{24}$$

where $\Omega_T$ is the tracked box and $\Omega_G$ is the ground truth box. If Eq. (24) is met, the object is thought to be tracked successfully. $\eta$ is a threshold controlling the tracking success rate.

### 4.2 Evaluation of Target Combined Feature Generation

The effectiveness of feature generation module is validated at first. The experimental sequence is Sequence Q3. **Fig. 4** gives the tracking results of frames 11, 48, 61, 87 and 167 by the proposed method with gray, edge, and combined features, respectivley. As can be seen, the combined feature can enhance the robustness compared to gray or edge feature.

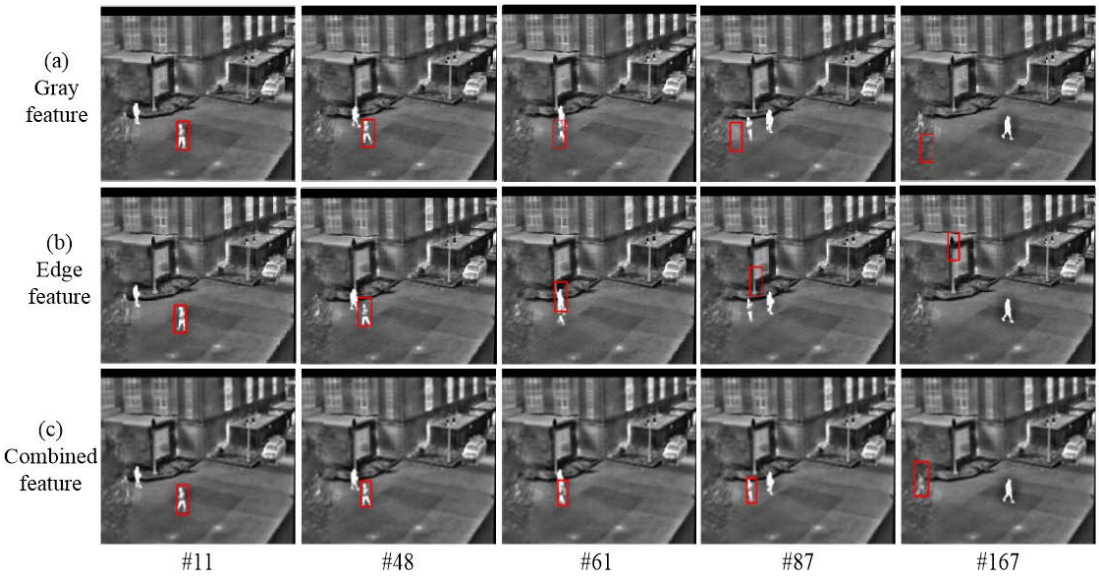



**Fig. 4.** Evaluation of combined feature generation.

### 4.3 Evaluation of Adaptive Dictionary Update

The effectiveness of adaptive dictionary update module is evaluated subsequently. Results with and without using the update step are compared, as shown in **Fig. 5**. As can be seen, our technique produces better tracking results, while drifting occurs when the method is without the dictionary update step.

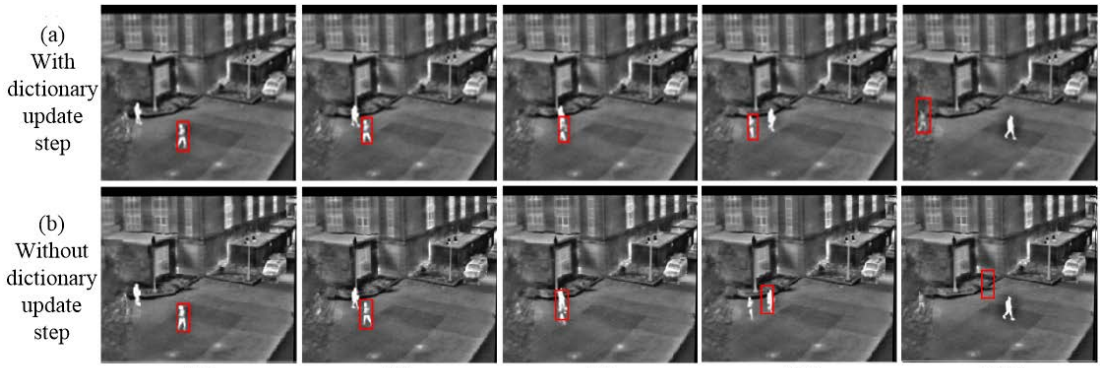



**Fig. 5.** Evaluation of adaptive dictioanry update scheme.

### 4.4 Qualitative Evaluation

Our approach is compared with two classical tracking algorithms in this section. The first one is the sparse representation based method [15]. We refer to it as SRT. The second one is the classical particle filter based tracking algorithm [2]. We refer to it as PFT. Both of these two comparing methods are based on gray characteristic for infrared target tracking.

