

Vehicle Face Re-identification Based on Nonnegative Matrix Factorization with Time Difference Constraint

Na Ma*, Tingxin Wen

School of Business Administration, Liaoning Technical University
Huludao, Liaoning Province - China
[e-mail: vicky070708@163.com]
*Corresponding author: Na Ma

*Received January 19, 2021; revised March 31, 2021; revised April 25, 2021; revised May 20, 2021;
accepted May 31, 2021; published June 30, 2021*

Abstract

Light intensity variation is one of the key factors which affect the accuracy of vehicle face re-identification, so in order to improve the robustness of vehicle face features to light intensity variation, a Nonnegative Matrix Factorization model with the constraint of image acquisition time difference is proposed. First, the original features vectors of all pairs of positive samples which are used for training are placed in two original feature matrices respectively, where the same columns of the two matrices represent the same vehicle; Then, the new features obtained after decomposition are divided into stable and variable features proportionally, where the constraints of intra-class similarity and inter-class difference are imposed on the stable feature, and the constraint of image acquisition time difference is imposed on the variable feature; At last, vehicle face matching is achieved through calculating the cosine distance of stable features. Experimental results show that the average False Reject Rate and the average False Accept Rate of the proposed algorithm can be reduced to 0.14 and 0.11 respectively on five different datasets, and even sometimes under the large difference of light intensities, the vehicle face image can be still recognized accurately, which verifies that the extracted features have good robustness to light variation.

Keywords: Vehicle face re-identification; Nonnegative Matrix Factorization; stable feature; variable feature; cosine distance

This research was supported in part by the National Natural Science Foundation of China (71371091), the Social Science Foundation of Liaoning Education Department (L14BTJ004).

1. Introduction

Traditionally, the vehicles with fake plates are mainly detected by viewing videos manually, however, with the rapid increase in the numbers of vehicles, the manual detection methods will face great difficulties due to the massive amounts of video data. Therefore, it is of great significance to propose an automatic and effective vehicle re-identification algorithm. For vehicle re-identification, the biggest challenge is that affected by the difference of light intensities, there may be large differences between the captured images representing the same vehicle as shown in Fig. 1, which brings great difficulties to re-identification, so it can be concluded that it is very critical to obtain stable and effective vehicle face features under various lighting conditions [1].



(a) Acquisition time: 10:27 (b) Acquisition time: 18:09

Fig. 1. The captured images of the same vehicle at different times

The remainder of this paper is organized as follows. In section 2, some related works are addressed. The proposed NMF model for vehicle face recognition is proposed in section 3. In section 4, a projected gradient algorithm is used to solve the objective function of the proposed NMF model, and the matching method of vehicle face features is given in section 5. In section 6, the proposed algorithm is proved effectively through experiments. Finally, the conclusion is drawn in section 7.

2. Related Work

Nowadays, the existing vehicle recognition methods mainly include the following two categories, one is based on artificial extracted features, the other is based on the features which are obtained automatically through deep learning. The artificial features can be divided into the low-level features and the high-level features, where the low-level image features include color feature [2-3], edge feature [4], texture feature [5], and shape feature [6], et al; and scale key point features [7-8] and 3D model feature [9-12] can be considered as the high-level features. However, the artificial features depend on human experience to a large extent, and the deep information of image is not easy to be mined, so the effectiveness of artificial features is hard to be ensured. Therefore, the deep learning based vehicle recognition algorithms are paid more attention in recent years, which include some traditional deep learning models such as Convolutional Neural Network model [13-15], Deep Belief Network model [16-17], Transfer learning model [18-20], Restricted Boltzmann Machine [21-23], and some improved models such as Conv5 [24], Teacher-Student Network [25], Parsing-based View-aware Embedding Network [26], Semantics-guided Part Attention Network [27], the model fused by multiple networks [28], and the network based on reconstruction [29], et al. For the supervised

vehicle classification problem, these deep learning methods have achieved good results, but for vehicle face matching problem under the conditions that the times of each vehicle being captured is very limited and the number of the training samples is too small, the universalities of these models are not very well. Therefore, under a limited number of vehicle face samples, it is very meaningful to propose a vehicle re-identification algorithm with good robustness and universality.

