IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 15, 2021, accepted May 31, 2021, date of publication June 7, 2021, date of current version June 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3087238

Evolutionary Rear-Lamp Tracking at Nighttime

HAITIAN SUN[1], TAKUMI NAKANE[2], NAIDAN ZHANG[1], AND CHAO ZHANG[2], (Member, IEEE)

[1]Department of Computer Science, Northwest A&F University, Yangling 712100, China
[2]Faculty of Engineering, University of Fukui, Fukui 910-8507, Japan

Corresponding author: Chao Zhang (zhang@u-fukui.ac.jp)

ABSTRACT Rear-lamp tracking at nighttime plays a momentous role in the advanced driver assistance system (ADAS), involving collision mitigation, automatic cruise control, automatic headlamp dimming, etc. Most of the existing tracking methods based on monocular camera leverage on color features. However, such tracking methods can be easily influenced by background clutter, illumination change, distance variation, and occlusion. In this paper, we propose an evolutionary adaptive rear-lamp tracking method at nighttime, in which a novel genetic algorithm powered by the probabilistic bitwise operation (PBO) is utilized. Also, to improve the robustness against various environments, a balanced fitness function is proposed by taking color information, symmetry, spatial relationship, and rigidity into account. Especially, a series of adaptive thresholds based on rear data in HSV color space is proposed to exploit color information reasonably with respect to our task. A strategy to deal with occlusion is also proposed, which relies on color information and rigidity. Moreover, to our knowledge, there is no publicly available dataset for rear-lamp tracking at nighttime. To fill the gap between the real-world application and the theoretical research, we create a novel dataset, which contains diverse traffic conditions at nighttime. The experimental results indicate that our method outperforms comparative online tracking methods in terms of success rate and center location error.

INDEX TERMS Nighttime rear-lamp tracking, evolutionary algorithm, probabilistic bitwise operation.

I. INTRODUCTION

More than half of all traffic deaths occur after dark and a large percentage of road injuries are attributed to lack of clarity of vision at nighttime [1], [2]. In this paper, we propose to automatically track the rear-lamps of the preceding vehicle at nighttime, in order to make contributions to the advanced driving assistance system (ADAS). For example, our proposed rear-lamp tracking method can be useful for the traffic status understanding (e.g., the status of the brake light or the turn signal light), collision mitigation, automatic cruise control, automatic headlamp dimming, etc. On the other hand, although the LIDAR sensor can obtain 3D information which can possibly increase the success rate of rear-lamp tracking, despite its high cost, it also has range limitations. In this paper, we aim to design a rear-lamp tracking system that depends on a single monocular camera, which can be a part of the driving recorder.

Previous studies have given less consideration on rear-lamp tracking at nighttime. Most existing researches on vehicle tracking or detection have been focused on the daytime environment (e.g., [3]). However, tracking or detection at nighttime can be more challenging because of the low-light environment and complex light sources. E.g., under the low-light environment, the contour of a vehicle can hardly be perceived. Instead of detecting the vehicles directly, rear-lamp detection can be indirectly applied for vehicle detection at nighttime [4]. Following the same idea, vehicle tracking can also be handled by rear-lamp tracking. On the other hand, owing to the lack of visual features, existing online tracking methods, such as [5]–[8], cannot deal with the problem well as shown in our experimental result. Fig. 1 shows some examples of tracking results from an image sequence. As the sequence progresses, the scale of rear-lamp regions changes and a variety of light sources appear to include street lamps, traffic lights, the headlight of opposite vehicles, reflection from the ground, etc., which makes the tracking task challenging.

Rear-lamp tracking can be realized in a tracking-by-detection fashion. Most of the existing detection methods consist of two steps: 1) candidate regions sampling; 2) optimization based on candidate regions. In the first step, since the rear-lamps emit red light, features are considered to be effective for sampling candidate regions, such as RGB [9]–[11], HSV [12], [13], and YC_bC_r [14]. The second step is to distinguish the rear-lamps from candidate regions

The associate editor coordinating the review of this manuscript and approving it for publication was Michail Makridis.

FIGURE 1. Examples of rear-lamp tracking under mixed light sources and scale variation at nighttime. (a)∼(d) are different frames from the same sequence, with the frame number shown at the top left. The red bounding boxes illustrate our tracking results. Cyan, magenta, yellow, blue bounding boxes illustrate the tracking results of [5]–[8] respectively.

in order to alleviate the drift problem of tracking caused by other light sources. Existing methods propose to detect the symmetry of two rear-lamps by normalized cross-correlation [12] and color information [10]. However, such methods rely on fixed thresholds, which could cause results to change with the environment.

There also exist some rear-lamp datasets at nighttime for the detection purpose (e.g., Sun Yatsen University Night-time Vehicle Dataset [15]), with the limited number of annotations based on independent images. In this paper, a nighttime rear-lamp tracking dataset is created. It contains 50 annotated sequences (19558 frames) with different types of vehicles, traffic situations, and environments (e.g., illumination variation, scale variation, occlusion, and background clutters).

To tackle the problems described above, in this paper, a robust evolutionary rear-lamp tracking method at nighttime is proposed. The main framework is powered by a variant of genetic algorithm (GA) embedded with probabilistic bit-wise operation (PBO) [16], which replaces the traditional genetic operations such as crossover and mutation, to increase the ability to prevent the algorithm from falling into local optimum. In order to adopt PBO to our problem, we propose a balanced fitness function by considering color information, symmetry, consistency of tracking results between adjacent frames, and rigidity. Here, color information is mainly extracted based on a series of adaptive thresholds in HSV color space and location relationships. Also, a strategy for dealing with occlusion is proposed. In conclusion, the main contributions of this paper can be summarized as follows:

- A novel evolutionary algorithm with a balanced fitness function is applied for rear-lamp tracking.
- A series of adaptive color thresholds and a strategy for dealing with occlusion are proposed for further improvement.
- A large-scale rear-lamp tracking dataset at nighttime is originally built with annotations.

II. RELATED WORKS

A. REAR-LAMP DETECTION AT NIGHTTIME

As the claim given by Schamm *et al.* [4] that vehicle detection at nighttime can be through rear-lamp detection, rear-lamp detection at nighttime is important, well studied, and can be the fundamental of rear-lamp tracking. The research field is usually divided into two steps: 1) candidate regions sampling; 2) candidate regions optimization, as is mentioned in Sect. I. We review related works for rear-lamp detection at nighttime from the aspects of the two steps.

