

Vehicle re-identification in tunnel scenes via synergistically cascade forests

Rixing Zhu^a, Jianwu Fang^{b,a,*}, Shuying Li^c, Qi Wang^d, Hongke Xu^a, Jianru Xue^b,
Hongkai Yu^e

^a School of Electronic & Control Engineering, Chang'an University, Xi'an, China

^b Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China

^c School of Automation, Xi'an University of Posts and Telecommunications, Xi'an, China

^d School of Computer Science and the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China

^e Department of Computer Science, University of Texas - Rio Grande Valley, Edinburg, TX, USA



ARTICLE INFO

Article history:

Received 24 May 2019

Revised 9 September 2019

Accepted 24 November 2019

Available online 4 December 2019

Communicated by Dr. D. Tao

Keywords:

Tunnel surveillance

Tunnel vehicle re-identification

Extremely randomized forest

Random forest

Curriculum learning

ABSTRACT

Nowadays, numerous cameras have been equipped in tunnels for monitoring the tunnel safety, such as detecting fire, vehicle stopping, crashes, and so forth. Nevertheless, safety events in tunnels may occur in the blind zones not covered by the multi-camera monitoring systems. Therefore, this paper opens the challenging problem, tunnel vehicle re-identification (abbr. tunnel vehicle Re-ID), to make a between-camera speculation. Different from the open road scenes focused by existing vehicle Re-ID methods, tunnel vehicle Re-ID is more challenging because of poor light condition, low resolution, frequent occlusion, severe motion blur, high between-vehicle similarity, and so on. To be specific, we propose a synergistically cascade forests (SCF) model which aims to gradually construct the linking relation between vehicle samples with an increasing of alternative layers of random forest and extremely randomized forest. Through the modeling of SCF, we can restrict the influence of little inter-variation of different vehicle identities and large intra-variation of the same identities. This paper constructs a new and challenging tunnel vehicle dataset (Tunnel-VRelD), consisting of 1000 pairs of tunnel vehicle images. Extensive experiments on our Tunnel-VRelD demonstrate that the proposed method can outperform current state-of-the-art methods. Besides, in order to prove the adaptation ability of SCF, we also verify the superiority of SCF on a large-scale vehicle Re-ID dataset, named as VehicleID, collected in open road scenes.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Tunnels have become the important component in the road network, and own the apparent superiority for shortening the transportation distance. However, as a double-edged sword, because of the limited space, narrow fields of view and dim illumination condition, there is significantly higher risk and severer injuries of accidents in tunnels than on open stretches of road [1,2]. Therefore, multi-camera monitoring systems are universally equipped in tunnels for guaranteeing safety control. The intention is to discover the safety events timely, thereby traffic managers can rescue immediately to avoid severe hazards and secondary accidents.

Although many cameras are equipped in current tunnels, the camera array is always with blind zones, as shown in Fig. 1. If there are safety events in the blind zones, such as vehicle stopping or collisions, the observed vehicles in one camera will not appear in the next one, which plants serious hazard seeds to tunnel safety. In this work, we tackle this problem by introducing vehicle re-identification (Re-ID) technique, while it owns a sufficient consideration of the vehicle characteristics in tunnel scenes.

Tunnel vehicle Re-ID aims to match the corresponding same vehicles from different cameras, so it can detect whether the identity of vehicle lost or not after passing through the blind zone. Another promising application of tunnel vehicle Re-ID is to keep tracking of interested vehicles, such as buses taking too many passengers and trucks carrying dangerous goods with high risk for tunnel safety. By deducing the possible position of safety events or interested vehicles in the tunnel, suitable arrangements can be implemented timely. However, different from the traditional vehicle Re-ID works [3,4] focusing on the vehicle samples captured under favourable

* Corresponding author.

E-mail addresses: rixingzhu@chd.edu.cn (R. Zhu), fangjianwu@chd.edu.cn (J. Fang), lishuying@xupt.edu.cn (S. Li), crabwq@nwpu.edu.cn (Q. Wang), xuhongke@chd.edu.cn (H. Xu), jrxue@mail.xjtu.edu.cn (J. Xue), hongkai.yu@utrgv.edu (H. Yu).

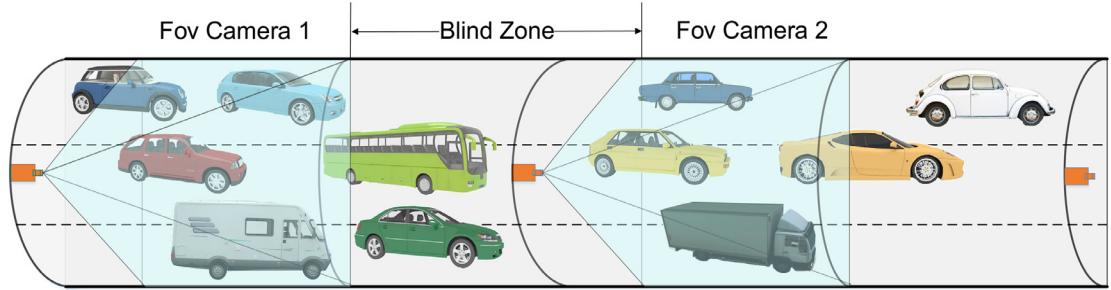


Fig. 1. Illustration of multi-camera setup and blind zones in tunnel monitoring system.



Fig. 2. Examples of vehicle image pairs from our Tunnel-VReID dataset. The vehicles in tunnels have large intra-variation and little inter-variation. Each column of (a) shows some examples with large intra-variation between same identities of vehicles in each column, and (b) demonstrates some typical vehicles with quite similar appearance but different identities of vehicles in each column.

light condition and observing view in open road scenes, tunnel vehicle Re-ID is more challenging owing to following factors.

- Due to dim illumination and fast movement of vehicles, the vehicles in images captured by the cameras are almost all blurred. License plate information and fine-grained appearance features are not available.
- Non-uniform distributed illumination, dazzling rear lights and reflection of tunnel lamps on the vehicle surface cause drastic influence for vehicle appearance.
- Because the viewing fields of cameras are limited by the special linear constraint of the tunnel, pose change and occlusion of vehicles are also intractable.
- There are so many vehicles with similar type that the vehicles with same identity are difficult to be picked out exactly, even by humans.

All these disturbances entangle and degrade the vehicle appearance, and make the vehicles have rather large intra-variation and little inter-variation, which is more challenging than the ones in open road scenes. Fig. 2 demonstrates some pairs of vehicle samples. Note that the cameras are all with high resolution. From

Fig. 2, we can observe that tunnel vehicle Re-ID is rather challenging.

Considering above challenges, we aims to tackle this problem from the exploration of the between-vehicle structure relation. In other words, when matching identities, we not only consider their own similarity, but also introduce the contextual relationships with other vehicle samples. In order to fulfill it, this paper proposes a novel synergistically cascade forest model (SCF) for partitioning the different vehicles and correlating the same vehicle by relationship measurement. Specifically, SCF is constructed by alternative layers of random forests and extremely randomized forests, and is gradually modeled by putting each pair of training identities into the construction process. Similarly, the positive samples are the vehicle pairs with same identities while the vehicle pairs with different identities are negative ones. Particularly, the positive-negative classification procedure is modeled as a progressive multi-layer forests ensemble, mainly targeting the deep correlation of samples with the increasing of the layers. By this procedure, the classification margin between positive and negative pairs of vehicle samples can be enlarged gradually (will be illustrated in the discussion).

In order to boost the training efficiency and effectiveness, we further introduce curriculum learning to pave an easy-to-difficult learning framework, in which the degree of difficulty is determined by Euclidean distance of feature vectors of vehicle pairs. In this paper, we construct a dataset for tunnel vehicle Re-ID problem (named as Tunnel-VReID), which contains 1,000 pairs of identities collected from *nine pairs of high-resolution cameras* in three different tunnels. It is worth noting that these 1,000 pairs of identities took laborious labeling work by ourselves because of the rough appearance. The proposed method is evaluated to be effective through exhaustive experiments on our Tunnel-VReID. In addition, in order to prove the adaptation ability of our model for vehicle Re-ID task in open road scenes, we also validate its superiority on a large-scale VehicleID [5] dataset.

The main **contributions** of this paper are as follows:

- This work aims to address a large-scale tunnel vehicle Re-ID problem with rather diverse and challenging tunnel scenes, and extends the scenarios of vehicle Re-ID task.
- An effective SCF model is proposed to model the deep correlation of vehicle samples and enhance the classification capacity of positive and negative vehicle pairs with the increasing of the forest ensemble layers.
- A new and challenging tunnel vehicle Re-ID dataset (Tunnel-VReID) with 1000 pairs of vehicle identities in tunnel scenes is constructed. Exhaustive experiments demonstrate that the proposed method can generate a state-of-the-art performance.

The rest of the paper is organized as follows. **Section 2** presents the related works. A detailed description of the tunnel vehicle Re-ID dataset is presented in **Section 3**. In **Section 4**, we present the proposed method, and **Section 5** gives the evaluation of our method and the comparison with previous works. Finally, the conclusion is stated in **Section 6**.

2. Related works

In recent years, the Re-ID problem has been extensively studied in computer vision. Most existing Re-ID methods focused on person Re-ID while vehicle Re-ID has drawn fewer attention.

Vehicle Re-ID. Vehicle Re-ID is an attractive topic for traffic safety. Early study [6] employed automatic vehicle identification (AVI) tags to carry out vehicle Re-ID through information exchange between transponders equipped on vehicles and AVI readers equipped on roadside. Since market penetration of in-vehicle equipment is extremely low, AVI is difficult to be widely put into use. In the works of [7–9], vehicle Re-ID is implemented by matching electromagnetic signatures captured from inductive/magnetic sensors without in-vehicle equipment, which requires extensive detector installation and maintenance efforts in road network. The procedure of signature waveform feature selection and matching is too complex that it is incapable for heavy traffic.

The license plate is the unique identity of individual vehicle, which is a crucial clue for vehicle re-ID task. Generally speaking, vehicle license plate recognition [10–12] is a practical technology. However, low visibility caused by rainy and foggy weather, dim illumination and unavoidable occlusions, vehicle license plate information is usually indistinct and unobtainable. Therefore, license plate Re-ID methods only can work in restricted conditions, such as favorable illumination and low vehicle speed.