Nonnegative Matrix Factorization (NMF) can obtain effective basis feature images for image classification, and it has also achieved good results in vehicle face recognition in recent years [1, 30]. Therefore, in the paper we propose a new vehicle face re-identification method based on improved NMF, which takes into account the image difference caused by the capturing times. In the proposed algorithm, the variable features and stable features which are easily affected and not easily affected respectively by light variation can be obtained after model training, then the stable features are used to judge whether two vehicle face images can match or not.

3. Proposed Model

After the vehicle image is captured by the surveillance cameras on the highway, in order to remove the useless information in the image and obtain the effective vehicle face features, it is necessary to segment the vehicle face region firstly in the captured image. Yolo models have been proven to be very effective in target detection, and since YOLOv5 model has the advantages such as small network structure and fast processing speed, it is selected to segment the vehicle face region in the image in the proposed algorithm, where the codes of YOLOv5 model and the weight files can be downloaded from <https://github.com/ultralytics/yolov5> and <https://github.com/ultralytics/yolov5/releases/tag/v5.0> respectively, and the segmentation results are shown in Fig. 2.



Fig. 2. Vehicle face segmentation results

As we know, it is very important to obtain effective feature basis vectors through dimensionality reduction for object recognition, where the common dimensionality reduction methods include principal component analysis (PCA), linear discriminant analysis (LDA), local preserving projection (LPP), et al. From the principles of these methods, we know that the elements in the feature basis vectors and coefficient vectors can be either positive or negative, and the negative elements are reasonable from in terms of mathematical operation. However, for image processing, the negative elements are difficult to explain reasonably, for example, the pixels in the basis images and the weights can both not be negative. Therefore, we use NMF model to ensure the non-negativity of the decomposed matrices, where the original NMF model is shown as (1),

$$\mathbf{F}_{m \times n} \approx \mathbf{U}_{m \times r} \mathbf{V}_{r \times n}, \quad u_{ik}, v_{kj} \geq 0 \quad (1)$$

where the columns of \mathbf{F} , \mathbf{U} and \mathbf{V} represent original feature vectors, basis vectors and coefficient vectors respectively, and each coefficient vector is usually regarded as the new feature vector [31].

In the proposed algorithm, two images which represent the same vehicle can be seen as a pair of training samples, and all pairs of training samples are placed in two original feature matrices, i.e., \mathbf{F}_1 and \mathbf{F}_2 , where the column vectors at the same position in \mathbf{F}_1 and \mathbf{F}_2 represent the same vehicle. Therefore, the whole decomposition error can be obtained through fusing the decomposed errors of \mathbf{F}_1 and \mathbf{F}_2 as shown in (2),

$$J_{err} = \frac{1}{2} \sum_{i=1}^2 \|\mathbf{F}_i - \mathbf{U}\mathbf{V}_i\|_2 \quad (2)$$

where the smaller the decomposition error, the more accurate the vehicle face feature.

From the perspective of feature stability, it can be considered that each vehicle face image contains two types of features as shown in Fig. 3. One is the stable features which are robust to illumination variation such as the shape of vehicle windows and grid, the other is the variable features such as vehicle color and vehicle light brightness, et al, which will change with the light intensity.

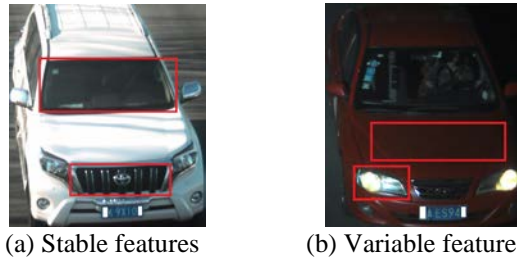


Fig. 3. Two types of features in vehicle face image

From the above analysis, we can see that it is very important to distinguish stable features and variable features from vehicle face image for vehicle re-identification. Therefore, in addition to ensuring the non-negativity of the decomposition results, some constraints which are conducive to accurate identification should be imposed on the stable and variable features, where the constraints are shown as follows:

1) The orthogonal constraint of stable features. After matrix factorization, the j th column in \mathbf{V}_i can be regarded as the new feature of the j th image in \mathbf{F}_i , i.e., $\mathbf{V}_i^{(j)}$. In our proposed algorithm, the first k dimensions in the feature vector $\mathbf{V}_i^{(j)}$ can be supposed as the stable feature, i.e., $\mathbf{S}_i^{(j)}$, where the solution of $\mathbf{S}_i^{(j)}$ is shown as (3),

$$\mathbf{S}_i^{(j)} = \mathbf{Z}\mathbf{V}_i^{(j)} \quad (3)$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{I}_{k \times k} & \mathbf{0}_{k \times (r-k)} \\ \mathbf{0}_{(r-k) \times k} & \mathbf{0}_{(r-k) \times (r-k)} \end{bmatrix}_{r \times r} \quad (4)$$

$\mathbf{I}_{k \times k}$ is k identity matrix, $\mathbf{0}_{k \times (r-k)}$ is $k \times (r-k)$ zero matrix, and r is the dimension of coefficient vector.