In the first step, because rear-lamps are red, which is an important property, most methods employed various color spaces to solve the problem. Chen *et al.* [9] proposed to exploit the contrast between the rear-lamp and the background so that the candidate regions can be extracted and sampled. The contrast was emphasized by transfer RGB images to single-channel images. O'Malley *et al.* [12] decided thresholds in HSV color space, which were determined according to worldwide regulations and real-world conditions, to sample candidate regions. Casares *et al.* [10] proposed to sample the candidate regions by locating the region that is red or red surrounded by white. Jeong *et al.* [11] proposed a tone-mapping process to locate rear-lamps. Nakane *et al.* [13] combined PBO with the thresholds in [12] and the strategy in [10]. The method of [13] is our previous work, which aims at detecting the rear-lamps from a single image, rather than videos.

The second step is to cluster the candidate regions into the positive (rear-lamp) or the negative (other luminous sources). Firstly, the shape of the rear-lamp depends on the design, thus it cannot be handled by geometric rules. Some researchers concentrated on the property of symmetry, estimated by normalized cross-correlation [12] and color information [10]. Secondly, the part-based models that measure the spatial relationships are also effective. E.g., rear-lamps and license plates have determined location relationships, which can be detected by Markov [17] random field or Gaussian mixture model [18]. Thirdly, this step can also be tackled with by deep learning [19].

Apart from the special researches, other object detection methods are supposed to be applied to rear-lamp detection. 1) Feature-based detection. The histogram of oriented gradient (HOG) [20] was good at measuring the whole feature and could be combined with supported vector machine (SVM) [21], which was applied to vehicle detection [22]. Scale-invariant feature transform (SIFT) [23] and speeded-up robust feature (SURF) [24] measured local point features. 2) Template matching. As representative methods, GA [25], particle swarm optimization (PSO) [26], and deterministic crowding (DC) [27] based template matching was able to locate objects. Best-buddies similarity (BBS) [28] and deformable diversity similarity (DDIS) [29] had the ability

to deal with deformation between the template and target. 3) Machine learning. Once the training dataset is large enough, convolutional neural networks (CNN) based object detection, such as the one-stage you look only once (YOLO) [30] and single shot multibox detector (SSD) [31], as well as two-stage Faster R-CNN [32], performed much better than other methods, including other machine learning methods. But they also usually face the challenging of training dataset creation.

B. REAR-LAMP TRACKING

As an object tracking problem, rear-lamp tracking can be solved by online tracking methods. Sevilla *et al.* [7] proposed distribution fields for tracking (DFT), which built an image descriptor using distribution fields. Zhang *et al.* [6] proposed compressive tracking (CT), which compressed the samples by a sparse matrix and employed naive Bayes classifier to classify haar-like features. Henriques *et al.* [5] proposed circulant structure of tracking-by-detection with kernels (CSK). Oron *et al.* [8] proposed locally orderless tracking (LOT), which automatically estimated the amount of local (dis)order in the target. Nevertheless, the online tracking methods are not the special methods for rear-lamp tracking and are difficult to deal with the problem in real traffic environments.

Existing special researches on rear-lamp tracking at nighttime are based on the detection. O'Malley [12] extended their detection method to tracking. Li *et al.* [33] proposed to track rear-lamps by vehicle detection and symmetry analysis, but their method was not conducted experiments at nighttime. Vancea *et al.* [19] proposed to sample candidates by thresholding and pair the candidates by deep learning, then extended the detection to tracking. However, existing rear-lamp tracking methods are experimented with simple backgrounds and without real traffic environments.

In this paper, we propose a novel robust evolutionary rear-lamp tracking method at nighttime, which is based on refined GA with a balanced fitness function. The function also considers color information, but the thresholds are adaptive according to the environment. And a strategy to cope with occlusion is also proposed. The proposed method is evaluated in diverse real complex conditions.

III. METHODOLOGY

A. OVERVIEW

Given an image sequence of a driving record at nighttime, the locations of two rear-lamps in the first frame are given by the user (denoted by r_1), which follows the problem setting of online tracking [5]–[8]. The objective of tracking is to automatically locate the rear-lamps from the second frame, with the locations of two lamps denoted by $r_{2,\ldots,n}$. The tracking result of t-th frame r_t is defined as a set consists of the left bounding box r_t^l and the right bounding box r_t^r. With located r_t, in $(t+1)$-th frame, a region of interest (ROI) is located according to t-th frame as shown in Fig. 2. The ROI is represented by a green bounding box, in which there are randomly generated candidate boxes at the initialization step

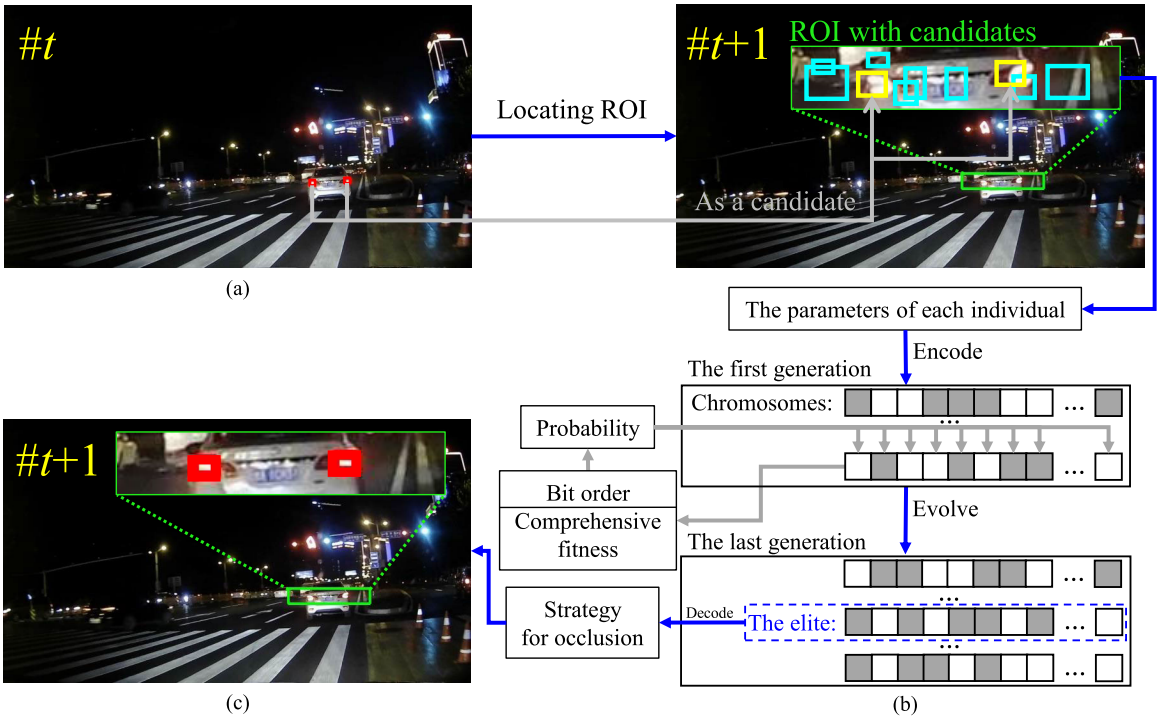


FIGURE 2. The overview of our method. (a) t-th frame in the sequence with the tracking result represented by two red bounding boxes. (b) The optimization procedure of frame $t+1$ by PBO. The cyan boxes are the randomly generated initial candidates, which form the first generation in PBO. (c) The tracking result of $(t+1)$-th frame.