With the rapid development of computer vision, more works mainly employ visual appearance feature in vehicle Re-ID, which focuses on learning discriminative features [4,13–15]. Liu et al. [16] extracted GoogLeNet deep features, color and SIFT features, and applied the late fusion strategy to calculate similarity scores for vehicle Re-ID. Tang et al. [17] designed a multi-modal architecture, which integrated LBP texture map and Bag-of-Word-based

Color Name feature into an end-to-end convolutional neural network. Liu et al. [5] proposed a VAMI model and an adversarial architecture to conduct multi-view feature inference to address the multi-view re-ID problem. Apart from global visual features, special marks can also be valuable for Re-ID task [18]. Peng et al. [19] proposed a multi-region model to extract local features from subtle regions to identify vehicles.

In addition, Liu et al. [20] combined visual appearance feature with license plate. This method firstly adopted visual appearance features of vehicle for coarse filtering, and then utilized the Siamese neural network for license plate verification, so as to make a further confirmation. Furthermore, Shen et al. [21] and Wang et al. [13] utilized visually spatio-temporal proposal to address the vehicle Re-ID task. Taking visual similarity into consideration, the spatial-temporal constraints should also be satisfied.

As aforementioned, these methods are based on open road, while vehicle Re-ID in tunnel scene is challenging, resulting from lacking license plate information, low resolution, dim illumination and drastic changing of the appearance of vehicles. Frias et al. [22] proposed Bayesian framework which combined motion and appearance features to solve tunnel Re-ID problem. They employed shape complexity of vehicles, spatial discrepancy, and lane changes, which required a great deal of annotated information. Besides, they paved this problem with three adjacent cameras in a single tunnel and captured 11 minutes of videos in total for experiment, which manifestly cannot display the diversity of tunnel scenes.

Person Re-ID. Compared with vehicle Re-ID, person Re-ID is a similar topic but more widely studied recently. The existing methods can be commonly divided into two categories, namely hand-crafted metric learning based methods and deep learning based methods.

Hand-crafted metric learning methods usually focus on designing distinct features and learning an effective similarity metric. Ideal features are expected to be robust to illumination, pose, and occlusion [23], where many features have been applied, such as variations on color histograms [24,25], local binary patterns [25,26], SIFT [27], and so on. Metric learning aims to learn an effective similarity metric to compare the features across images. Particularly, XQDA [25], MLAPG [28] have shown to be favorable for person Re-ID.

Owning to prominent advances of deep convolutional neural networks on computer vision, it is popular to apply the deep architecture to person Re-ID task [29–33]. Common deep learning based methods are end-to-end systems by automatically learning features and metrics for person Re-ID. These methods are significantly different in their network architectures structured by flexible convolution layers, fully connected layers and loss functions. In addition, there are many other approaches for Re-ID, such as supervised smoothed manifold [34,35], graph theoretic algorithm [36], saliency embedding matching [37], symmetry-driven accumulation of local features [38], generative adversarial network [39], and so on, where each of them explored its own functional characteristics, respectively.

Compared with person Re-ID, the interference from plenty of similar samples in vehicle Re-ID datasets should be paid more attention, because the vehicle type is not private and is shared by different persons universally. In particular, in our Tunnel-VReID dataset, this interference is strongly aggravated by the poor light condition and motion blur.

3. Tunnel-VReID dataset description

In this paper, a new dataset, called Tunnel-VReID, is constructed to evaluate the tunnel vehicle Re-ID performance. The vehicle images in Tunnel-VReID were captured by 9 pairs of

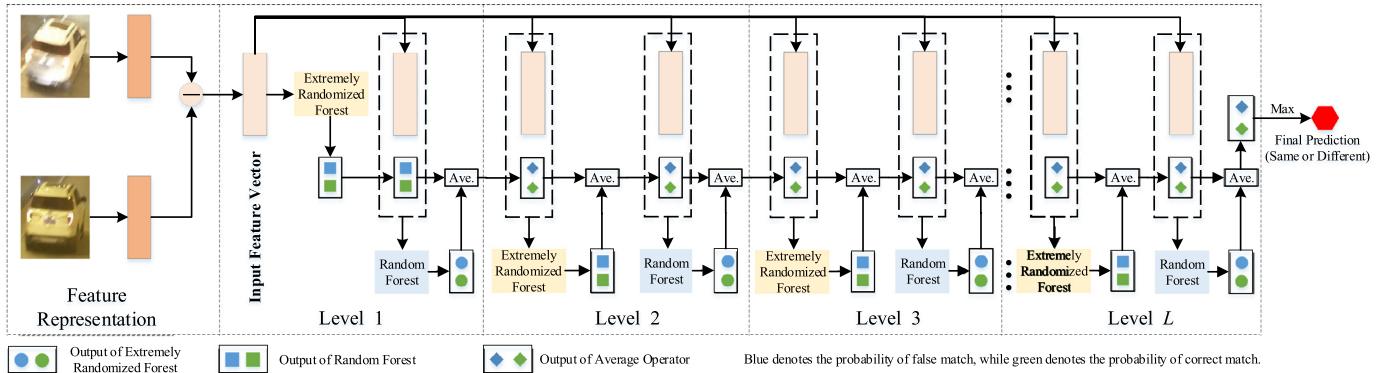


Fig. 3. An overview of the synergistically cascade forests structure. Tunnel vehicle Re-ID is formulated as a progressive multi-layer same-different binary classification framework. Input feature vector is obtained by difference operation of vehicle features from different cameras. Each forest will output a two-dimensional class vector of probabilities of correct match and false match. Each level of the cascade consists of two layers, i.e., the extremely randomized forest and the random forest. The output of the previous layer is used as the augmented feature concatenating with the original feature vector, and taken as the input of the next layer. Extremely randomized forest and random forest are performed alternatively.

1920 × 1080 HD surveillance cameras in three different tunnels located at Xi'han Highway, Shaanxi province, China, which crosses the famous Qinling mountains. Tunnel-VReID dataset includes 1,000 pairs of vehicle identities, which consists of 739 cars, 60 midsize cars (e.g., SUV) and 201 trucks. Because of the severe disturbance of poor light condition, fast motion blur and high between-vehicle similarity, the annotation is rather laborious and has been checked for several times.

3.1. Construction procedure

With the eighteen cameras, the total time duration of the captured video sequences is 340 minutes. Considering the resolution of vehicle images, we pick out images of vehicle identities as clear as possible. Each vehicle identity contains two images captured by two non-overlapping cameras. Then we annotate bounding boxes for vehicles in each image that surrounds the whole vehicle body and suppresses the background as much as possible. Finally, all vehicle images are normalized to 300 × 200 pixels. We take one camera view as the probe set, another one as the gallery set, and then randomly divide the pairs into equal half for training and the other half for testing. We will release Tunnel-VReID dataset in the near future.

3.2. Tunnel-VReID dataset at a glance

In order to guarantee the complexity of Tunnel-VReID, we take full account of various scenes and challenging factors when picking the samples, such as illumination, viewpoint, motion blur, and occlusion. These eighteen cameras in this work are distributed at the exit, entrance and inside of tunnel where illumination conditions are rather different. The illumination near entrance is brighter than the one inside of tunnel. In view of mounting positions and angles of cameras in tunnel, the capturing views of cameras mainly have two orientations. One mostly photographs at the rear of vehicle, while the other may pay more attention to the roof and side of vehicle. Because the designed speed range ([60,80] km/h) in the tunnels of Xi'han Highway, motion blur is universal for all of the vehicles. The challenges of illumination, viewpoint and motion blur give rise to drastic change of vehicle appearance. Besides, the tunnel traffic flow is very large, and many similar vehicles often appear in the same camera view. From these factors, the samples in this dataset have rather little inter-variation and severe large intra-variation.

4. The proposed method

The tunnel vehicle Re-ID is a challenging problem due to large intra-variation and little inter-variation. For distinguishing different vehicles and recognizing same vehicles better, we model a cascade forests structure to explore the deep correlation between vehicles inspired by gcForest [40]. GcForest generates a deep forest ensemble with a cascade structure to do representation learning, which is an alternative to deep neural networks, and is much easier in hyper-parameter tuning and has favorable performance on both small-scale and large-scale datasets. However, the structure of gcForest directly treats the output of the preceding level as the augmented input of the next level, which will cause drastic fluctuation of results between two adjacent levels (demonstrated in our experiments in Section 5.5.1). One underlying reason of this phenomena can be interpreted as follows. Because the challenging factors aforementioned before, many of them are hard to be re-identified, and can be treated as hard samples. Based on the observation by [41], the predicting labels for hard samples by different random forests in the cascade structure may appear divergence and cause large prediction variance, which may cause a confused and fluctuant prediction.

Considering this issue, we propose a new synergistically cascade forests (SCF) structure for tunnel vehicle Re-ID adopting an average strategy of two adjacent levels to restrict this phenomena. Actually, the average strategy can play a role for restricting the influence from the large predicting variance, and transfer the samples without clear distinction to the next level for further distinguishing (We will verify this function in the experiments). The proposed SCF structure is constructed by alternative layers of random forest and extremely randomized forest, which has more powerful ability for distinguishing different identities of vehicles with rather little inter-variation and recognizing same identities of vehicles with large intra-variation with the increasing of the number of layers. The method flowchart is demonstrated in Fig. 3.