**Fig. 6** shows some tracking results of Q1, in which a pedestrian target is walking outdoors on campus. As the target moves on, it undergoes illumination variation. From **Fig. 6** (c), we find that PFT drifts from the 27th frame, and there is no correction in the following tracking process. SRT tracks the target a little better as shown in **Fig. 6** (b). But there is a sign of drifting starting from the 70th frame, and then the tracking fails as the errors increase. On the contrary, the proposed algorithm overcomes the influence of the illumination variation and presents satisfactory tracking performance for infrared pedestrian target (**Fig. 6** (a)).

Q2's results are given in **Fig. 7**. In this sequence, an infrared pedestrian target is walking under a forest background. During tracking, it is occluded by tree. From **Fig. 7** (a), it can be seen that our algorithm recovers from occlusion. However, SRT and PFT fail in the tracking process (**Fig. 7** (b) and **Fig. 7** (c)).

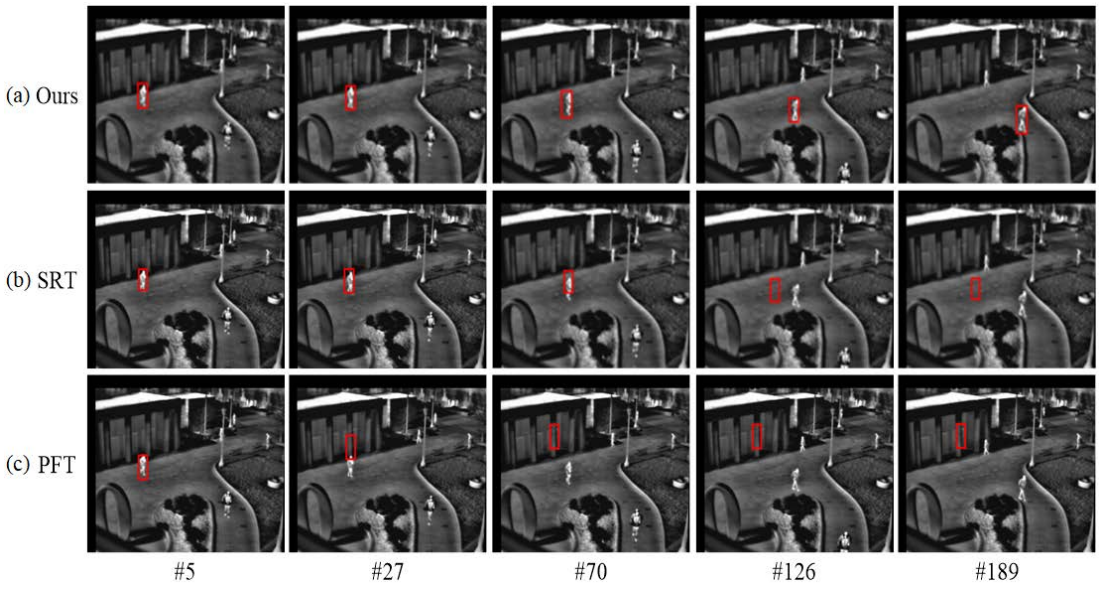



**Fig. 6.** Comparison results of Q1 by different algorithms.

**Fig. 8** gives the results of Q3, where the pedestrian is moving from middle to left. The target is first disturbed by another pedestrian with similar gray and then occluded by tree. Under such circumstances, both of SRT and PFT can hardly handle the problems, and then lose the target (**Fig. 8** (b) and **Fig. 8** (c)), while our method tracks the target successfully (**Fig. 8** (a)) for it fuses gray and edge features into the local sparse representation framework.