of PBO. Especially, the tracking result of t-th frame is also inherited as a candidate to ensure the consistency of tracking results. Each individual in PBO consists of parameters to define a boxed pair, and parameters are encoded into binary chromosomes. During the generation iteration of PBO, genes are flipped according to a probability calculated from the bit order and the fitness value. To estimate r_{t+1}, the elite of the last generation is decoded back to parameters in real number and then passed through the strategy for dealing with occlusion at last. We will then introduce each step in detail.

B. LOCATING ROI

Given the $(t+1)$-th frame and the tracking result of the last frame r_t, r_{t+1} is considered to be near r_t because rear-lamps should not be displaced significantly in two adjacent frames. In our algorithm, the ROI of $(t+1)$-th frame, which potentially contains the rear-lamps, is defined by

$$\mathrm{ROI}_{t+1} = \lambda r_t^m, \tag{1}$$

where r_t^m is the minimum bounding box that can contain both r_t^l and r_t^r. λ is a scaling factor which is set to 2 throughout this paper. That is, the ROI of frame $t+1$ is determined by enlarging the neighborhood region of r_t.

C. PBO BASED GA

1) SIMPLE GA

Simple GA [34] encodes parameters into chromosomes and explores the optimum solution by iterating the genetic operations, including selection, crossover, and mutation. Specifically, GA begins with a potential solution space, named population, which consists of a number of individuals. Each individual corresponds to a chromosome, while each chromosome is made up of a series of genes. Usually, the chromosome is encoded into binary codes, in order to simply simulate heredity and evolution. The population evolves for a predetermined number of generations. The evolution of each generation involves the computation of fitness and the operation of selection, crossover, and mutation. Meanwhile, the elite of each generation is saved and inherited from the next generation. The decoded parameters from the elite in the last generation are the optimum solution given by GA.

2) PBO

GA could fall into local optimums easily, owing to the insufficient ability of diversity preserving when the solution space is large [35]. Hence, preserving genetic diversity during the evolution process is a crucial factor for effective solution search [36]. We adopt PBO [16], which is an individual-independent bit flip operation calculated according to the fitness and bit order, to replace crossover and mutation. PBO flips each bit in the chromosome with the probability

$$P = \omega_{max} \exp(-\frac{1}{2}(\frac{b^2}{s_b^2} + \frac{f^2}{s_f^2})), \tag{2}$$

where b denotes the normalized bit order (the maximal bit order is defined as the left-most bit) in one parameter, f denotes the fitness value, s_b and s_f are the parameters for controlling the smoothness of distribution, and ω_{max} denotes the max value of the distribution. The description of PBO is shown in Fig. 3.

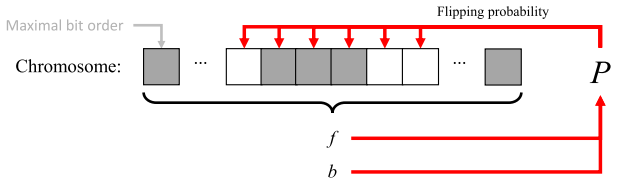


FIGURE 3. Illustration of PBO. PBO flips each binary bit by probability P. P is calculated according to bit order b and fitness f, where the maximal b is defined as the left-most bit.

PBO also employs annealing selection with the rate of P_a for controlling the convergence, which represents the probability of an individual to be selected. For the i-th generation, P_a is calculated by

$$P_a = \frac{1.0 - \exp(\frac{m}{G}i)}{\exp(m) - 1.0}(1.0 - P_{min}) + 1.0, \tag{3}$$

where m is the parameter to control the smoothness of the distribution, G denotes the predetermined number of generations, and P_{min} is a constant.

3) OPTIMIZATION

Each chromosome in our method is encoded with six parameters as shown in Fig. 4. Both rear-lamps are supposed to be the same size. Minimal width w_{min} and height h_{min} are defined. Hence, $w = \alpha w_{min}$, $h = \beta h_{min}$. In this way, in stead of w and h, only integers α and β need to be encoded and optimized.

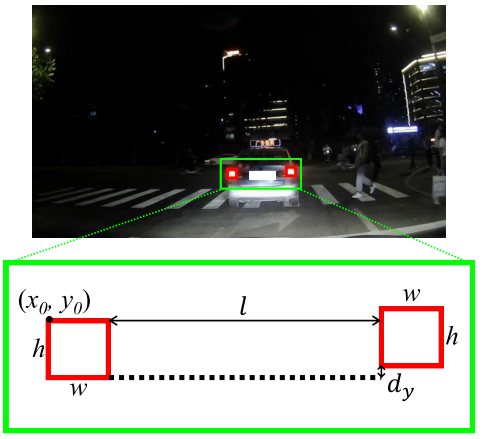


FIGURE 4. Parameters to represent a tracking result in optimization. The ROI is illustrated by a green bounding box, and the rear-lamp regions are represented by two red bounding boxes. There are six parameters in total, including the upper-left coordinates of the left rear lamp (x_0, y_0), the width of both rear-lamps w, the height of both rear-lamps h, the horizontal distance between the two rear-lamps l, and the vertical distance between the two rear-lamps d_y.

Also, r_t is added to the first generation to accelerate the convergence.

We design a balanced fitness function, which is defined by

$$f = \sum_{j=1}^{6} \omega_j f_j, \tag{4}$$

where f_j denotes the j-th of six terms, and ω_j denotes the corresponding weight. With letting $\sum_{j=1}^{6} \omega_j = 1$ and normalizing each f_j to $[0, 1]$, the problem is defined as a maximization problem of Eq. 4 within $[0, 1]$. The weights are empirically determined and fixed throughout the experiments: $\omega_1 = 0.10$, $\omega_2 = 0.60$, $\omega_3 = 0.10$, $\omega_4 = 0.05$, $\omega_5 = 0.10$, and $\omega_6 = 0.05$.

The first term f_1 considers the spatial relationship, which is calculated by

$$f_1 = 1.0 - 2|d'_y - 0.5|, \tag{5}$$

where $d'_y = \frac{d_y - d_{min}}{d_{max} - d_{min}}$ is the normalized d_y to satisfy $f_1 \in [0, 1]$, in which $d_{max} = h - h_{min}$, and $d_{min} = -d_{max}$. Eq. 5 turns out to be closer to 1 when $d_y \to 0$ because we suppose the heights between the left and the right rear-lamps are the same.