4.1. Synergistically cascade forests

4.1.1. Tunnel vehicle Re-ID formulation

We treat tunnel vehicle Re-ID problem as a same-different binary classification problem. Assume the label of a pair of vehicle images ($\mathbf{I}_i, \mathbf{I}_j$) from different cameras is l_{ij} . If \mathbf{I}_i and \mathbf{I}_j belong to the same identity, then $l_{ij} = 1$ (positive sample), and $l_{ij} = 0$ (negative sample) vice versa. We perform difference operation on feature vectors of two vehicle images ($\mathbf{I}_i, \mathbf{I}_j$), and treat it as the input vectors of SCF model. Therefore, training samples \mathbf{X} are formulated

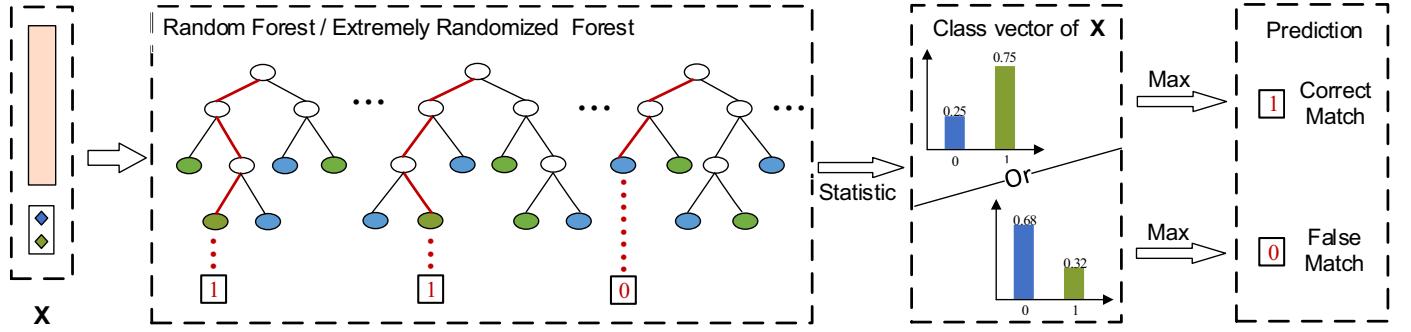


Fig. 4. Illustration of classification process via random forest/extremely randomized forest. The linking line marked by red color highlights paths, along which the instance traverses to leaf nodes. The same instance will be classified as different classes at the leaf node by different decision trees. Through statistics, the percentage of two classes constitutes a two-dimension class vector representing probabilities of false match (blue) and correct match (green), and then the class of larger probability is determined as the prediction.

as follows:

$$\mathbf{X} = \{\mathbf{d}_{ij}\}_{i,j=1}^N = \{F(\mathbf{I}_i) - F(\mathbf{I}_j)\}_{i,j=1}^N, \quad (1)$$

where \mathbf{d}_{ij} is the difference between feature vectors of $(\mathbf{I}_i, \mathbf{I}_j)$, N is the number of training identities of vehicles, and the operator $F(\cdot)$ is the function to extract feature vector of vehicle image \mathbf{I} .

After constructing \mathbf{X} , we feed them into the construction of the forest ensembles in SCF model gradually by maximizing the classification accuracy of different identities and same identities. Therefore, the key factor to achieve effective tunnel vehicle Re-ID is how to design $F(\cdot)$ and SCF model.

4.1.2. Feature representation

Since lighting condition of tunnel scene is poor and camera calibration information is scarce, the work is purely based on vehicle appearances without license plate and spatio-temporal information. Color and texture are efficient for image feature representation, and are commonly utilized for the complementary properties of multiple features. In order to make a discriminative feature representation, we combine the low-level and high-level semantic features together. Specifically, for the low-level feature, this paper uses Local Maximal Occurrence (LOMO) [25] feature because of its good discriminative ability. LOMO combines HSV color histogram with Scale Invariant Local Ternary Pattern (SILTP) descriptor. When extracting features, it not only needs to acquire rich image details, but also should avoid too large dimension and excessive interference. We choose a subwindow size of 20×20 with overlapping step of 15 pixels to describe details. We utilize a three-scale pyramid representation down-sampling by average pooling operation, then extract SILTP_{4,3}^{0,3} histogram and $8 \times 8 \times 8$ -bin HSV histogram within each subwindow. It not only considers the multi-scale information, but also gains abundant color and texture information. For 300×200 vehicle image, all computed histograms are concatenated to form LOMO feature vector with 18976-dimension. Although the feature dimension is high, it can be efficiently classified by following SCF model because of the powerful processing ability of random forest for high-dimensional data [42,43].

Due to different view settings and illumination across cameras, there is drastic change of the perceived color of the same vehicles under different cameras. In the preprocessing of LOMO feature extraction, we perform Retinex [44] to acquire a positive color approximation, generating a more consistent color representation than original one. Actually, for the image representation, there has been many successful efforts, such as the spatial pyramid-enhanced structure [45], multi-view feature learning [46], and hierarchical feature selection [47]. In this work, we focus on a combination of low-level and high-level features to make a discriminative feature representation.

For the high-level semantic features, we adopt the VGG-19, VGG-F and ResNet-152 models pre-trained on ImageNet to obtain CNN feature vectors after the last pooling layer of the CNN models, respectively. The CNN feature is concatenated with LOMO feature with a [0,1]-normalization, and is fed into our SCF model.

4.1.3. SCF components

SCF model is a multi-layer ensemble of different kinds of forests, which is composed of random forests and extremely randomized forests. The diversity of different kinds of forests is crucial for ensembles. Random forest is a classification ensemble with a set of decision trees [48,49]. During forest construction, bagging method is commonly adopted to generate training sample subsets for building decision trees. In an individual decision tree, a subset of attributes of feature vector is randomly selected for tree growing, and the attribute with best Gini value is chosen as the splitting node. Commonly, subset size is set as \sqrt{d} , where d is the dimension of input feature vector. The tree stops growing when leaf nodes only contain the same class of samples. Finally, all trees are combined to form random forest model. Compared with random forest, extremely randomized forest [50] adopts all training samples to build decision trees. It randomly selects one attribute while the splitting value of the attribute with minimum Gini is chosen at each node splitting. The introduction of these random strategies from random forest and extremely randomized forest encourages the diversity of SCF model, which is inherent power to enhance the performance for ensemble method [40,51].

In forest, the same instance will be classified as positive or negative classes at the leaf node by different decision trees. A two-dimension class vector representing probabilities of false and correct match is obtained by calculating the percentage of two classes. Then the class with larger probability is determined as the prediction, as illustrated in Fig. 4.

4.1.4. SCF Structure

SCF is a layer-by-layer cascade structure constituted by random forest and extremely randomized forest, and each level of the cascade is an ensemble of the two kinds of forest alternatively, as illustrated in Fig. 3.

In the first level of SCF model, each original input feature vector is fed into extremely randomized forest, and it outputs class vector of correct match and false match probabilities. Then the original feature vector is concatenated with the class vector of extremely randomized forest as the input of random forest for the second layer. Specially, the order of extremely randomized forest and random forest is not important and can be switched in our model. The two-dimension average of output class vectors of extremely ran-

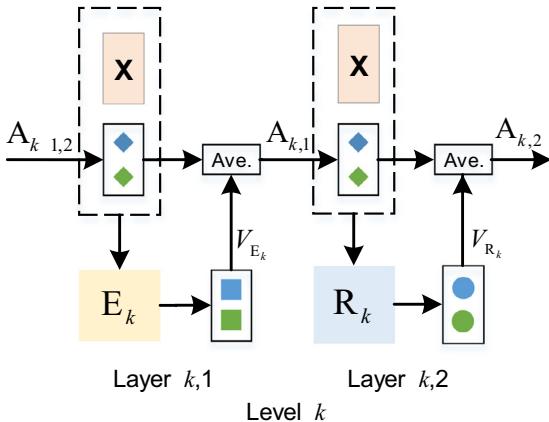


Fig. 5. An illustration of the k th level of recursive SCF structure. The output class vector of each layer is further averaged with the average of preceding layer, and then the newly obtained average is put into the next layer as a new augmented vector.

domized forest and random forest is computed as an augmented feature for the input of the second level.

Starting from the second level to the end of SCF model, the output class vector (two-dimension) of each layer is further averaged with the two-dimensional average of the preceding layer, and then the newly obtained average is put into the next layer as a new augmented vector. Each level of the SCF structure is constructed by former extremely randomized forest and latter random forest, namely two layers. To be concise and easy for understanding, the k th level ($k \geq 2$) of the SCF model can be formulated by a recursive form as follows:

$$\mathbf{A}_{k,1} = \frac{1}{2}(\mathbf{A}_{k-1,2} + \mathbf{V}_{E_k}), \quad (2)$$

$$\mathbf{A}_{k,2} = \frac{1}{2}(\mathbf{A}_{k,1} + \mathbf{V}_{R_k}), \quad (3)$$

$$\mathbf{V}_{E_k} = E_k([\mathbf{X}, \mathbf{A}_{k-1,2}]), \quad (4)$$

$$\mathbf{V}_{R_k} = R_k([\mathbf{X}, \mathbf{A}_{k,1}]), \quad (5)$$

where $\mathbf{A}_{k,1}$ denotes the average vector of the former layer of the k th level, while $\mathbf{A}_{k,2}$ denotes the average vector of the latter layer of the k th level. The average vector concatenated with the original feature vector \mathbf{X} is fed into the latter layer as its input vector. $E_k(\cdot)$ and $R_k(\cdot)$ generate the class vectors \mathbf{V}_{E_k} and \mathbf{V}_{R_k} by extremely randomized forest and random forest, respectively. An illustration of the k th level of the SCF model is shown in Fig. 5. Specially, $\mathbf{A}_{1,1} = \mathbf{V}_{E_1} = E_1(\mathbf{X})$ at the first level of the SCF model.

By alternating random forest and extremely randomized forest layers in SCF model, the averaging operation bridging them is beneficial to improve classification capacity and smooth the fluctuation of classification probability between two adjacent layers. With the increasing of model levels, classification accuracy is gradually improved, which can be visually perceived in the training process, as illustrated in Fig. 6. In the first few levels, some samples with same vehicle identities are misclassified into different vehicles, while some samples with different vehicle identities are misclassified into same vehicles, but these mis-classifications are corrected in later levels. In addition, the probability of correct match increases gradually while the probability of false match reduces gradually (will be discussed in the experiments), which demonstrates that SCF structure has the property to stably distinguish different identities of vehicles with little inter-variation and recognize same identities of vehicles with large intra-variation.

4.1.5. Overfitting restriction

In spite of the randomization in the construction of random forest and extremely randomized forest, finite number of training samples easily give rise to overfitting. For this issue, m -fold cross-validation is utilized to divide training samples into m subsets, and one subset is reserved as verification set, the other $m - 1$ subsets are used for training set. Cross-validation is repeated m times and each subset is verified once, then m classification accuracies are averaged to produce the classification accuracy of the current layer, which is denoted by Acc_{E_k} and Acc_{R_k} , representing the classification accuracy of extremely randomized forest and random forest, respectively. The classification accuracy Acc_k of the k th level is estimated by the average of Acc_{E_k} and Acc_{R_k} .