In sequence Q4, the target is moving from right to left under a dense forest background. As shown in **Fig. 9**, the target of interest meets a person who comes from the left side, and then they separate. This adds difficulties to the target tracking. Moreover, the target also undergoes the size and posture variations during tracking. Our technique tracks the object accurately throughout the video (**Fig. 9** (a)). On the other hand, the traditional SRT and PFT methods lose the target completely (**Fig. 9** (b) and **Fig. 9** (c)).

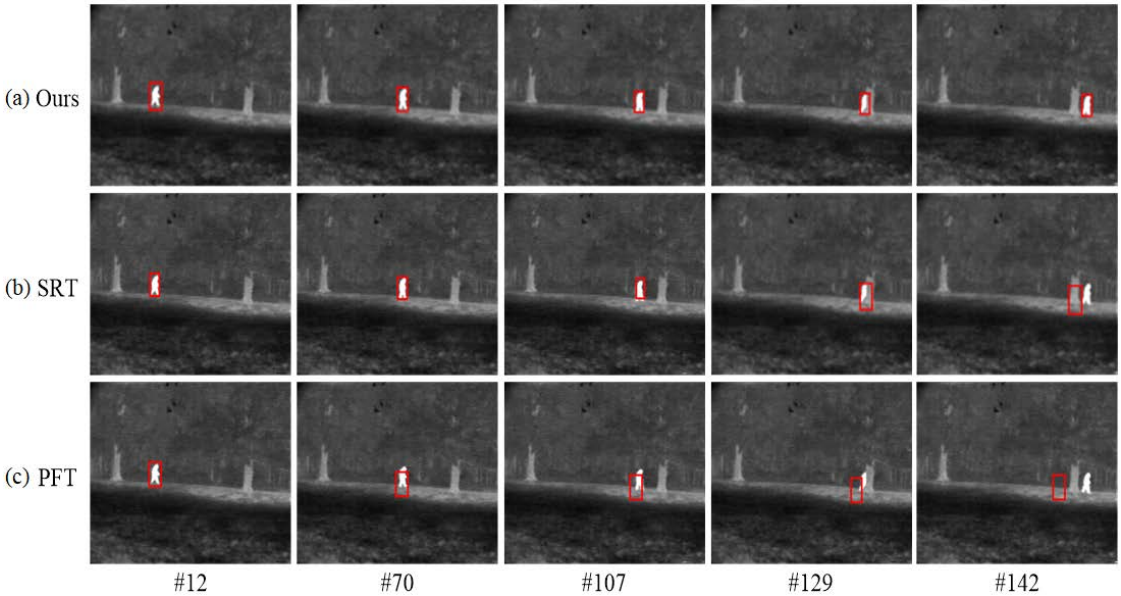



**Fig. 7.** Comparison results of Q2 by different algorithms.

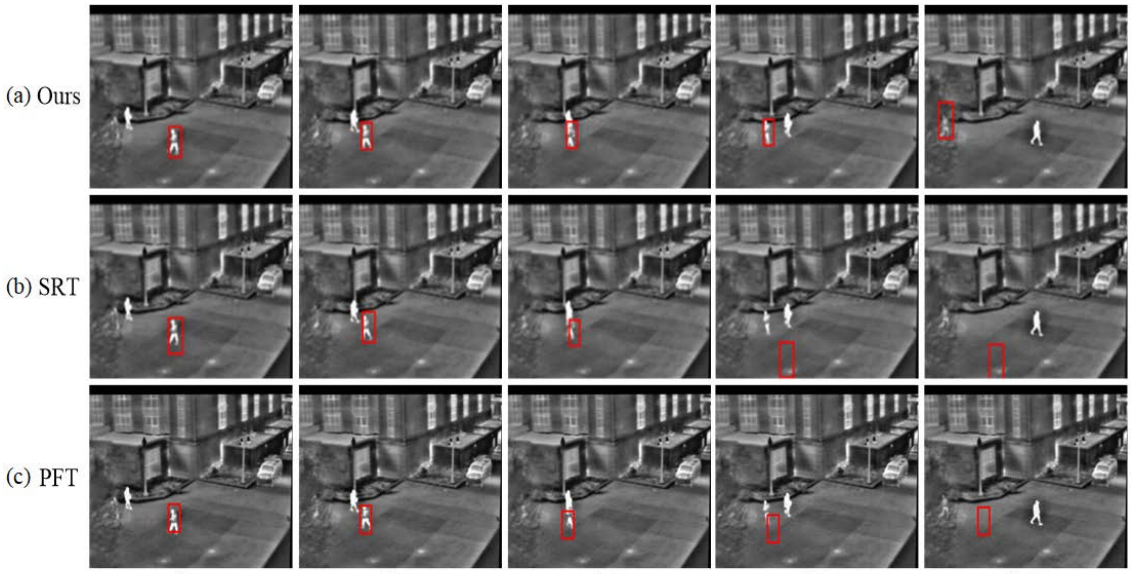



**Fig. 8.** Comparison results of Q3 by different algorithms.

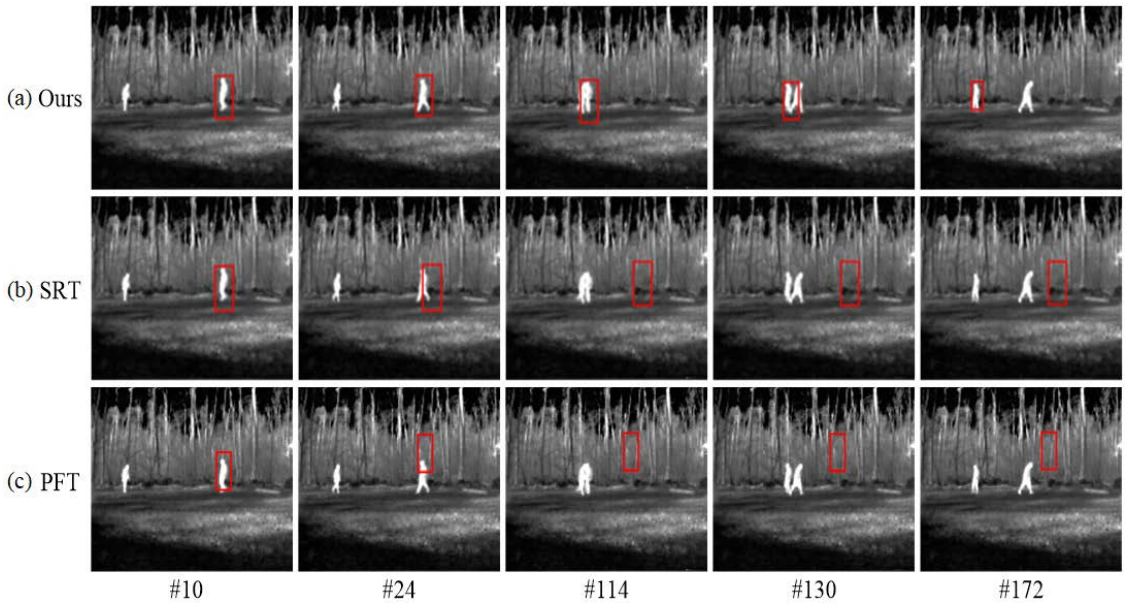



**Fig. 9.** Comparison results of Q4 by different algorithms.

### 4.5 Quantitative Evaluation

Quantitative results are also compared for the above four sequences in **Table 2** and **Table 3**.

First, from **Table 2**, we can see that the CLE values of our algorithm are lower than those of the other algorithms, which indicates that our method has higher tracking accuracy.

Second, from **Table 3**, it can be seen that the TSR values of our algorithm are much higher than those of the other algorithms, which further reveals that our algorithm has a better tracking performance.

**Table 2.** Center location errors (pixles) of three tracking methods for four different image sequences.

| Image sequences | Ours | SRT | PFT |
|---|---|---|---|
| S1 | 8 | 26 | 51 |
| S2 | 6 | 13 | 21 |
| S3 | 11 | 39 | 32 |
| S4 | 12 | 47 | 55 |

**Table 3.** Tracking Success Rates (%) of three tracking methods for four different image sequences.

| Image sequences | Ours | SRT | PFT |
|---|---|---|---|
| S1 | 91.2 | 38.2 | 11.7 |
| S2 | 92.6 | 85.6 | 68.7 |
| S3 | 90.1 | 33.4 | 33.8 |
| S4 | 89.4 | 13.8 | 12.7 |

## 5. Conclusion

A multi-feature local sparse representation scheme is proposed for infrared pedestrian tracking problems. Frist, we extract the gray and the edge features for the tracked infrared pedestrian target and fuse them together to learn an effective multi-feature local sparse appearance model, which is well used for describing the characteristics of the tracked target. Then, based on the learned model, a robust tracker with an adaptive dictionary learning technique is presented to track the object over time. The results show that our algorithm works well for infrared pedestrian target tracking problems. Future area for research includes the investigation of alternative features and tracking multiple infrared targets.