TABLE 1. Adaptive thresholds.

	H	S	V
Cluster 1 of the "red"	$[156, 180]$	$[75, 150]$	$[200, 255]$
Cluster 2 of the "red"	$[0, 5] \cup [171, 180]$	$[118, 255]$	$[51, 255]$
Cluster 3 of the "red"	$[0, 13]$	$[53, 128]$	$[130, 230]$
The "white"	$[0, 255]$	$[0, 117]$	$[179, 255]$

The second term f_2 considers color information, in which adaptive thresholds are proposed. Casares *et al.* [10] and Nakane *et al.* [13] utilized the fixed thresholds to extract red and white regions, while the extraction result of rear-lamp regions by color can change due to different illuminations. To solve this problem, based on rear-lamp images, we statistically cluster the red pixels in rear-lamp regions into three clusters, as shown in Table 1. In the ROI of each frame, more pixels within a cluster of the "red" represents the cluster is more probably adaptive. For this reason, the cluster of the "red" is selected by

$$\max \begin{cases} C(\{(x, y)|\text{HSV}(x, y) \in M_1, (x, y) \in \text{ROI}'\}), \\ C(\{(x, y)|\text{HSV}(x, y) \in M_2, (x, y) \in \text{ROI}'\}), \\ C(\{(x, y)|\text{HSV}(x, y) \in M_3, (x, y) \in \text{ROI}'\}). \end{cases} \tag{6}$$

In Eq. 6, C$(\cdot)$ is the function of counting pixels, HSV$(\cdot)$ is the function of extracting HSV values, $M_{1\sim3}$ are cluster 1∼3 of the "red" respectively shown in Table 1, and ROI$'$ is the center partition of the ROI, where the ROI$'$ remains the region from the height of $\frac{1}{4}$ to $\frac{3}{4}$. After selecting the cluster, f_2 is calculated by

$$f_2 = \frac{1}{4}\left(\frac{C_{iL} + C_{iR}}{2S_i} + \frac{C_{oL} + C_{oR}}{2S_o}\right), \tag{7}$$

Algorithm 1 The Strategy for Occlusion

input: r_t, E_{t+1}
output: r_{t+1}
  if $C_{iL} = 0$ & $C_{oL} = 0$ **then**
    $r^l_{t+1} = r^l_t, r^r_{t+1} = E^r_{t+1}$
  end if
  if $C_{iR} = 0$ & $C_{oR} = 0$ **then**
    $r^l_{t+1} = E^l_{t+1}, r^r_{t+1} = r^r_t$
  end if
  if $\rho(r^l_t, r^l_{t+1}) \geq 0.8$ **then**
    $r^l_{t+1} = E^r_{t+1} - E^r_{t+1}\frac{r^r_t - r^l_t}{r^r_t}, r^r_{t+1} = E^r_{t+1}$
  else if $\rho(r^r_t, r^r_{t+1}) \geq 0.8$ **then**
    $r^l_{t+1} = E^l_{t+1}, r^r_{t+1} = E^l_{t+1} - E^l_{t+1}\frac{r^l_t - r^r_t}{r^l_t}$
  else
    $r_{t+1} = E_{t+1}$
  end if
return: r_{t+1}

where C_{iL} and C_{iR} are the counting number of the "red" pixels inside the left and the right rear-lamp bounding boxes, respectively, S_i denotes the sum pixel number within the candidate region, C_{oL} and C_{oR} are the counting number of the "white" pixels on the bounding box of the left and the right bounding boxes, respectively, and S_o denotes the sum pixel number on the bounding box of candidate region.

f_3 and f_4 consider the symmetry of rear-lamps. Among them, f_3 evaluates the fitness by the visual (color) symmetry, which is calculated by

$$f_3 = 1.0 - |\frac{C_{iL}}{S_i} - \frac{C_{iR}}{S_i}|, \tag{8}$$

i.e., a closer difference of effective pixels between the left and right rear-lamps leads to a larger f_3. Owing to the rear-lamp in this research belongs to the preceding vehicle, f_4 utilizes the zero-means normalized cross-correlation (ZNCC) computed from the grayscale image, which is calculated by

$$\begin{aligned} f_4 &= \frac{\text{ZNCC}(I_1, I_2) + 1.0}{2.0}, \\ \text{ZNCC}(L, R) &= \frac{\sum_{p=1}^{D} (I_1(p) - \text{mean}(I_1))(I_2(p) - \text{mean}(I_2))}{\sqrt{\sum_{p=1}^{D} (I_1(p) - \text{mean}(I_1))^2 (I_2(p) - \text{mean}(I_2))^2}}, \end{aligned} \tag{9}$$

where D represents the number of pixels in both the left patch I_1 and the right patch I_2 which are illustrated in Fig. 4. mean$(\cdot)$ is a function to calculate the mean intensity value of an input patch and p denotes a pixel. Patches with the larger ZNCC value are considered a better match. The fourth part f_4 makes contribution to preventing the bounding box from the false pairing between rear-lamp and the background.

The fifth part f_5 evaluates the relationship to the last frame. The result r_{t+1} is considered more probably to be located within a range surrounded with r_t, which is a set that consists

TABLE 2. List of the attributes tagged to the sequences.

Attribute	Description
IV	**Illumination Variation** - the illumination in or near the target region changes.
SV	**Scale Variation** - the scale of the target region changes overtime.
OC	**Occlusion** - the target is partially or fully occluded.
BC	**Background Clutters** - the background near the target has the similar color or texture as the target.
FA	**Far** - the target region is small over the sequence.
CL	**Close** - the target region is large over the sequence.

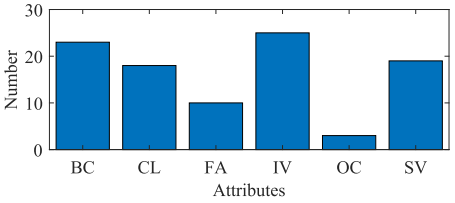


FIGURE 5. The distribution of the attributes. The attribute that concerns most sequences is IV (25 sequences), the least is OC (3 sequences).

of the left and right bounding boxes, thus a threshold of this range $\gamma = \frac{1}{4} \cdot r_t^l = \frac{1}{4} \cdot r_t^r$. Then f_5 can be described as

$$f_5 = \begin{cases} 1 - 2\dfrac{|r_{t+1} - r_t|_2}{f_5^{\max}}, & \text{if } |r_{t+1} - r_t|_2 > \gamma \\ 1, & \text{otherwise} \end{cases} \tag{10}$$

in which $f_5^{\max}$ is a constant to satisfy $f_5 \in [0, 1]$, r_t and r_{t+1} are vectors that contain both the center points of the left and the right bounding boxes in two consecutive frames. The sixth part f_6 evaluates the rigidity. The change of the structure of left and right rear-lamps from frame t to $t + 1$, such as mutual distance and length-width ratio, would not be radical, i.e., rear-lamps in adjacent frames have rigidity. The rigidity is evaluated by

$$\begin{aligned} f_6 &= \frac{1}{w_I}\left(\frac{p_{t+1}}{p_{t+1} - p_t} + \frac{q_{t+1}}{q_{t+1} - q_t}\right), \\ p &= \frac{l}{w}, q = \frac{w}{h}, \end{aligned} \tag{11}$$

where w_I denotes the width of the whole image, and guarantees $f_6 \in [0, 1]$.