The stable feature of vehicle face image should have the following characteristics, that is, even if there is a large illumination difference when capturing the same vehicle two times, the stable features in two captured vehicle images still have strong similarities, on the contrary, the stable features which represent different vehicles should be different from each other, i.e., they should be as orthogonal as possible as shown in (5),

$$\begin{cases} \langle \mathbf{S}_1^{(j)}, \mathbf{S}_2^{(g)} \rangle \rightarrow 1, & j = g \\ \langle \mathbf{S}_1^{(j)}, \mathbf{S}_2^{(g)} \rangle \rightarrow 0, & j \neq g \end{cases} \quad (5)$$

$$(\mathbf{S}_1)^T \mathbf{S}_2 \rightarrow \mathbf{I}_{m \times m} \quad (6)$$

$$\mathbf{S}_1 = [\mathbf{S}_1^{(1)} \quad \mathbf{S}_1^{(2)} \quad \dots \quad \mathbf{S}_1^{(m)}] \quad (7)$$

where m is the number of columns in \mathbf{F}_i , i.e., the number of the pairs of training samples.

From the above analysis, the function of measuring the orthogonality of stable features can be obtained as shown in (8).

$$J_{sta} = \|(\mathbf{S}_1)^T \mathbf{S}_2 - \mathbf{I}_{m \times m}\|_2 \quad (8)$$

2) The similar constraint of weighted variable features. Since the feature vector \mathbf{V}_i is composed of stable features and variable features after dimensionality reduction, the last $r-k$ dimensions in \mathbf{V}_i can be regarded as the variable features, whose solution is shown in (9),

$$\mathbf{C}_i^{(j)} = (\mathbf{I}_{r \times r} - \mathbf{Z}) \mathbf{V}_i^{(j)} \quad (9)$$

We know that the light intensity changes continuously with time, so the larger the interval of the times of capturing two images, the more uncertain the difference between variable features will be. On the contrary, if the time interval is small, the variable features will be similar relatively. Therefore, we can conclude that if the interval of the times of capturing the same vehicle is large, the negative effects of variable features on re-identification will be obvious. In order to reduce the impact of variable features, we will assign different weights to the variable features according to the time difference of image acquisitions as shown in (10), where the capturing time can be obtained through surveillance camera,

$$T_j = 1 - \frac{|T_{j1} - T_{j2}|}{T_d} \quad (10)$$

T_j represents the similarity of the capturing times of the j th pair of vehicle face images, i.e., the weight, T_{j1} and T_{j2} are the capturing times of two images, and T_d is the largest time difference, $0 \leq T_j \leq 1$. Therefore, the difference between the variable features of the j th pair of vehicle face images can be measured by (11),

$$J_{j-var} = T_j \left(C_1^{(j)} \right)^T C_2^{(j)} \quad (11)$$

and the differences between the variable features of all pairs of vehicle face images can be obtained through (12).

$$J_{var} = tr \left(T \left(C_1 \right)^T C_2 \right) \quad (12)$$

$$T = \begin{bmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_m \end{bmatrix}_{m \times m} \quad (13)$$

$$C_1 = \begin{bmatrix} C_1^{(1)} & C_1^{(2)} & \cdots & C_1^{(m)} \end{bmatrix}_{r \times m} \quad (14)$$

When the capturing times of the two images which represent the same vehicle are close enough, the variable features of the two images should be similar, i.e., if $T_j \rightarrow 1$, $\left(C_1^{(j)} \right)^T C_2^{(j)} \rightarrow 1$; On the contrary, if the difference of capturing times is large, the similarity between the variable features of the two images will be uncertain, i.e., if $T_j \rightarrow 0$, $0 \leq \left(C_1^{(j)} \right)^T C_2^{(j)} \leq 1$. From the above analysis, the greater the difference measure function J_{var} , the more helpful for re-identification.