## References

[1] Masahiro Yasuno, Noboru Yasuda and Masayoshi Aoki, "Pedestrian Detection and Tracking in Far Infrared Images," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshop*, Washington, pp. 125, June 27-31, 2004. Article (CrossRef Link).

[2] Xin Wang and Zhenmin Tang, "Modified particle filter-based infrared pedestrian tracking", *Infrared Physics & Technology*, vol. 53, no. 4, pp. 280-287, July, 2010. Article (CrossRef Link).

[3] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon and Tim Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174-188, February, 2002. Article (CrossRef Link).

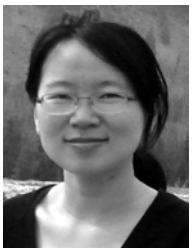
[4] Katja Nummiaro, Esther Koller-Meier and Luc Van Gool, "An adaptive color-based particle filter," *Image and Vision Computing*, vol. 21, no. 1, pp. 99-110, January, 2003. Article (CrossRef Link).

[5] Jiangtao Wang, Debao Chen, Haiyan Chen and Jingyu Yang, "On pedestrian detection and tracking in infrared videos", *Pattern Recognition Letters*, vol. 33, no. 6, pp. 775-785, April, 2012. Article (CrossRef Link).

[6] Xin Wang, Lei Liu and Zhenmin Tang, "Infrared human tracking with improved Mean Shift algorithm based on multi-cue fusion", *Applied Optics*, vol. 48, no. 21, pp. 4201-4212, July, 2009. Article (CrossRef Link).

[7] Dorin Comaniciu, Visvanathan Ramesh and Peter Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 142-149, June 15, 2000. Article (CrossRef Link).
[8] Changjiang Yang, R. Duraiswami and L. Davis, "Efficient mean-shift tracking via a new similarity measure," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 176-183, June 20-25, 2005. Article (CrossRef Link).
[9] Fabrizio Lamberti, Andrea Sanna and Gianluca Paravati, "Improving robustness of infrared target tracking algorithms based on template matching", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 1467-1480, April, 2011. Article (CrossRef Link).
[10] Suk Jin Lee, Gaurav Shah, Arka Aloke Bhattacharya and Yuichi Motai, "Human tracking with an infrared camera using a curve matching framework", *EURASIP Journal on Advances in Signal Processing*, vol. 2012, pp. 99, May, 2012. Article (CrossRef Link).
[11] Xin Wang, Chen Ning and Lizhong Xu, "Spatiotemporal Difference-of-Gaussians filters for robust infrared small target tracking in various complex scenes," *Applied Optics*, vol. 54, no. 7, pp. 1573-1586, July, 2015. Article (CrossRef Link).
[12] Xianguo Yu, Qifeng Yu, Yang Shang and Hongliang Zhang, "Dense structural learning for infrared object tracking at 200+ Frames per Second," *Pattern Recognition Letters*, vol. 100, pp. 152-159, December, 2017. Article (CrossRef Link).
[13] C. S. Asha and A. V. Narasimhadhan, "Robust infrared target tracking using discriminative and generative approaches," *Infrared Physics & Technology*, vol. 85, pp. 114-127, June, 2017. Article (CrossRef Link).
[14] Xue Mei and Haibin Ling, "Robust visual tracking using l1 minimization," in *Proc. Of IEEE Conference on Computer Vision*, pp. 1436-1443, 2009. Article (CrossRef Link).
[15] Xu Jia, Huchuan Lu and MingHsuan Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proc. Of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1822-1829, June 16-21, 2012. Article (CrossRef Link).
[16] John Wright, Yi Ma, Julien Mairal, Guillermo Sapiro, Thomas S. Huang and Shuicheng Yan, "Sparse Representation for Computer Vision and Pattern Recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031-1044, June, 2010. Article (CrossRef Link).
[17] Guang Han, Xingyue Wang, Jixin Liu, Ning Sun and Cailing Wang, "Robust object tracking based on local region sparse appearance model," *Neurocomputing*, vol. 184, pp. 145-167, April, 2016. Article (CrossRef Link).
[18] Bohan Zhuang, Huchuan Lu, Ziyang Xiao and Dong Wang, "Visual tracking via discriminative sparse similarity map," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1872-1881, April, 2014. Article (CrossRef Link).
[19] Xue Mei and Haibin Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no.11, pp. 2259-2272, April, 2011. Article (CrossRef Link).
[20] Michael Elad and Michal Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736-3745, November, 2006. Article (CrossRef Link).
[21] Xin Wang, Siqiu Shen, Chen Ning, Mengxi Xu and Xijun Yan, "A sparse representation-based method for infrared dim target detection under sea-sky background," *Infrared Physics & Technologu*, vol. 71, pp. 347-355, July, 2015. Article (CrossRef Link).
[22] Tanaya Guha and Rabab K Ward, "Learning sparse representations for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 8, pp. 1576-1588, August, 2012. Article (CrossRef Link).
[23] Xin Wang, Siqiu Shen, Chen Ning, Fengchen Huang and Hongmin Gao, "Multi-class remote sensing object recognition based on discriminative sparse representation," *Applied Optics*, vol. 55, no. 6, pp. 1381-1394, 2016. Article (CrossRef Link).
[24] Jian Zhang, Debin Zhao and Wen Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336-3351, May, 2014. Article (CrossRef Link).