At a certain level of SCF, if its classification accuracy is larger than the ones of previous level and there is no gain in the next T extra levels, the growth of SCF level will stop. The number of SCF levels can be adaptively determined according to the terminal threshold T . Detailed training process is shown in Algorithm 1. The

Algorithm 1 Synergistically Cascade Forests.

Require: Training samples: \mathbf{X} with labels $y = \{l_{ij}\}_{i,j=1}^N$; The terminal threshold of SCF growing: T .
Ensure: The optimal level of SCF: L_{opt}

- 1: Build the initial SCF level $k = 1$:
- 2: Construct extremely randomized forest E_1 based on \mathbf{X} , and output \mathbf{V}_{E_1} and Acc_{E_1} ;
- 3: Construct random forest R_1 based on $[\mathbf{X}, \mathbf{V}_{E_1}]$, and output \mathbf{V}_{R_1} and Acc_{R_1} ;
- 4: $\mathbf{A}_{1,2} \leftarrow (\mathbf{V}_{E_1} + \mathbf{V}_{R_1})/2$;
- 5: $Acc_1 \leftarrow (Acc_{E_1} + Acc_{R_1})/2$;
- 6: **repeat**
- 7: Increase SCF level $k \leftarrow k + 1$;
- 8: Construct E_k based on $[\mathbf{X}, \mathbf{A}_{k-1,2}]$, and output \mathbf{V}_{E_k} and Acc_{E_k} ;
- 9: $\mathbf{A}_{k,1} \leftarrow (\mathbf{A}_{k-1,2} + \mathbf{V}_{E_k})/2$;
- 10: Construct R_k based on $[\mathbf{X}, \mathbf{A}_{k,1}]$, and output \mathbf{V}_{R_k} and Acc_{R_k} ;
- 11: $\mathbf{A}_{k,2} \leftarrow (\mathbf{A}_{k,1} + \mathbf{V}_{R_k})/2$;
- 12: $Acc_k \leftarrow (Acc_{E_k} + Acc_{R_k})/2$;
- 13: Update $Acc_{max} = \text{Max}\{Acc_n\}_{n=1}^k$ and corresponding index $L_{opt} \leftarrow k^*$;
- 14: **until** $Acc_k < Acc_{max}$ and $(k - L_{opt}) \geq T$

level of maximal classification accuracy will be treated as the optimal level L_{opt} of SCF, and T is the terminal threshold of SCF growing. A reasonable setting of T not only saves computing cost, but also prevents accuracy declining. Therefore, the entire SCF model is defined as follows:

$$L_{opt} = k^* = \arg \max_k Acc_k = \arg \max_k (Acc_{E_k} + Acc_{R_k}). \quad (6)$$

For testing process, the classification result will be predicted by the constructed SCF model with L_{opt} levels. The next T levels followed the k^* th layer, without significant performance gain, are discarded to reduce the total number of levels of the model, which can help shorten training time.

4.2. Synergistically cascade forests with curriculum learning

In view of the complex tunnel scenes, intra-variation in same identities and inter-variation in different identities are widely divergent, that is, tunnel vehicle samples have different difficulty levels. Therefore, in this work, we also want to check whether the performance can be boosted by some easy-to-difficult learning methods, such as curriculum learning [52,53]. The larger the

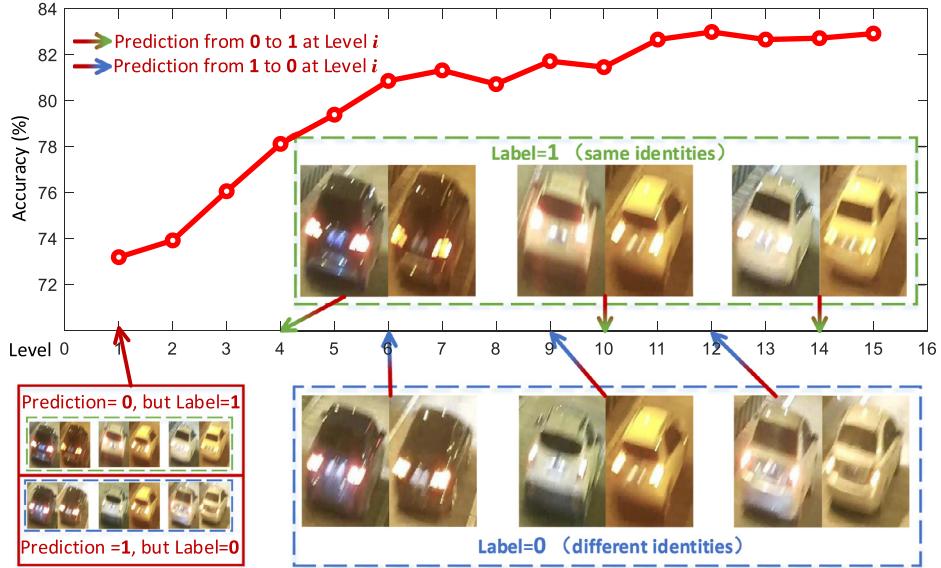


Fig. 6. Examples of prediction change of several samples with the increasing of model levels. The prediction of three samples with same identities (in green box) and three samples with different identities (in blue box) are wrong in the first level, but they are predicted correctly after some levels. From this figure, we further see the challenges for tunnel ReID. These samples are hard to be distinguished even by humans. This figure should be viewed in color mode.

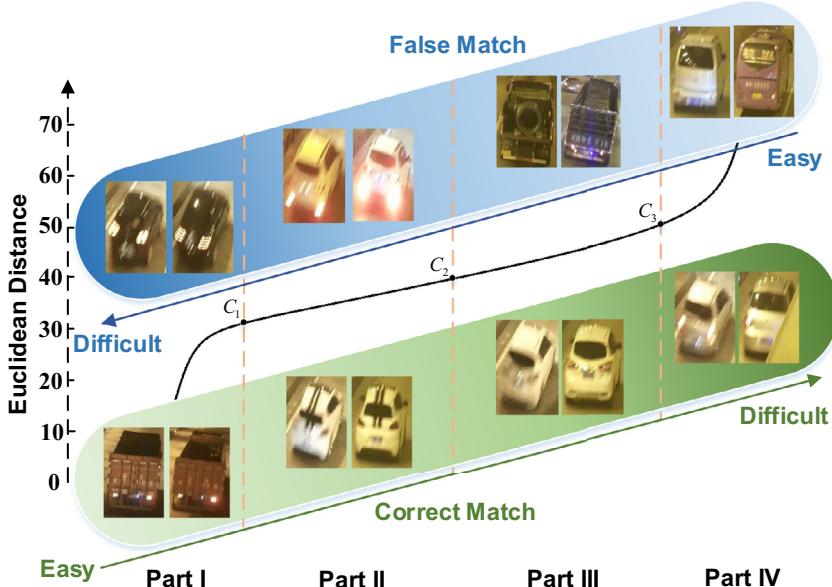


Fig. 7. Illustration of tunnel vehicle Re-ID from easy to difficult. As Euclidean distance increases, the difficulty level of correct match changes from easy to difficult, while false match is opposite.

intra-variation is, the more difficult picking out the correct match is. On the contrary, the smaller the inter-variation is, the more difficult distinguishing the false match is. In this paper, we divide the samples of tunnel vehicle Re-ID into four difficulty levels according to the sorted Euclidean distance of vehicle features from different cameras, as shown in Fig. 7. With Euclidean distance increasing, the difficulty curve of samples appear two apparent inflection points. Training samples are divided into four parts as follows:

$$\text{correct match (label=1)} \cdot \begin{cases} \text{if } d \leq C_1 , x_i \in \text{Part I} \\ \text{if } C_1 < d \leq C_2 , x_i \in \text{Part II} \\ \text{if } C_2 < d \leq C_3 , x_i \in \text{Part III} \\ \text{if } C_3 < d , x_i \in \text{Part IV} \end{cases} \quad (7)$$

$$\text{false match (label=0)} \cdot \begin{cases} \text{if } C_3 < d , x_i \in \text{Part I} \\ \text{if } C_2 < d \leq C_3 , x_i \in \text{Part II} \\ \text{if } C_1 < d \leq C_2 , x_i \in \text{Part III} \\ \text{if } d \leq C_1 , x_i \in \text{Part IV} \end{cases} \quad (8)$$

where d is the Euclidean distance of feature vectors of vehicle image pairs, and C_1 , C_2 , C_3 are the partitioning thresholds of four difficulty levels of training samples. Observing the curve of Euclidean distance of all samples, there are two obvious inflection points near $d = 30$ and $d = 50$. Large amounts of Euclidean distance fall into range of [30,50], and we divide it into two parts by the middle value 40. Therefore, the partitioning thresholds $[C_1, C_2, C_3]$ of difficulty levels are set as [30,40,50] to divide training samples into four parts according to the distribution of Euclidean distance. The

Table 1

The obtained SCF levels, training time and accuracy when setting terminal threshold T as 1, 3, and 5.

T	1	3	5
Levels	1–5	12–20	15–30
Training Time (mins)	5–10	20–40	25–70
Training Accuracy (%)	70–75	80–85	82–86

percentages of samples of four difficulty levels are 12.62%, 26.90%, 39.07%, and 21.41%, respectively.

During training process of SCF with curriculum learning (SCFCL), we firstly feed samples of Part I to SCF model. Then we feed samples of Part II to training process until there is no significant performance gain of Part I. For the samples of previous step, class vectors of the optimal level are adopted to augmented input of next step. In the same way, the samples of four parts are feed into training process in sequence.

5. Experiments and discussions

5.1. Implementation details

5.1.1. Training procedure

Structural parameters of the proposed SCF model have great effect on training time and testing accuracy, which mainly depends on the number of trees in each forest and the fold of cross-validation. Through extensive experiments on Tunnel-VReID dataset, we select the two important parameters as 50 and 20. Detailed analysis of parameter selection is described in Section 5.2. Furthermore, it is worth noting that the terminal threshold T of SCF growing cannot be too large, otherwise SCF model will generate too many levels and spend much training time without significant performance gain. We have checked this influence by setting T as 1, 3 and 5, and observe that $T = 3$ is reasonable, as shown in Table 1.