In summary, the objective function of the proposed model is shown in (15), and the optimal parameters U^* , V_1^* , V_2^* will be obtained by (16).

$$J = J_{err} + \alpha J_{sta} - \beta J_{var} \quad (15)$$

$$U^*, V_1^*, V_2^* = \arg \min_{U, V_1, V_2} \{ J_{err} + \alpha J_{sta} - \beta J_{var} \} \quad (16)$$

4. Objective Function Solution Based on Gradient Descent Method

First, in order to solve the objective function more conveniently, Equation (2), Equation (8) and Equation (12) can be further transformed into (17), (18) and (19),

$$\begin{aligned} J_{err} &= \frac{1}{2} \sum_{i=1}^2 \| F_i - UV_i \|_2^2 \\ &= \frac{1}{2} \sum_{k=1}^2 tr \left((F_i - UV_i)^T (F_i - UV_i) \right) \\ &= \frac{1}{2} \sum_{k=1}^2 tr \left(F_i^T F_i - V_i^T U^T F_i - F_i^T UV_i + V_i^T U^T UV_i \right) \end{aligned} \quad (17)$$

$$\begin{aligned}
J_{sta} &= \left\| \left(\mathbf{V}_1^{sta} \right)^T \mathbf{V}_2^{sta} - \mathbf{I}_{m \times m} \right\|_2 \\
&= \left\| \left(\mathbf{ZV}_1 \right)^T \left(\mathbf{ZV}_2 \right) - \mathbf{I}_{m \times m} \right\|_2
\end{aligned} \tag{18}$$

$$\begin{aligned}
&= tr \left(\left(\mathbf{V}_1^T \mathbf{ZV}_2 - \mathbf{I}_{m \times m} \right)^T \left(\mathbf{V}_1^T \mathbf{ZV}_2 - \mathbf{I}_{m \times m} \right) \right) \\
&= tr \left(\mathbf{V}_2^T \mathbf{ZV}_1 \mathbf{V}_1^T \mathbf{ZV}_2 - \mathbf{V}_1^T \mathbf{ZV}_2 - \mathbf{V}_2^T \mathbf{ZV}_1 + \mathbf{I}_{m \times m} \right) \\
J_{var} &= tr \left(\mathbf{T} \left(\mathbf{V}_1^{var} \right)^T \mathbf{V}_2^{var} \right) \\
&= tr \left(\left(\left(\mathbf{I}_{r \times r} - \mathbf{Z} \right) \mathbf{V}_1 \right)^T \left(\mathbf{I}_{r \times r} - \mathbf{Z} \right) \mathbf{V}_2 \right) \\
&= tr \left(\mathbf{V}_1^T \left(\mathbf{I}_{r \times r} - \mathbf{Z} \right) \mathbf{V}_2 \right)
\end{aligned} \tag{19}$$

and the derivatives of (15) with respect to \mathbf{U} , \mathbf{V}_1 and \mathbf{V}_2 are solved as shown in (20), (21) and (22).

$$\frac{\partial J}{\partial \mathbf{U}} = \sum_{i=1}^2 \left(-\mathbf{F}_i \mathbf{V}_i^T + \mathbf{U} \mathbf{V}_i \mathbf{V}_i^T \right) \tag{20}$$

$$\frac{\partial J}{\partial \mathbf{V}_1} = \mathbf{U}^T \mathbf{F}_1 + \mathbf{U}^T \mathbf{U} \mathbf{V}_1 + 2\alpha \left(\mathbf{V}_1^T \mathbf{ZV}_2 \mathbf{V}_2^T \mathbf{Z} - \mathbf{ZV}_2 \right) - \beta \left(\mathbf{I}_{r \times r} - \mathbf{Z} \right) \mathbf{V}_2 \tag{21}$$

$$\frac{\partial J}{\partial \mathbf{V}_2} = \mathbf{U}^T \mathbf{F}_2 + \mathbf{U}^T \mathbf{U} \mathbf{V}_2 + 2\alpha \left(\mathbf{V}_2^T \mathbf{ZV}_1 \mathbf{V}_1^T \mathbf{Z} - \mathbf{ZV}_1 \right) - \beta \left(\mathbf{I}_{r \times r} - \mathbf{Z} \right) \mathbf{V}_1 \tag{22}$$

Then, according to (20), (21) and (22), the iterative rules of \mathbf{U} , \mathbf{V}_1 and \mathbf{V}_2 can be obtained as shown in (23), (24) and (25).