The rigidity also helps to improve the results. The results given by GA without rigidity are not permitted. Specifically, when the results given by optimal parameters occur $\rho(r_t^l, r_{t+1}^l) \geq 0.8$ or $\rho(r_t^r, r_{t+1}^r) \geq 0.8$, where $\rho(\cdot, \cdot)$ denotes Euclidean distance, the corresponding r_{t+1}^l or r_{t+1}^r is meliorated by rigidity. Specifically,

$$(x, y)_{t+1}^l = (x, y)_{t+1}^r - (x, y)_{t+1}^r \frac{(x, y)_t^r - (x, y)_t^l}{(x, y)_t^r}, \quad \text{if } \rho(r_t^l, r_{t+1}^l) \geq 0.8, \tag{12}$$

similarly, $(x, y)_{t+1}^r$ can be meliorated when $\rho(r_t^r, r_{t+1}^r) \geq 0.8$.

TABLE 3. Comparison between proposed method and online tracking methods with respect to average success rate (%).

Attribute	Ours	CSK [5]	CT [6]	DFT [7]	LOT [8]
IV	**89.54**	75.09	75.76	49.27	50.04
SV	**93.81**	77.63	75.47	60.68	52.03
OC	**94.91**	71.71	87.40	56.64	62.50
BC	**90.53**	83.04	79.58	64.28	48.06
FA	**85.66**	80.52	59.99	62.13	39.86
CL	**94.29**	81.77	85.76	69.34	51.70
ALL	**91.29**	78.87	79.42	62.52	48.77

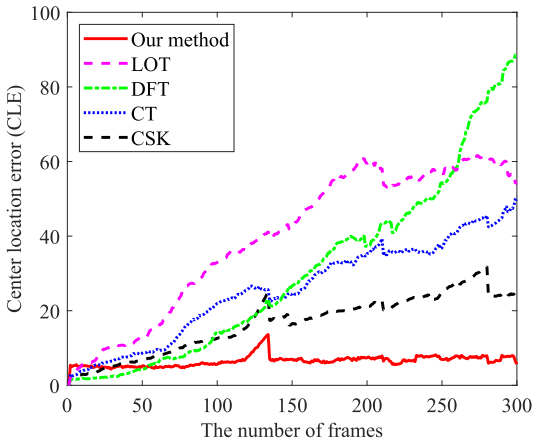


FIGURE 6. Comparison of average CLE over the whole dataset (300 frames × 50 sequences). Best view in color.

4) THE STRATEGY FOR OCCLUSION

If the occlusion occurs partially, the part that is not occluded can be the evidence for tracking, so the complete occlusion is mainly discussed here. When the complete occlusion happens, there could be two conditions: 1) when our method locates in other lamp-like regions that can be resolved by Eq. 12; 2) when our method suffers a series of negative chromosomes, meaning $C_{iL} = 0$ & $C_{oL} = 0$ or $C_{iR} = 0$ & $C_{oR} = 0$. For the second condition, r_{t+1} is set the same as r_t. The strategy is summarized in Algorithm 1, in which E_{t+1} denotes the results of frame $t + 1$ that are not processed with the strategy for occlusion, and the superscript l and r denote left and right, respectively.

IV. EXPERIMENT

A. NIGHTTIME REAR-LAMP TRACKING DATASET

In order to evaluate rear-lamp tracking at nighttime, a novel dataset is created in this research. The dataset contains 50 sequences consist of 19558 frames and 39116 annotated bounding boxes. Each sequence is taken by a driving recorder in 1280 × 720 pixels resolution. Sequences are collected under different traffic situations and annotated with attributes, including illumination variation (IV), scale variation (SV), occlusion (OC), background center location error (CLE) and success rate (SR) are employed as evaluation criteria, which are common under the scenario of object tracking (e.g., [37]–[39]). In the experiment, the average processing