In addition, there is an important factor on the proportion of positive and negative samples in training process. The number of positive samples equals to the number of image pairs with same identities for training process, while there are too many negative samples produced by alternant combination of different identities, which far exceeds the number of positive samples. If all negative samples are put into training, it will lead that almost all inputs will be classified as negative. To be balancing, we finally choose a suitable 1:2 ratio of positive and negative samples for training through adequate test validation.

5.1.2. Evaluation procedure

Re-ID results are commonly evaluated by Rank-1, 5, 10, 20 accuracy and the average cumulative match characteristic (CMC) curve. Rank- n and CMC curve are calculated according to the probability of correct match in final prediction, where Rank- n represents the expectation of finding the correct match from the top n matches. The CMC curve represents the chance of the correct match appearing in the top 1, 2, ..., n of the ranked list, and every point on the CMC curve is corresponding to Rank-1, 2, ..., n , respectively. All evaluation processes are repeated for 10 times to get an average performance as the final result.

5.2. Parameter analysis

In the proposed SCF model, there are two key parameters, which are the number of trees in each forest and the fold of cross validation, having the important impact on matching accuracy and training time.

Table 2

Matching rates (in percentage) of different methods on our Tunnel-VReID dataset. The best one is highlighted in bold fonts and the second one is marked by italic fonts.

Method	Rank-1	Rank-5	Rank-10
gcForest [40]	32.5	61.4	74.1
XQDA [25]	40.0	71.1	81.5
MLAPG [28]	41.6	71.9	83.1
VGG+Triplet Loss [32]	43.6	69.6	77.8
CNN Embedding [54]	45.8	71.1	79.2
PCB [55]	48.9	73.2	80.8
SCF+ResNet-152	13.0	32.1	47.7
SCF+VGG-19	19.4	40.6	55.5
SCF+VGG-F	23.4	48.2	61.6
SCF+LOMO	53.8	66.2	75.1
SCFCL+LOMO	55.1	68.1	76.5
SCF+LOMO+VGG-F	57.6	72.8	81.1
SCFCL+LOMO+VGG-F	61.4	<i>77.8</i>	84.8

5.2.1. Influence of the number of trees in each forest

Each forest is constructed by multiple decision trees, and each decision tree will generate a predicted value. By voting, classification result is determined by the decision results of majority of trees. Therefore, the number of trees n_{Trees} in each forest is crucial to final result. To evaluate the influence of n_{Trees} , we conduct experiments on tunnel vehicle Re-ID dataset by initializing the number of decision trees as [10,20,50,100,150,200] with three different folds [10,20,30] of cross-validation. In Fig. 8 (a), we can observe that the matching rate of rank-1 have the highest and the robustest performance when $n_{\text{Trees}} = 50$. When n_{Trees} is greater than 50, rank-1 does not ascend any longer and fluctuates severely under different folds. Besides, computational cost for training will increase drastically with the increasing of n_{Trees} . Therefore, we finally choose $n_{\text{Trees}} = 50$ as default setting.

5.2.2. Influence of the fold of cross-validation

In order to achieve a reliable and stable model, we adopt m -fold cross-validation to get useful information as much as possible from limited data, and avoid overfitting to a certain extent. We test varying values of m -fold of cross-validation by the list [1,5,10,20,30,40,50] while fixing n_{Trees} as 50. Fig. 8 (b) plots the rank-1 matching rate w.r.t. the value of m -fold. The value of m -fold also have strong impact on training time. Through extensive experiments, we choose $m = 20$ to balance performance and training time.

5.3. Performance comparison on Tunnel-VReID Dataset

We evaluate our methods on Tunnel-VReID dataset in comparison with several state-of-the-art methods. Fig. 9 shows the CMC curves of various methods, and Table 2 shows the rank-1, 5, and 10 values.

From the results, we can observe that the proposed method manifestly outperforms the other methods, and SCFCL+LOMO+VGG-F achieves the highest rank-1 score of 61.4% among the competitors. The proposed SCF model is inspired by gcForest, but SCF+LOMO improves the accuracy with a dramatic increase of 21.3% at rank-1, which proves effectiveness of the average strategy of SCF model. XQDA [25] and MLAPG [28] combined with LOMO feature are outstanding hand-crafted metric learning methods. Similarly, SCF+LOMO clearly outperforms XQDA and MLAPG with an improvement of 13.8% and 12.2%, respectively. VGG+Triplet Loss [32], CNN Embedding [54] and PCB [55] are deep learning frameworks. SCF+LOMO, with hand-crafted feature, achieves relative increase of 5–10% compared with these three methods. That is because of the challenging vehicle images in tunnel and the limited number of samples. However, deep learning

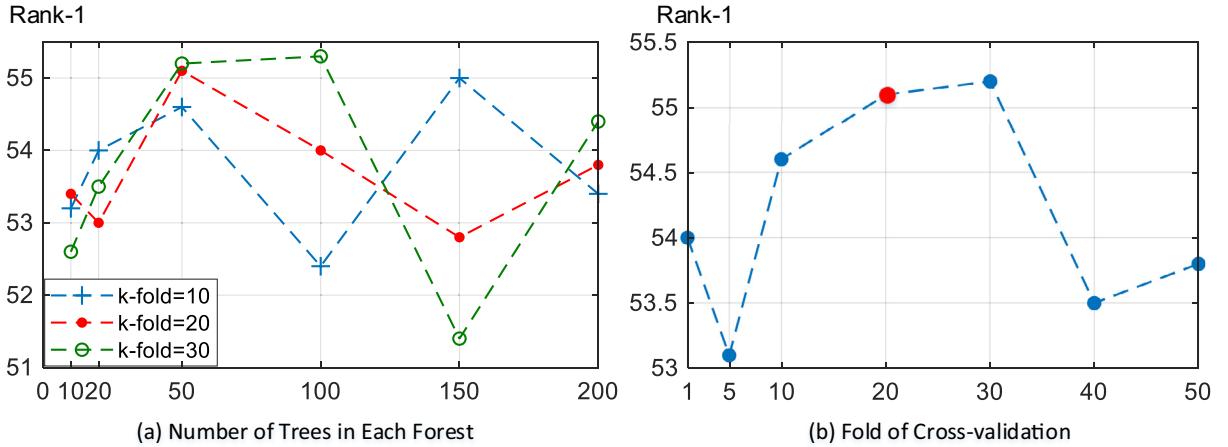


Fig. 8. (a) Matching rate of Rank-1 with increasing number of trees in each forest. The fold of cross-validation is taken into account with three different values [10,20,30], marked by blue, red and green lines, respectively. (b) Matching rate of Rank-1 with increasing fold of cross-validation. The performance of default setting is marked by red color.

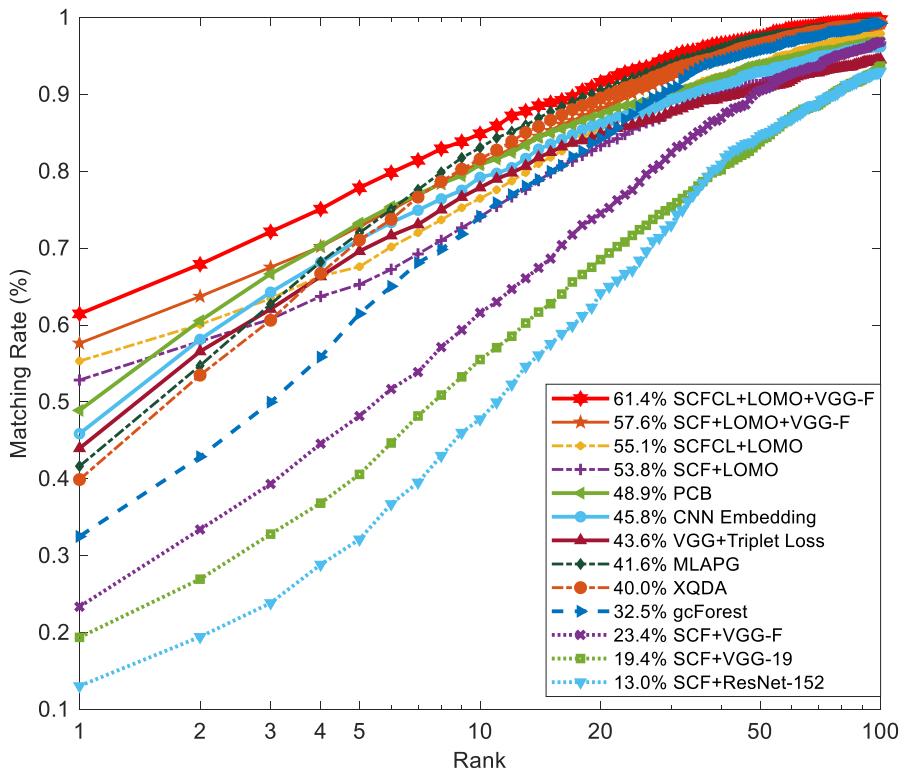


Fig. 9. Performance comparison with state-of-the-art methods using CMC curves on the Tunnel-VReID dataset.

requires a large amount of labeled data to facilitate the estimation of model parameters.

Compared with SCF+LOMO, SCF+LOMO+VGG-F improves the accuracy with an increase of 3.8%. Interestingly, only CNN features have poor performance obtained by SCF+ResNet-152/VGG-19/VGG-F. However, CNN feature still has a positive impact, we choose to concatenate preferable VGG-F feature with LOMO as final feature. It can be found that the accuracy of SCFCL+LOMO+VGG-F is 6.3% higher than that of SCFCL+LOMO at rank-1. Furthermore, SCFCL+LOMO surpasses SCF+LOMO by 3.1%, and SCFCL+LOMO+VGG-F surpasses SCF+LOMO+VGG-F by 2.4%. These two contrast experiments reveal that curriculum learning introduced to SCF can enhance model classification ability.