$$u^{(x,y),t+1} \leftarrow u^{(x,y),t} \left(\frac{\sum_{i=1}^2 \mathbf{F}_i \left(\mathbf{V}_i^t \right)^T}{\sum_{i=1}^2 \mathbf{U}^t \mathbf{V}_i^t \left(\mathbf{V}_i^t \right)^T} \right)^{(x,y)} \tag{23}$$

$$v_1^{(x,y),t+1} \leftarrow v_1^{(x,y),t} \left(\frac{2\alpha \mathbf{ZV}_2^t + \beta \mathbf{V}_2^t}{\left(\mathbf{U}^t \right)^T \mathbf{F}_1 + \left(\mathbf{U}^t \right)^T \mathbf{U}^t \mathbf{V}_1^t + 2\alpha \left(\mathbf{V}_1^t \right)^T \mathbf{ZV}_2^t \left(\mathbf{V}_2^t \right)^T \mathbf{Z} + \beta \mathbf{ZV}_2^t} \right)^{(x,y)} \tag{24}$$

$$v_2^{(x,y),t+1} \leftarrow v_2^{(x,y),t} \left(\frac{2\alpha \mathbf{ZV}_1^t + \beta \mathbf{V}_1^t}{\left(\mathbf{U}^t \right)^T \mathbf{F}_2 + \left(\mathbf{U}^t \right)^T \mathbf{U}^t \mathbf{V}_2^t + 2\alpha \left(\mathbf{V}_2^t \right)^T \mathbf{ZV}_1^t \left(\mathbf{V}_1^t \right)^T \mathbf{Z} + \beta \mathbf{ZV}_1^t} \right)^{(x,y)} \tag{25}$$

In final, we will optimize the parameters \mathbf{U} , \mathbf{V}_1 and \mathbf{V}_2 according to their iterative rules, where the optimization process is as follows:

Step.1 Given the training data \mathbf{F}_1 and \mathbf{F}_2 , the balance coefficients α and β , the error threshold ξ , the maximum number of iterations N_{\max} ;

Step.2 Initializing the parameters U^0 , V_1^0 , V_2^0 and the number of iterations $t = 0$;

Step.3 $t = t + 1$, and update the parameters U , V_1 and V_2 accord to (23), (24) and (25);

Step.4

if $\left| J(U^{t+1}, V_1^{t+1}, V_2^{t+1}) - J(U^t, V_1^t, V_2^t) \right| < \xi$ or $t > N_{\max}$:

goto **Step.5**;

else:

goto **Step.3**;

Step.5 Output the optimal parameters U^* , V_1^* , V_2^* .

5. Feature Matching Based on Cosine Distance

When judging whether two images I_{test1} and I_{test2} represent the same vehicle, we should extract their stable features by decomposing their original feature vectors F_{test1} and F_{test2} as shown in (26) [32],

$$S_{new} = Z \left((U^*)^T U^* \right)^{-1} (U^*)^T F_{new} \quad (26)$$

then, we will measure the similarity of two vehicle face features by using (27),

$$d(S_{test1}, S_{test2}) = \frac{\langle S_{test1}, S_{test2} \rangle}{\|S_{test1}\|_2 \|S_{test2}\|_2} \quad (27)$$

if $d > \eta$, the two images can be considered to represent the same vehicle, on the contrary, they represent different vehicles, where η is the similarity threshold.

6. Experimental Results and Analysis

6.1 Dataset and Experimental Environment

In the experiment there are five datasets to be used which include “BITVehicle” dataset [33], “VeRI776” dataset [34], “VeRI-Wild” dataset [35], “VehicleID” dataset [36] and the self-built dataset.

1) “BITVehicle” dataset. There are 9850 vehicle images in the “BITVehicle” dataset, but not all vehicles have been captured more than twice, therefore, we selected 1500 pairs of vehicle images as positive samples from the dataset, where each pair of images represent the same vehicle; at the same time, we selected 2000 pairs of vehicle images as negative samples, where each pair of images represent the different vehicle. Some pairs of positive and negative samples are shown in Fig. 4.