[25] Julien Mairal, Francis Bach and Jean Ponce, "Task-driven dictionary learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4): 791-804, 2012. Article (CrossRef Link).
[26] Michal Aharon, Michael Elad and Alfred Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311-4322, November, 2006. Article (CrossRef Link).
[27] StCphane G. Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397-3415, 1993. Article (CrossRef Link).
[28] Joel A. Tropp and Anna C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655-4666, December, 2007. Article (CrossRef Link).
[29] Baiyang Liu, Junzhou Huang, Casimir Kulikowski and Lin Yang, "Robust visual tracking using local sparse appearance model and K-selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2968-2981, December, 2013. Article (CrossRef Link).
[30] Paul Brasnett, Lyudmila Mihaylova, David Bull and Nishan Canagarajah, "Sequential Monte Carlo tracking by fusing multiple cues in video sequences," *Image and Vision Computing*, vol. 25, no. 8, pp. 1217-1227, August, 2007. Article (CrossRef Link).
[31] David A. Ross, Jongwoo Lim, RueiSung Lin and MingHsuan Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125-141, May, 2008. Article (CrossRef Link).
[32] J. Davis and M. Keck, "A two-stage approach to person detection in thermal imagery, " in *Proc.of Workshop on Applications of Computer Vision*, January, pp. 364-369, 2005. Article (CrossRef Link).
[33] Dilip K. Prasad and Michael S. Brown, "Online tracking of deformable objects under occlusion using dominant points," *Journal of the Optical Society of America A*, vol. 30, no. 8, pp. 1484-1491, 2013. Article (CrossRef Link).
[34] Xin Wang, Siqiu Shen, Chen Ning, Yuzhen Zhang and Guofang Lv, "Robust object tracking via local discriminative sparse representation," *Journal of the Optical Society of America A*, vol. 34, no. 4, pp. 533-544, 2017. Article (CrossRef Link).



**Xin Wang** received the Ph.D. degree in Computer Application Technology from Nanjing University of Science and Technology, Nanjing, China, in 2010. She is currently an Associate Professor with the College of Computer and Information, Hohai University, Nanjing, China. She has published more than 50 papers in journals and referred conferences. Her current research interests include image processing, computer vision, and pattern recognition.

**Lingling Xu** received the B.S. degree in Communication Engineering from Nanjing University of Science and Technology Zijin College, Nanjing, China, in 2015. Now she is working toward the M.S. degree in the College of Computer and Information, Hohai University. Her current research interests include image processing, target detection and tracking.

**Chen Ning** received the M.S. degree in Signal and Information Processing from University of Science and Technology of China, Hefei, China, in 2003. He is currently working in the School of Physics and Technology, Nanjing Normal University, Nanjing, China. His current research interests include signal and image processing, and computer vision.