5.4. Performance comparison on VehicleID Dataset

In addition to the constructed Tunnel-VReID dataset, we also evaluate the performance on the public VehicleID [5] dataset collected during daytime in open roads scenes. Different from our Tunnel-VReID, it is easy to distinguish color, local details and shape of vehicles of VehicleID. Besides, the license plate information in VehicleID is blacked out. There are 221,567 images of 26,328 vehicles in total (8.42 images/vehicle in average) in VehicleID dataset. The vehicle image is also resized to 300×200 , then the LOMO feature is extracted by subwindow size of 20×20 with overlapping step of 15 pixels sampling step. During training, the same SCF settings $nTrees = 50$ and $m\text{-fold}=20$ are employed to

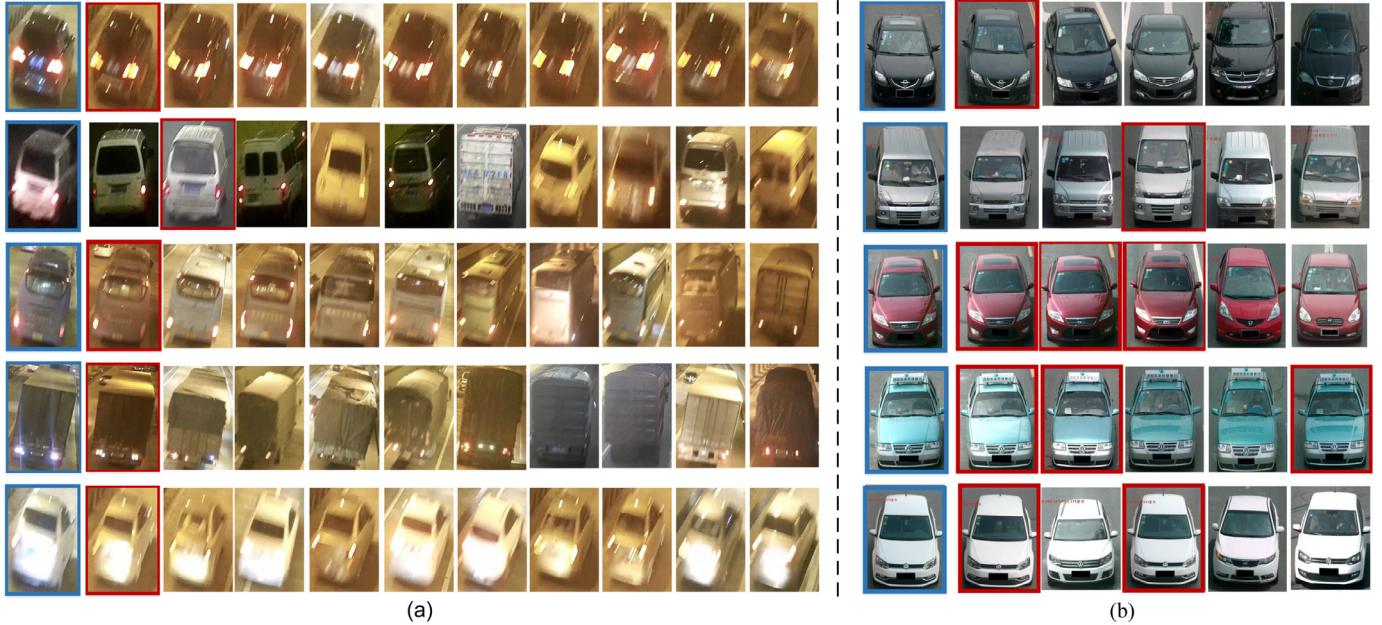


Fig. 10. Examples of Re-ID results by our model on the Tunnel-VReID and VehicleID datasets. For each dataset, each row represents a ranking result with the first image being the query and the rest images being the returned list. The query images and its correct match images are respectively marked by blue and red rectangle boxes. (a) Tunnel-VReID dataset. (b) VehicleID dataset.

Table 3

Matching rates (in percentage) of different methods on our VehicleID dataset. The best one is highlighted in bold fonts and the second one is marked by italic fonts.

Method	Rank-1	Rank-5	Rank-20
XQDA [25]	15.3	25.3	36.0
MLAPG [28]	18.1	28.2	43.4
gcForest [40]	27.7	34.5	43.5
VGG+Triplet Loss [32]	31.9	50.3	57.8
VGG+CCL [5]	35.6	56.2	68.4
CNN Embedding [54]	37.3	57.8	70.2
FACT [20]	39.9	60.3	72.9
PCB [55]	44.8	59.5	64.3
VAMI [15]	47.3	70.3	79.9
SCF+VGG-F	33.1	46.8	55.0
SCF+LOMO	43.3	60.7	71.1
SCFCL+LOMO	46.4	64.6	76.1
SCF+LOMO+VGG-F	50.7	67.0	77.8
SCFCL+LOMO+VGG-F	53.1	69.7	81.5

VehicleID. Besides, the partitioning thresholds $[C_1, C_2, C_3]$ are set as [35,40,45] for SCFCL. As the same difficulty partitioning criterion to Tunnel-VReID, the percentages of training samples of four difficulty levels are 19.79%, 19.25%, 27.62%, and 33.34%, respectively.

On VehicleID dataset, we compare the performance of the proposed method against several state-of-the-art algorithms. We utilize the same sample partitioning protocol with [15], where 13,134 vehicles are used for training and 2400 vehicles are used for testing. As listed in Table 3, the results show that SCFCL+LOMO+VGG-F achieves the highest rank-1 accuracy of 53.1%. Compared with Tunnel-VReID, XQDA and MLAPG based on hand-crafted metric learning have serious performance degradation on VehicleID, while the performance of gcForest is relatively stable. VGG+Triplet Loss, CNN Embedding and PCB maintain a good performance. In addition, VGG+CCL, FACT and VAMI based on deep learning also achieve prominent performance, especially for VAMI. Owning to multi-level ensemble structure, the performance of our SCF+LOMO+VGG-F on VehicleID dataset is quite outstanding. VGG-F feature concatenated with LOMO obtains nearly 5% promotion. SCFCL employing curriculum learning achieves about 3% improvement over SCF at rank-1.

In Fig. 10 (a) and (b), we show five typical query results of the Tunnel-VReID and VehicleID datasets by our model, respectively. The query images of vehicles are quite similar to the ranking results. Specially, there are some indiscernible vehicles with extremely similar appearance in Tunnel-VReID dataset, which is extremely difficult to recognize the correct match, even by human eyes. Comparatively speaking, the image quality of VehicleID dataset is further better than that of Tunnel-VReID. Distinct color and texture of vehicle images in VehicleID dataset make it relatively easy to find correct match. It is worth noting that the number of vehicle images of VehicleID dataset is 100 times more than that of Tunnel-VReID. Massive data volume can improve the capacity of deep learning models, while the proposed SCF model still achieves good performance on VehicleID dataset.

5.5. Discussions

5.5.1. SCF versus GcForest

The proposed SCF is inspired by gcForest, and it significantly improves the rank-1 Re-ID rate with an increase of nearly 30%. For the behind meaning, we compare these two models in training and testing process, and find a credible interpretation.

In the cascade structure of gcForest and SCF, each level receives the prediction information from its preceding level, and transmits its processed result to the next level. This operation of feature augmentation can boost feature enrichment, and guide the model to obtain better classification capacity in deeper levels. For interpreting the different power of classification, we adopt t-SNE method [56] to visualize the feature distribution of different levels in training process on our Tunnel-VReID dataset, as shown in Fig. 11. Generally, the feature distributions between positive and negative samples are getting separable in gcForest and our SCF model, which is in consistent with the ascending classification accuracy of model. However, our SCF has more powerful ability for separating the positive and negative samples than original gcForest. Especially, with the increasing of layers, gcForest seems powerless. The behind meaning is that our averaging operation between different layers in SCF model acts as a buffering role for hard samples, and takes them into the following layers for further determination. On

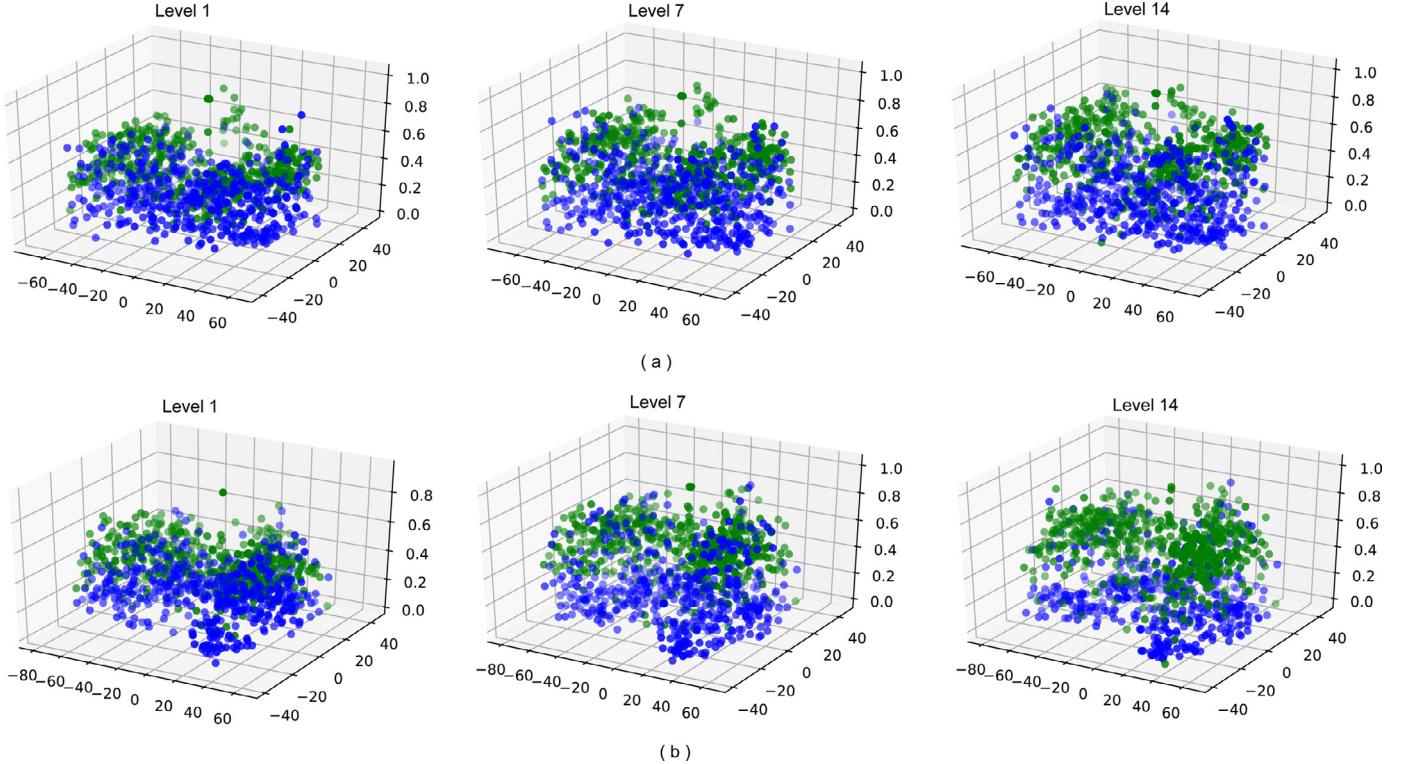


Fig. 11. Multi-level feature distribution visualization of gcForest (a) and SCF (b) in training process. Intuitively, inter-class separability is getting greater as the level becomes deeper. Green points denote the positive samples, while blue points denote the negative samples.