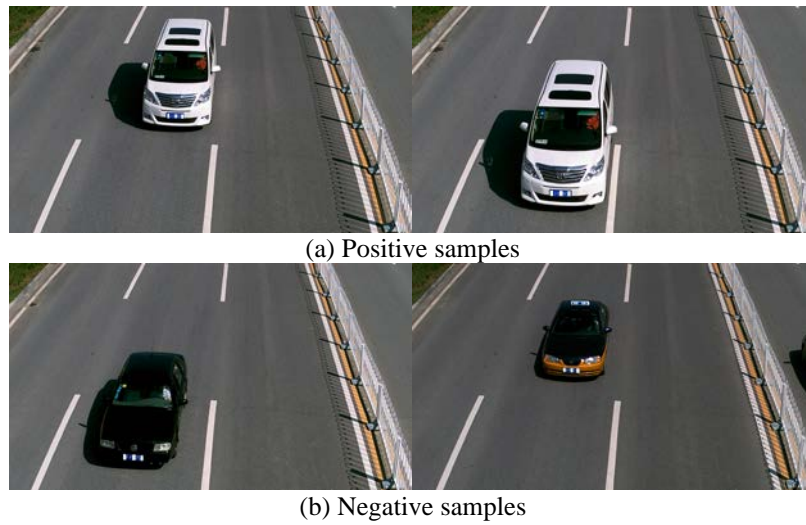


Fig. 4. Some positive samples and negative samples in the “BITVehicle” dataset

2) “VeRI776” dataset. In the “VeRI776” dataset, a total of 776 vehicles are annotated, and each vehicle contains 60 samples approximately which were captured from different angles. Since the problem we studied is based on vehicle face images, we deleted all the images which captured from invalid perspectives when selecting samples. Finally, in the experiment we selected 200 pairs of positive samples and 200 pairs of negative samples, where some pairs of positive and negative samples are shown in **Fig. 5**.



Fig. 5. Some positive samples and negative samples in the “VeRI776” dataset

3) “VeRI-Wild” dataset. This dataset contains 40,672 vehicles and a total of 416,314 images, which are captured from different perspectives. In order to meet the test requirements of vehicle face recognition algorithm, 7,056 pairs of vehicle face images which represent the same vehicles and meet the acquisition angle are selected as the positive samples for testing, and 9,000 pairs of vehicle face images which represent different vehicles are selected as the negative samples, where some samples are shown in **Fig. 6**.

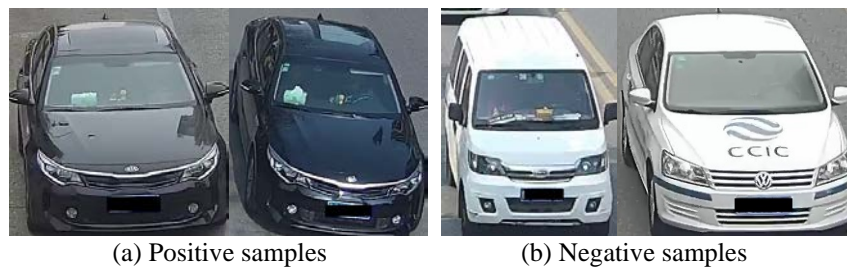


Fig. 6. Some positive samples and negative samples in the “VeRI-Wild” dataset

4) “VehicleID” dataset. There are 221763 images of 26267 vehicles in the dataset, and the vehicle in each image is either captured from the front or the back. Since vehicle face images are useful for us, 4200 pairs of vehicle face images which represent the same vehicles are selected as the positive samples, in addition, 4000 pairs of images are selected as the negative samples, some positive samples and negative samples are shown in [Fig. 7](#).



Fig. 7. Some positive samples and negative samples in the “VehicleID” dataset

5) Self-built dataset. The self-built dataset is formed by the vehicle face images which were captured by 22 surveillance cameras, and there are 4898 pairs of positive samples and 6204 pairs of negative samples, where some samples are shown in [Fig. 8](#).

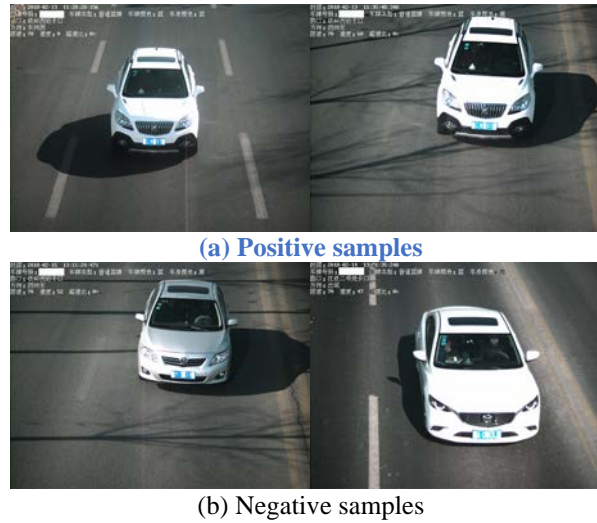


Fig. 8. Some positive samples and negative samples in the self-built dataset

In addition, in the experiment for the algorithms based on NMF, we use the PC machine whose configuration is Intel i5-10300H CPU, 16G RAM and Matlab 2017b; and for the algorithms based on deep learning, we use the server whose configuration is 16G RAM, two Geforce 1080Ti GPUs, and Tensorflow 1.0.