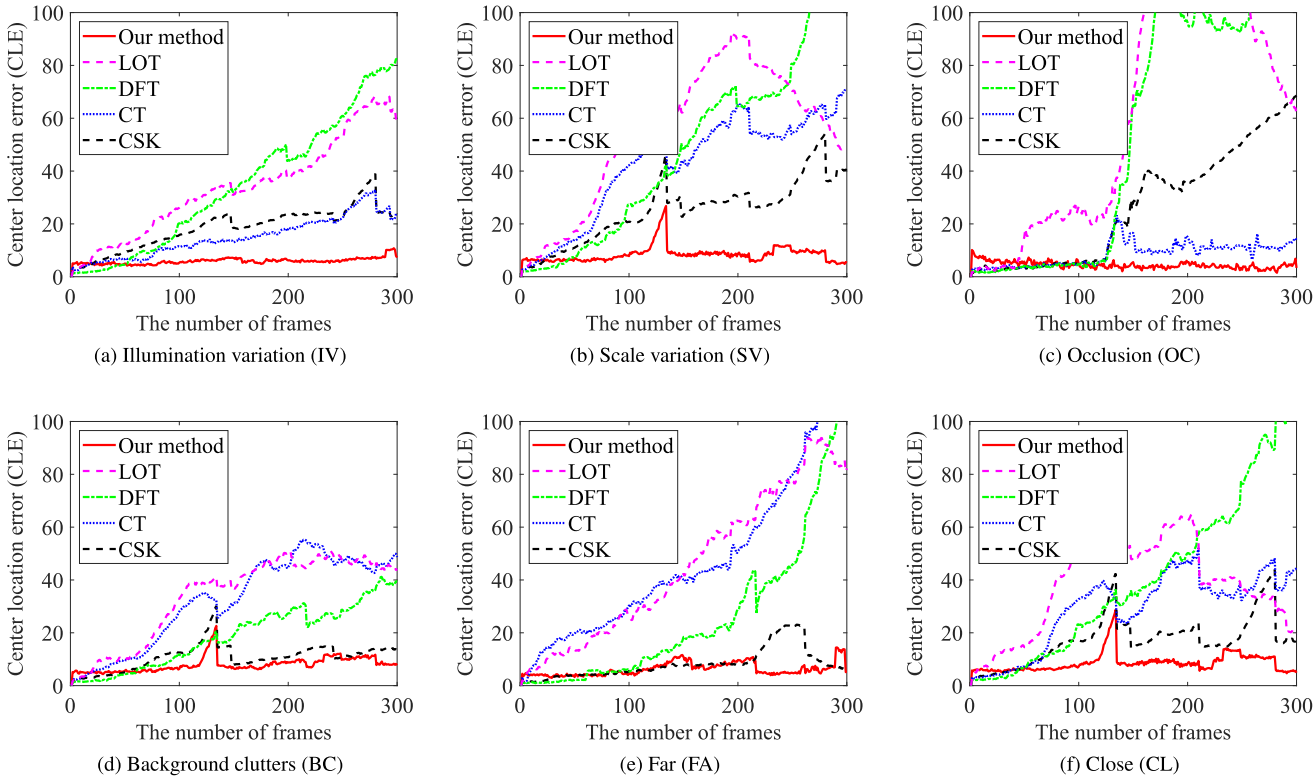


FIGURE 7. Comparison on attribute-wise subset of the dataset. Best view in color.

time of our method with respect to all the sequences is 3.57fps (implemented on Windows with Intel i9-10900X 3.7GHz and 32GB RAM, the programming language is C++). Both CLE and SR are defined on a single rear lamp. CLE is the distance between the center of ground truth and the result. In a single frame, the tracking result of a single rear lamp is treated as successful if the overlap coefficient of bounding boxes is more than 0.5. In this paper, the overlap coefficient is calculated by dividing the intersection of the ground-truth box and the result box by the smaller one. Because rather than the sizes of bounding boxes, we care more about the position of the tracking result in this task. Background clutters (BC), far (FA), and close (CL) for analysis. The information of each attribute is summarized in Table 2. The number of sequences with respect to each attribute is shown in Fig. 5.

B. EVALUATION CRITERIA

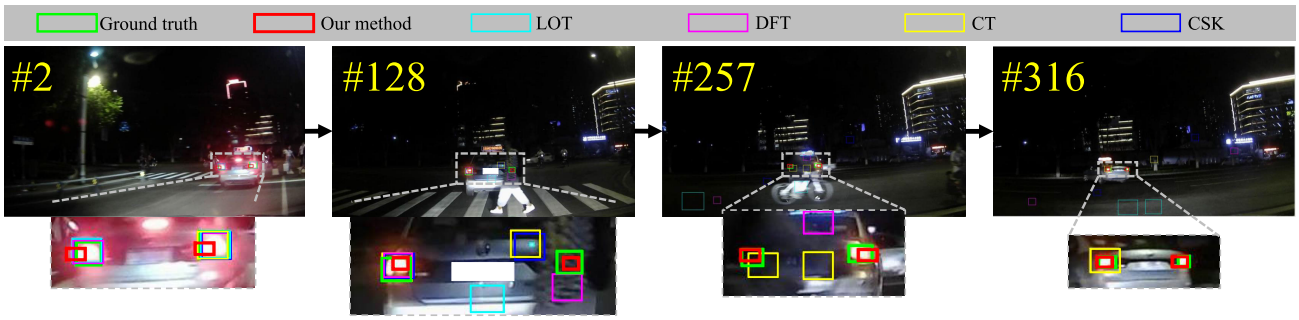
Center location error (CLE) and success rate (SR) are employed as evaluation criteria, which are common under the scenario of object tracking (e.g., [37]–[39]). Both CLE and SR are defined on a single rear lamp. CLE is the distance between the center of ground truth and the result. In a single frame, the tracking result of a single rear lamp is treated as successful if the overlap coefficient of bounding boxes is more than 0.5. In this paper, the overlap coefficient is calculated by dividing the intersection of the ground-truth box and the result box by the smaller one. Because rather than the sizes of bounding boxes, we care more about the position of the tracking result in this task.

C. RESULTS AND ANALYSIS

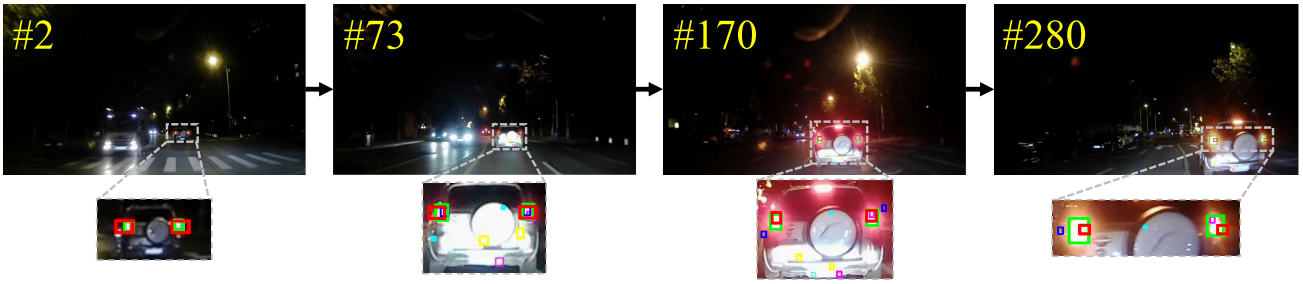
The quantitative experiment consists of the evaluation of overall performance and attribute-wise performance. Online tracking algorithms CSK [8], DFT [7], CT [6], and LOT [5] are employed for comparison.

1) OVERALL PERFORMANCE ANALYSIS

The average success rate and CLE plots over the whole dataset are shown in the last row of Table 3 and Fig. 6 respectively. Generally, our method outperforms other online trackers with respect to both criteria. It is worth pointing out that during the first 50 frames, especially the beginning of the frames, the curve of our method is inferior to some of the other methods. The reason is that our method tends to track the part of the rear-lamp which fits the color model best, rather than the whole region recognized by the human eye, which results in a relatively larger average CLE. On the other hand, as can be observed from the trend of the CLE curves, our method tracks the target well over a long period of time comparing against other trackers as the CLE increases after losing the target. There exists a peek at the red curve around the 130-th frame because a challenging tracking scenario that leads to failure happens.