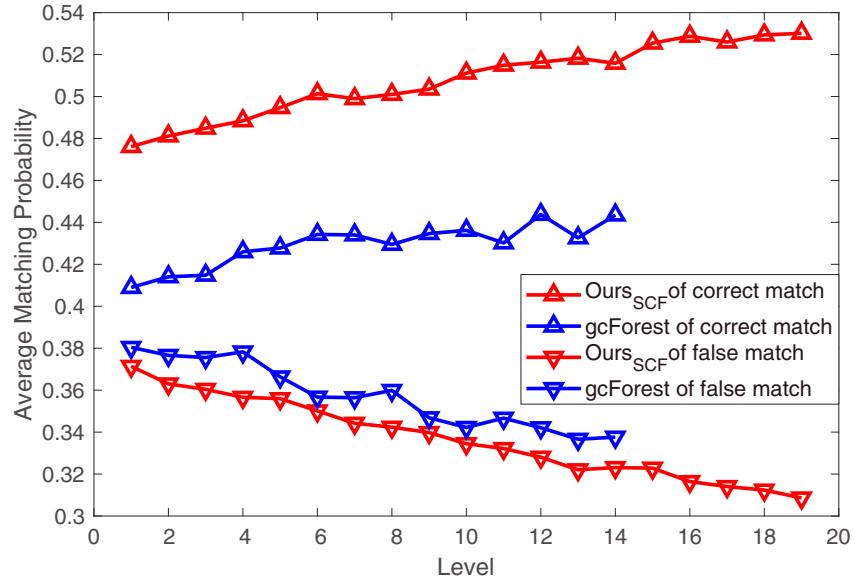


Fig. 12. Average matching probability change trend of testing with multi-level SCF and gcForest. The red line represents the probability change of SCF, while the blue line represents the probability change of gcForest. The line marked by upper triangular plots the probability of correct match, whereas the line marked by lower triangular plots the probability of false match.

the contrary, gcForest may get into confused. Obviously, feature separability of SCF is better than gcForest as the level becomes deeper.

In addition, in the testing stage, predicted class vectors of the current level in gcForest are directly inputted as augmented features of the next level, so the fluctuation amplitude of prediction accuracy is large between the two adjacent levels. We retrofit it by synergistic cascade structure that adopts the averaging bridge of alternative layers to serve as the augmented feature inputting

to next level. This average strategy promotes a steady increase in classification capacity with deep levels. To demonstrate the effectiveness, we randomly and respectively choose 100 test samples to compare the averaged change trend of match probability of SCF and gcForest, i.e., samples of correct match and false match for half of each. The tendency curve comparison is shown in Fig. 12.

We can observe that the average matching probability of correct match is upward while false match is downward. However, change trend of Ours_{SCF} is more smooth than gcForest. On the other hand,

the average matching probability curve of correct match of Ours_{SCF} is higher than gcForest, while false match of Ours_{SCF} is lower than gcForest. This also demonstrates that the ability of Ours_{SCF} is more powerful to distinguish the identity of two vehicles. Moreover, the gradual change trend of SCF model also give rise to about 5–10 more levels than gcForest model, and each level has two layers in SCF model. Multi-level structure yields desirable classification capacity. Therefore, the strategy of synergistic cascade structure can smooth the sudden jump of the two adjacent levels, it not only avoids performance declining, but also makes it possible to distinguish different identities of vehicles with little inter-variation and recognize same identities of vehicles with large intra-variation with the increasing of the layers.

5.5.2. SCF versus deep learning methods

Generally, the performances of deep learning methods are favorable on the large-scale labeled dataset. That is because a large number of labeled data for training can effectively facilitate the estimation of model parameters. However, it is very difficult and time-consuming to collect large-scale and high-quality labeled data. In contrast, the number of parameters in our SCF model is relatively small, but it still can work well on small size of dataset. Furthermore, the performance of CNN feature without end-to-end training is weaker than LOMO, especially for our challenging tunnel vehicle images. The proposed SCF model is mainly affected by two parameters, the number of trees in each forest and the fold of cross-validation. When the same settings are applied to different datasets across different domains, promising performance still can be achieved.

6. Conclusion

In this paper, we attempted to address the tunnel vehicle Re-ID problem by computer vision technique. Different from the most of the vehicle Re-ID methods, we aimed to explore the deep correlation between-vehicle samples in tunnel scenes, and construct the tunnel vehicle Re-ID model by synergistically cascade forests (SCF). In the construction process of SCF, the training pairs of tunnel vehicle samples were gradually fed into alternative layers of random forest and extremely randomized forest, and the linking relation between vehicle samples are gradually modeled by the forest ensembles. In addition, we pave an easy-to-difficult learning framework by embedding curriculum learning into the SCF construction. We evaluated the effectiveness of the proposed method by a new tunnel vehicle Re-ID dataset (Tunnel-VReID) laboriously labeled by ourselves. Besides, we also checked the adaptation ability of our model on a large-scale vehicle Re-ID dataset, i.e., VehicleID, collected in open road scenes, and the superiority is also proved apparently.

Declaration of Competing Interest

In order to make a fair peer-review, the domain “chd.edu.cn” is our conflict of interest.

Acknowledgments

This work was supported by the National Nature Science Foundation of China under Grant Nos. 61751308, 61603057, 61773311, 51578074, 61773316, U1864204, Project of Special Zone for National Defense Science and Technology Innovation, and China Postdoctoral Science Foundation under Grant 2017M613152.

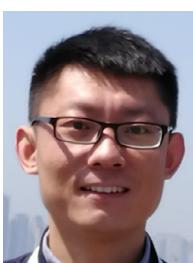
Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neucom.2019.11.069.

References

- [1] Z. Ma, C. Shao, S. Zhang, Characteristics of traffic accidents in chinese freeway tunnels, *Tunnell. Underground Space Technol.* 24 (3) (2009) 350–355.
- [2] J.S. Yeung, Y.D. Wong, Road traffic accidents in singapore expressway tunnels, *Tunnell. Underground Space Technol.* 38 (2013) 534–541.
- [3] X. Liu, W. Liu, T. Mei, H. Ma, PROVID: progressive and multimodal vehicle re-identification for large-scale urban surveillance, *IEEE Trans. Multimedia* 20 (3) (2018) 645–658.
- [4] X. Li, M. Yuan, Q. Jiang, G. Li, VRID-1: a basic vehicle re-identification dataset for similar vehicles, in: Proc. IEEE International Conference on Intelligent Transportation Systems, 2017, pp. 1–8.
- [5] H. Liu, Y. Tian, Y. Wang, L. Pang, T. Huang, Deep relative distance learning: tell the difference between similar vehicles, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2167–2175.
- [6] X. Zhou, H.S. Mahmassani, Dynamic origin-destination demand estimation using automatic vehicle identification data, *IEEE Trans. Intell. Transp. Syst.* 7 (1) (2006) 105–114.
- [7] C. Oh, S.G. Ritchie, S. Jeng, Anonymous vehicle reidentification using heterogeneous detection systems, *IEEE Trans. Intell. Transp. Syst.* 8 (3) (2007) 460–469.
- [8] S. Jeng, Y.C.A. Tok, S.G. Ritchie, Freeway corridor performance measurement based on vehicle reidentification, *IEEE Trans. Intell. Transp. Syst.* 11 (3) (2010) 639–646.
- [9] S. Charbonnier, A.C. Pitton, A. Vassilev, Vehicle re-identification with a single magnetic sensor, in: Proc. Instrumentation and Measurement Technology Conference, 2012, pp. 380–385.
- [10] C. Anagnostopoulos, I. Anagnostopoulos, I.D. Psoroulas, V. Loumos, E. Kayafas, License plate recognition from still images and video sequences: a survey, *IEEE Trans. Intell. Transp. Syst.* 9 (3) (2008) 377–391.
- [11] Y. Wen, Y. Lu, J. Yan, Z. Zhou, K.M. von Deneen, P. Shi, An algorithm for license plate recognition applied to intelligent transportation system, *IEEE Trans. Intell. Transp. Syst.* 12 (3) (2011) 830–845.
- [12] C. Gou, K. Wang, Y. Yao, Z. Li, Vehicle license plate recognition based on extremal regions and restricted boltzmann machines, *IEEE Trans. Intell. Transp. Syst.* 17 (4) (2016) 1096–1107.
- [13] Z. Wang, L. Tang, X. Liu, Z. Yao, S. Yi, J. Shao, J. Yan, S. Wang, H. Li, X. Wang, Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification, in: Proc. IEEE International Conference on Computer Vision, 2017, pp. 379–387.
- [14] Y. Zhang, D. Liu, Z.-J. Zha, Improving triplet-wise training of convolutional neural network for vehicle re-identification, in: Proc. IEEE International Conference on Multimedia and Expo, 2017, pp. 1386–1391.
- [15] Y. Zhou, L. Shao, Aware attentive multi-view inference for vehicle re-identification, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6489–6498.
- [16] X. Liu, W. Liu, H. Ma, H. Fu, Large-scale vehicle re-identification in urban surveillance videos, in: Proc. IEEE International Conference on Multimedia and Expo, 2016, pp. 1–6.
- [17] Y. Tang, D. Wu, Z. Jin, W. Zou, X. Li, Multi-modal metric learning for vehicle re-identification in traffic surveillance environment, in: Proc. IEEE International Conference on Image Processing, 2017, pp. 2254–2258.
- [18] Y. Bai, Y. Lou, F. Gao, S. Wang, Y. Wu, L.-Y. Duan, Group-sensitive triplet embedding for vehicle reidentification, *IEEE Trans. Multimedia* 20 (9) (2018) 2385–2399.
- [19] J. Peng, H. Wang, T. Zhao, X. Fu, Learning multi-region features for vehicle re-identification with context-based ranking method, *Neurocomputing*, (2019), doi:10.1016/j.neucom.2019.06.013.
- [20] X. Liu, W. Liu, T. Mei, H. Ma, A deep learning-based approach to progressive vehicle re-identification for urban surveillance, in: Proc. European Conference on Computer Vision, 2016, pp. 869–884.
- [21] Y. Shen, T. Xiao, H. Li, S. Yi, X. Wang, Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals, in: Proc. IEEE International Conference on Computer Vision, 2017, pp. 1918–1927.
- [22] A. Frias-Velazquez, P.V. Hese, A. Pizurica, W. Philips, Split-and-match: a bayesian framework for vehicle re-identification in road tunnels, *Eng. Appl. Artif. Intell.* 45 (2015) 220–233.
- [23] L. An, X. Chen, S. Yang, Multi-graph feature level fusion for person re-identification, *Neurocomputing* 259 (2017) 39–45.
- [24] R.R. Varior, G. Wang, J. Lu, T. Liu, Learning invariant color features for person reidentification, *IEEE Trans. Image Process.* 25 (7) (2016) 3395–3410.
- [25] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2197–2206.
- [26] S. Khamis, C. Kuo, V.K. Singh, V.D. Shet, L.S. Davis, Joint learning for attribute-consistent person re-identification, in: Proc. European Conference on Computer Vision, 2014, pp. 134–146.
- [27] W. Ouyang, R. Zhao, X. Wang, Learning mid-level filters for person re-identification, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 144–151.
- [28] S. Liao, S.Z. Li, Efficient PSD constrained asymmetric metric learning for person re-identification, in: Proc. IEEE International Conference on Computer Vision, 2015, pp. 3685–3693.
- [29] Y. Yuan, J. Zhang, Q. Wang, Bike-person re-identification: a benchmark and a comprehensive evaluation, *IEEE Access* 6 (2018) 56059–56068.