6.2 Parameter Setting

In the proposed algorithm, some parameters need to be set appropriately, which include the dimension m of the original feature F , the number of training samples n , the dimension of coefficient vector r , the dimension of stable feature k , the balance factors α and β , and the similarity threshold η , where some of parameters are set according to experience, and the others will be obtained through experiment.

1) Among the above parameters, m , n and r can be set according to experience. The vehicle face image is resized to $160 \times 120 \times 3$ after normalization, and the original feature vector \mathbf{F} will be obtained through stacking all columns, i.e., $m = 57600$. In addition, n and r are set as 3000 and 500 respectively.

2) From the clustering theory in pattern recognition, if the parameters k , α , β and η are appropriate enough, the positive samples should be very similar, on the contrary, there are large differences between the negative samples. Therefore, we propose a measurement function to measure the clustering property as shown in (28),

$$H = \frac{1}{N_{pos}} \sum_{i=1}^{N_{pos}} d(S_{pos-i,1}, S_{pos-i,2}) - \frac{1}{N_{neg}} \sum_{i=1}^{N_{neg}} d(S_{neg-i,1}, S_{neg-i,2}) \quad (28)$$

where 1000 pairs of positive samples and 1000 pairs of negative samples are selected from the self-built dataset randomly in the experiment, i.e., $N_{pos}=1000$ and $N_{neg}=1000$, and the clustering properties under different parameters are shown in [Table 1](#).

Table 1. The clustering properties under different parameters

Parameters	$k = 0.9r$	$k = 0.8r$	$k = 0.7r$
$\alpha = 0.1, \beta = 0.1$	0.21	0.18	0.14
$\alpha = 0.1, \beta = 0.5$	0.23	0.28	0.14
$\alpha = 0.1, \beta = 1$	0.21	0.23	0.12
$\alpha = 0.5, \beta = 0.1$	0.31	0.30	0.24
$\alpha = 0.5, \beta = 0.5$	0.33	0.33	0.25
$\alpha = 0.5, \beta = 1$	0.32	0.29	0.24
$\alpha = 1, \beta = 0.1$	0.41	0.45	0.39
$\alpha = 1, \beta = 0.5$	0.45	0.47	0.42
$\alpha = 1, \beta = 1$	0.43	0.44	0.40

From [Table 1](#), we can see that when $k = 0.8r$, $\alpha = 1$, $\beta = 0.5$, the best clustering property can be obtained.

In addition, the similarity threshold η needs to be optimized to achieve the best recognition performance, and the purpose is to make False Reject Rate (FRR) and False Accept Rate (FAR) as low as possible, where the recognition performance is measured by (29),

$$\begin{aligned} \eta^* &= \arg \min_{\eta} P(\eta) \\ &= \arg \min_{\eta} \left[\frac{N_{neg}}{N_{pos} + N_{neg}} FRR(\eta) + \frac{N_{pos}}{N_{pos} + N_{neg}} FAR(\eta) \right] \end{aligned} \quad (29)$$

The curve of $P(\eta)$ under different η is shown in [Fig. 9](#), and it can be seen that the best recognition performance of the proposed algorithm is achieved when $\eta = 0.92$.