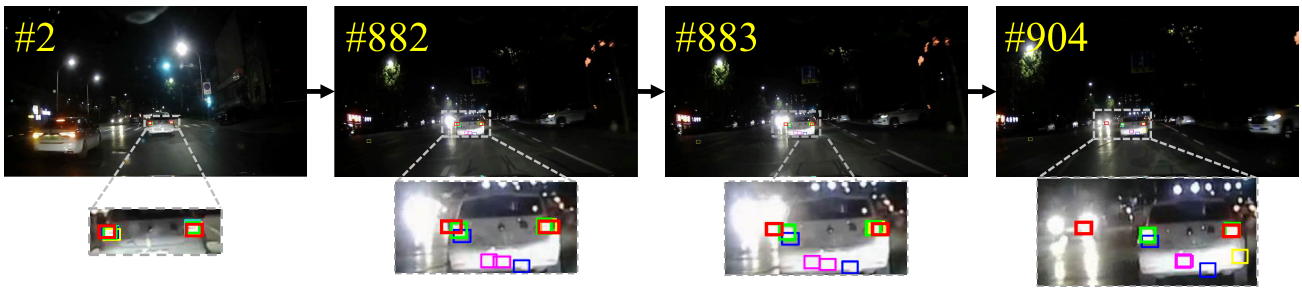


(a) Sequence with labels: scale variation (SV), occlusion (OC).



(b) Sequence with labels: illumination variation (IV), scale variation (SV).

(c) Sequence with labels: illumination variation (IV), background clutters (BC), and far (FA).



(d) Sequence with labels: illumination variation (IV).

FIGURE 8. Visual comparison between our method and online trackers. Number at the top left of each image indicates the frame ID. Legends are shown at the top. Best view in color.

2) ATTRIBUTE-WISE PERFORMANCE ANALYSIS

Our dataset is annotated with the attributes in Table 2, in order to facilitate a more comprehensive analysis on the performance. Because each sequence could have more than one label, each subset probably has an intersection with each other. The attribute-wise SR is shown in Table 3, and CLE curves are shown in Fig. 7. As can be observed, our method outperforms other online trackers in different aspects with respect to both SR and CLE in the rear-lamp tracking problem.

3) QUALITATIVE ANALYSIS

The visual results are shown in Fig. 8. Overall, all the methods can track the rear-lamps to some extent. With the challenging

TABLE 4. SR (%) comparison by varing the weights.

Value	0.10	0.30	0.50	0.70	0.90
ω_1	91.29	74.33	72.22	75.02	74.73
ω_2	54.24	74.27	83.02	68.66	45.38
ω_3	91.29	72.81	66.57	67.33	51.77
ω_4	76.19	79.87	68.01	59.45	24.56
ω_5	91.29	81.40	81.80	75.03	68.43
ω_6	73.43	48.80	26.26	29.60	24.58

factors involved (e.g., from the second column in Fig. 8), some of the methods start to fail. Our proposed method can track the rear-lamps from first to last in these three sequences. Especially, Fig. 8(a) evaluates the SV and OC. In frame #128 and #257, the rear-lamp is occluded by a pedestrian and a bicycle respectively, which leads other methods to failure. Our proposed method benefits from the strategy for occlusion to achieve complete tracking results. Fig. 8(b) is a challenging sequence because of intense IV and SV. In frame #2, the preceding vehicle is far and small, while in frame 280, the preceding vehicle is close. The illumination change is mainly caused by the light from the opposite lane, street light, and brake light. Our proposed method benefits from adaptive thresholds to achieve complete tracking results.

Figure 8(c) suffers intense IV and BC, especially in frame #285. Also, the preceding vehicle keeps its distance from the camera. In frame #266, the result of our method (left rear-lamp) drifts due to BC. However, in the following frames, by benefiting from the balanced fitness function, our method successes in relocating the target.

4) LIMITATION ANALYSIS

Figure 8(d) suffers intense IV from the headlight of the opposite vehicle, which is closed to the left rear-lamp as shown in frame #882 and #883. Although our method performs well in the previous frames before frame #882, it fails in tracking due to the strong illumination change. It indicates the limitation of our method that intense IV nearby the rear-lamps is still a challenging condition for our method.

To show our parameters are reasonably set, we compare different parameter settings by varying the empirically determined one ($\omega_1 = 0.10$, $\omega_2 = 0.60$, $\omega_3 = 0.10$, $\omega_4 = 0.05$, $\omega_5 = 0.10$, $\omega_6 = 0.05$) and the results are summarized in Tab. 4 with respect to the whole dataset. Each weight is incrementally increased by 0.20 from 0.10 to 0.90 to show if a more optimal parameter setting can be found. When a certain weight is varied, other weights are adjusted with an equal proportion to make sure that the sum of all weights equals 1. As can be observed in Tab. 4, as long as any of the weights is varied, the success ratio will decrease. The 91.29% of SR is achieved with our empirically determined parameter setting.

V. CONCLUSION

In this paper, an evolutionary algorithm based rear-lamp tracking method for nighttime usage is proposed. PBO based GA is adopted for optimization, with a balanced fitness function proposed, considering the consistency, color information, symmetry, and rigidity. Especially, as to the color information, adaptive thresholds are proposed. The proposed method processed occlusion by the analysis of color information as well. In evaluation, the proposed method performed better than the existing online tracking methods.

For future works, firstly, the proposed method sometimes tracked part of the rear-lamps, without bounding all the regions of the rear-lamps. Thus to solve the problem by region growing or rear-lamp edge detection is one of the future works. Secondly, as one limitation of this paper is that all the hyperparameters are empirically determined, we would like to introduce an optimal or adaptive tuning method for determining the weighting coefficients.

REFERENCES

[1] S. Plainis, "Road traffic casualties: Understanding the night-time death toll," *Injury Prevention*, vol. 12, no. 2, pp. 125–138, Apr. 2006.

[2] N. Stamatiadis, B. Psarianos, K. Apostoleris, and P. Taliouras, "Nighttime versus daytime horizontal curve design consistency: Issues and concerns," *J. Transp. Eng. A, Syst.*, vol. 146, no. 3, Mar. 2020, Art. no. 04019080.

[3] C. Gan, H. Zhao, P. Chen, D. Cox, and A. Torralba, "Self-supervised moving vehicle tracking with stereo sound," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7053–7062.

[4] T. Schamm, C. von Carlowitz, and J. M. Zöllner, "On-road vehicle detection during dusk and at night," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2010, pp. 418–423.

[5] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 702–715.

[6] K. Zhang, L. Zhang, and M.-H. Yang, "Real-time compressive tracking," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 864–877.

[7] L. Sevilla-Lara and E. Learned-Miller, "Distribution fields for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1910–1917.

[8] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," *Int. J. Comput. Vis.*, vol. 111, no. 2, pp. 213–228, Jan. 2015.

[9] D.-Y. Chen, Y.-H. Lin, and Y.-J. Peng, "Nighttime brake-light detection by nakagami imaging," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1627–1637, Dec. 2012.