- [30] W. Chen, X. Chen, J. Zhang, K. Huang, Beyond triplet loss: a deep quadruplet network for person re-identification, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1320–1329.
- [31] Y. Wang, Z. Chen, F. Wu, G. Wang, Person re-identification with cascaded pairwise convolutions, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1470–1478.
- [32] S. Ding, L. Lin, G. Wang, H. Chao, Deep feature learning with relative distance comparison for person re-identification, *Pattern Recognit.* 48 (10) (2015) 2993–3003.
- [33] F. Wang, W. Zuo, L. Lin, D. Zhang, L. Zhang, Joint learning of single-image and cross-image representations for person re-identification, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1288–1296.
- [34] S. Bai, X. Bai, Q. Tian, Scalable person re-identification on supervised smoothed manifold, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3356–3365.
- [35] S.-U. Rehman, Z. Chen, M. Raza, P. Wang, Q. Zhang, Person re-identification post-rank optimization via hypergraph-based learning, *Neurocomputing* 287 (2018) 143–153.
- [36] A. Barman, S.K. Shah, Shape: a novel graph theoretic algorithm for making consensus-based decisions in person re-identification systems, in: Proc. IEEE International Conference on Computer Vision, 2017, pp. 1124–1133.
- [37] R. Zhao, W. Ouyang, X. Wang, Person re-identification by salience matching, in: Proc. IEEE International Conference on Computer Vision, 2013, pp. 2528–2535.
- [38] L. Bazzani, M. Cristani, V. Murino, Symmetry-driven accumulation of local features for human characterization and re-identification, *Comput. Vision Image Understanding* 117 (2) (2013) 130–144.
- [39] Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by GAN improve the person re-identification baseline in vitro, in: Proc. IEEE International Conference on Computer Vision, 2017, pp. 3774–3782.
- [40] Z. Zhou, J. Feng, Deep forest: towards an alternative to deep neural networks, in: Proc. International Joint Conference on Artificial Intelligence, 2017, pp. 3553–3559.
- [41] X. Ren, C. Long, L. Zhang, Y. Wei, D. Du, J. Liang, W. Li, S. Zhang, A dynamic boosted ensemble learning based on random forest, 2018 arXiv preprint arXiv:1804.07270.
- [42] R. Genier, J.-M. Poggi, C. Tuleau-Malot, Variable selection using random forests, *Pattern Recognit. Lett.* 31 (14) (2010) 2225–2236.
- [43] A. Behnamian, K. Millard, S.N. Banks, L. White, M. Richardson, J. Pasher, A systematic approach for variable selection with random forests: achieving stable variable importance values, *IEEE Geosci. Remote Sens. Lett.* 14 (11) (2017) 1988–1992.
- [44] D.J. Jobson, Z. Rahman, G.A. Woodell, A multiscale retinex for bridging the gap between color images and the human observation of scenes, *IEEE Trans. Image Process.* 6 (7) (1997) 965–976.
- [45] J. Yu, C. Zhu, J. Zhang, Q. Huang, D. Tao, Spatial pyramid-enhanced netvlad with weighted triplet loss for place recognition, *IEEE Trans. Neural Netw. Learn. Syst.* To Be Published, (2019), doi:10.1109/TNNLS.2019.2908982.
- [46] J. Yu, Y. Rui, Y.Y. Tang, D. Tao, High-order distance-based multiview stochastic learning in image classification, *IEEE Trans. Cybern.* 44 (12) (2014) 2431–2442.
- [47] Q. Wang, J. Wan, F. Nie, B. Liu, C. Yan, X. Li, Hierarchical feature selection for random projection, *IEEE Trans. Neural Netw. Learn. Syst.* 30 (5) (2019) 1581–1586.
- [48] L. Breiman, Random forests, *Machine Learn.* 45 (1) (2001) 5–32.
- [49] Bosch, Anna, Zisserman, Andrew, X. Munoz, Image classification using random forests and ferns, in: Proc. IEEE International Conference on Computer Vision, 2007, pp. 1–8.
- [50] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, *Machine Learn.* 63 (1) (2006) 3–42.
- [51] Z. Zhou, *Ensemble Methods: Foundations and Algorithms*, Chapman and Hall/CRC, 2012.
- [52] Y. Bengio, J. Louradour, R. Collobert, J. Weston, Curriculum learning, in: Proc. International Conference on Machine Learning, 2009, pp. 41–48.
- [53] J. Lu, D. Meng, Q. Zhao, S. Shan, A.G. Hauptmann, Self-paced curriculum learning, in: Proc. AAAI Conference on Artificial Intelligence, 2015, pp. 2694–2700.
- [54] Z. Zheng, L. Zheng, Y. Yang, A discriminatively learned CNN embedding for person re-identification, *ACM Trans. Multimedia Comput. Commun. Appl. (TOMM)* 14 (1) (2018) 13:1–13:20.
- [55] Y. Sun, L. Zheng, Y. Yang, Q. Tian, S. Wang, Beyond part models: Person retrieval with refined part pooling (and A strong convolutional baseline), in: Proc. European Conference on Computer Vision, 2018, pp. 501–518.
- [56] L. Maaten, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* 9 (11) (2008) 2579–2605.



Rixing Zhu received the B.E. degree from the School of Electronic and Control Engineering, Chang'an University, China, in 2014. He is currently a Ph.D. candidate from the School of Electronic and Control Engineering, Chang'an University. His research interests include computer vision and traffic security.



Jianwu Fang received the Ph.D. degree in SIP (signal and information processing) from University of Chinese Academy of Sciences, China in 2015. He is currently an Associate Professor with Laboratory of Traffic Vision Safety (LOTVS) in the School of Electronic and Control Engineering, Chang'an University, Xi'an, China, and is also a postdoctoral researcher in the Institute of Artificial Intelligence and Robotics (IAIR), Xi'an Jiaotong University, Xi'an, China. He has published many papers on top-ranked journals and conferences, such as IEEE-TCYB, IEEE-TIE, IEEE-TCSV, AAAI, ICRA, etc. His research interests include computer vision and pattern recognition.



Shuying Li received the B.E. degree of Automation from University of Science and Technology of China in 2005, and the Ph.D. degree of Digital Circuit and System from Chinese Academy of Sciences in 2010. She is currently a professor with School of Automation, Xi'an University of Posts & Telecommunications, Xi'an, China. Her research interest includes remote sensing, computer vision and pattern recognition.



Qi Wang received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Computer Science, with the Unmanned System Research Institute, and with the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xian, China. His research interests include computer vision and pattern recognition.



Hongke Xu received his PhD degree from Chang'an University, China, in 2006. He is currently a full Professor of School of Electronic and Control Engineering, Chang'an University, Xi'an, China. His research interests include signal processing and traffic intelligent control engineering.



Jianru Xue received the masters and Ph.D. degrees from Xian Jiaotong University (XJTU), Xian, China, in 1999 and 2003, respectively. He was with FujiXerox, Tokyo, Japan, from 2002 to 2003, and visited the University of California at Los Angeles, Los Angeles, CA, USA, from 2008 to 2009. He is currently a Professor with the Institute of Artificial Intelligence and Robotics at XJTU. His research field includes computer vision, visual navigation, and scene understanding for autonomous system. Prof. Xue served as a Co-Organization Chair of the 2009 Asian Conference on Computer Vision and 2006 Virtual System and Multi-media conference. He also served as a PC Member of the 2012 Pattern Recognition conference, and the 2010 and 2012 Asian conference on Computer Vision.



Hongkai Yu received the Ph.D. degree in computer science and engineering from University of South Carolina, Columbia, SC, USA in 2018. He then joins the Department of Computer Science at University of Texas CRIO Grande Valley, Edinburg, TX, USA as an assistant professor. His research interests include computer vision, machine learning, deep learning and intelligent transportation system.