[10] M. Casares, A. Almagambetov, and S. Velipasalar, "A robust algorithm for the detection of vehicle turn signals and brake lights," in *Proc. IEEE 9th Int. Conf. Adv. Video Signal-Based Surveill.*, Sep. 2012, pp. 386–391.

[11] K. M. Jeong and B. C. Song, "Night time vehicle detection using rear-lamp intensity," in *Proc. IEEE Int. Conf. Consum. Electron.-Asia (ICCE-Asia)*, Oct. 2016, pp. 1–3.

[12] R. O'Malley, E. Jones, and M. Glavin, "Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 453–462, Jun. 2010.

[13] T. Nakane, H. Sun, and C. Zhang, "Rear-lamp localization with real-coded genetic algorithm at nighttime," in *Proc. IEEE 9th Global Conf. Consum. Electron. (GCCE)*, Oct. 2020, pp. 329–330.

[14] D.-Y. Chen and Y.-H. Lin, "Frequency-tuned nighttime brake-light detection," in *Proc. 6th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Oct. 2010, pp. 619–622.

[15] H. Kuang, L. Chen, F. Gu, J. Chen, L. Chan, and H. Yan, "Combining region-of-interest extraction and image enhancement for nighttime vehicle detection," *IEEE Intell. Syst.*, vol. 31, no. 3, pp. 57–65, May 2016.

[16] T. Nakane, T. Akashi, X. Lu, and C. Zhang, "A probabilistic bitwise genetic algorithm for B-spline based image deformation estimation," in *Proc. Genetic Evol. Comput. Conf. Companion*, Jul. 2019, pp. 300–301.

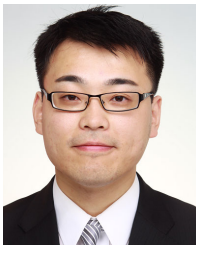
[17] B. Tian, Y. Li, B. Li, and D. Wen, "Rear-view vehicle detection and tracking by combining multiple parts for complex urban surveillance," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 597–606, Apr. 2014.

[18] J.-F. Song, "Vehicle detection using spatial relationship GMM for complex urban surveillance in daytime and nighttime," *Int. J. Parallel Program.*, vol. 46, no. 5, pp. 859–872, Oct. 2018.
[19] F. I. Vancea, A. D. Costea, and S. Nedevschi, "Vehicle taillight detection and tracking using deep learning and thresholding for candidate generation," in *Proc. 13th IEEE Int. Conf. Intell. Comput. Commun. Process. (ICCP)*, Sep. 2017, pp. 267–272.
[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005,pp. 886–893.
[21] S. Suthaharan, "Support vector machine," in *Machine Learning Models and Algorithms for Big Data Classification*. New York, NY, USA: Springer, 2016, pp. 207–235.
[22] S. Tuermer, F. Kurz, P. Reinartz, and U. Stilla, "Airborne vehicle detection in dense urban areas using HoG features and disparity maps," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 6, pp. 2327–2337, Dec. 2013.
[23] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1150–1157.
[24] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 404–417.
[25] S. Ozekes and A. Y. Camurcu, "Automatic lung nodule detection using template matching," in *Proc. Int. Conf. Adv. Inf. Syst.* Berlin, Germany: Springer, 2006, pp. 247–253.
[26] R. An, P. Gong, H. Wang, X. Feng, P. Xiao, Q. Chen, Q. Zhang, C. Chen, and P. Yan, "A modified PSO algorithm for remote sensing image template matching," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 4, pp. 379–389, Apr. 2010.
[27] J. Sato and T. Akashi, "Deterministic crowding introducing the distribution of population for template matching," *IEEJ Trans. Electr. Electron. Eng.*, vol. 13, no. 3, pp. 480–488, Mar. 2018.
[28] T. Dekel, S. Oron, M. Rubinstein, S. Avidan, and W. T. Freeman, "Best-buddies similarity for robust template matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2021–2029.
[29] I. Talmi, R. Mechrez, and L. Zelnik-Manor, "Template matching with deformable diversity similarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 175–183.
[30] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*. https://arxiv.org/abs/2004.10934
[31] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2016, pp. 21–37.
[32] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[33] Y. Li and Q. Yao, "Rear lamp based vehicle detection and tracking for complex traffic conditions," in *Proc. 9th IEEE Int. Conf. Netw., Sens. Control*, Apr. 2012, pp. 387–392.
[34] S. Mirjalili, "Genetic algorithm," in *Evolutionary Algorithms and Neural Networks*. Hertfordshire, U.K.: Springer, 2019, pp. 43–55.
[35] O. M. Shir and T. Bäck, "Niching in evolution strategies," in *Proc. Conf. Genetic Evol. Comput. (GECCO)*, 2005, pp. 915–916.
[36] M. Hutter, "Fitness uniform selection to preserve genetic diversity," in *Proc. Congr. Evol. Computation (CEC)*, vol. 1, 2002, pp. 783–788.
[37] Q. He, W. Zhang, W. Chen, G. Xie, and Y. Yao, "Target tracking algorithm combined part-based and redetection for UAV," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, pp. 1–17, Dec. 2020.
[38] C. Kwan, B. Chou, and L.-Y. M. Kwan, "A comparative study of conventional and deep learning target tracking algorithms for low quality videos," in *Proc. Int. Symp. Neural Netw.* Cham, Switzerland: Springer, 2018,pp. 521–531.
[39] Y. Chen, J. Wang, R. Xia, Q. Zhang, Z. Cao, and K. Yang, "The visual object tracking algorithm research based on adaptive combination kernel," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 12, pp. 4855–4867, Dec. 2019.

HAITIAN SUN received the B.E. degree from the College of Information Engineering, Northwest A&F University, China, in 2014, and the M.E. and Ph.D. degrees from Iwate University, Japan. He is currently a Lecturer with the Department of Computer Science, Northwest A&F University. His research interests include visual human pose retrieval, sketch-based image retrieval, and product recognition.

TAKUMI NAKANE received the B.E. and M.E. degrees from the University of Fukui, Japan, in 2017 and 2020, respectively, where he is currently pursuing the Ph.D. degree in advanced interdisciplinary science and technology with the Graduate School of Engineering. His research interests include evolutionary computing and vision-based optimization problems.

NAIDAN ZHANG is currently pursuing the B.S. degree in computer science and technology with the College of Information Engineering, Northwest A&F University, China. Her research interests include computer vision, pattern recognition, and semantic segmentation.



CHAO ZHANG (Member, IEEE) received the Ph.D. degree from Iwate University, Japan, in 2017. He is currently a Senior Lecturer with the Department of Engineering, University of Fukui, Japan. His research interests include computer vision, computer graphics, and evolutionary computing, mainly focused on applying optimization methods to solve visual computing problems.