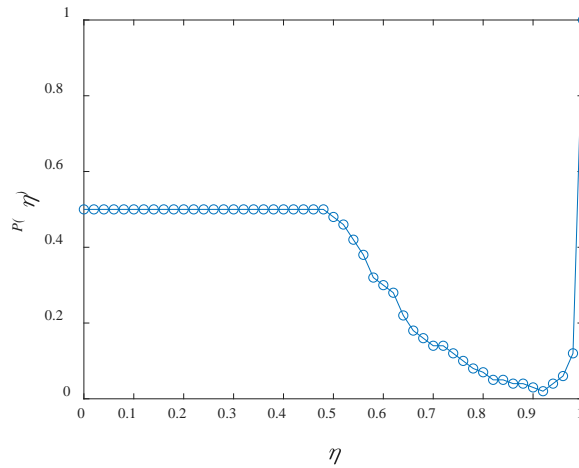


Fig. 9. The recognition performances under different similarity thresholds

6.3 Comparison of different algorithms

The algorithms which are selected in the comparison experiment are mainly divided into two categories: some algorithms are similar to the proposed algorithm [1, 30], they are all based on NMF model, and the others are based on the deep neural network model, i.e., AlexNet, Resnet50 and VGG19. Before the experiment, the training dataset needs to be constructed firstly as shown in Table 2.

Table 2. The numbers of the samples which are selected from different datasets when constructing the training dataset

Dataset	The number of the selected positive samples	The number of the selected negative samples
BITVehicle	100	100
VeRI776	50	50
VeRI-Wild	950	950
VehicleID	800	800
Self-built	600	600

Except for the images in the training dataset, the other pairs of samples will be used as the testing samples. Since the recognition principles of the selected two types of comparison algorithms are different, their training methods are also different. For the algorithms based on NMF model, the same training methods as the proposed algorithm will be used; for the algorithms based on deep learning, since the number of the vehicle face images which represent each vehicle is very small, the traditional training and classification methods may be not suitable. Therefore, their training methods have been modified in the experiment, i.e., each pair of three-channel samples are superimposed to form a six-channel sample, thus all pairs of positive samples, negative samples and testing samples are superimposed to form the new positive samples, the new negative samples and the new testing samples respectively, and the filters of the first layer of these deep neural networks are also adjusted accordingly.

In addition, the parameter T_j in the proposed algorithm is obtained based on the difference of the acquisition times of the two images, however, except for the self-built dataset, there is not image acquisition time information in the other datasets. Therefore, during the training

process, the parameters T_j from different datasets will be set differently, if the training samples are from the self-built dataset, the parameter T_j will be obtained according to (10); if the training samples are from the other datasets, the parameters T_j will be set to 1 or 0, i.e., when the two images are both captured in the day or in the night, the parameter T_j will be set to 1, and when the two images are captured respectively in the day and in the night, the parameter T_j will be set to 0.



After the models are trained, all the samples from different datasets will be tested, and the testing results are shown in Table 3, where **FRR** and **FAR** represent false reject rate and false accept rate respectively.



Table 3. The recognition results of different algorithms

	Alexnet		Resnet50		VGG16		[1]		[30]		Ours	
	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR
BITVehicle	0.16	0.12	0.16	0.12	0.13	0.09	0.20	0.17	0.19	0.14	0.12	0.10
VeRI776	0.17	0.11	0.14	0.11	0.16	0.13	0.15	0.14	0.17	0.12	0.13	0.11
VeRI-Wild	0.17	0.14	0.13	0.12	0.17	0.13	0.18	0.17	0.18	0.19	0.14	0.11
VehicleID	0.19	0.16	0.17	0.16	0.16	0.14	0.26	0.24	0.23	0.21	0.18	0.15
Self-built	0.24	0.20	0.22	0.15	0.24	0.19	0.26	0.23	0.25	0.22	0.12	0.09
Average	0.19	0.15	0.16	0.13	0.17	0.14	0.21	0.19	0.2	0.18	0.14	0.11

From Table 3, it can be seen that the proposed algorithm is slightly better than the other algorithms in terms of **FRR** and **FAR**. Since there are many pairs of samples in the self-built dataset which are captured under large illumination differences, the performances of the above comparison algorithms are significantly reduced, but the proposed algorithm can still achieve good recognition results. The reason is that during training the proposed algorithm ignores the variable features which are affected by the light intensity, and pays more attention to the stable features of vehicle face images, which makes the proposed algorithm have good robustness to light variation.

Table 4. The recognition results of some test samples

Light difference	Testing samples		The property of samples	The accuracy of recognition results
small			Positive	Alexnet: ✓ Resnet50: ✓ VGG16: ✓ [1]: ✓ [30]: ✓ Ours: ✓
small			Negative	Alexnet: ✓ Resnet50: ✓ VGG16: ✓ [1]: ✗ [30]: ✓ Ours: ✓

large		Positive	Alexnet: × Resnet50: ✓ VGG16: × [1]: × [30]: × Ours: ✓
large		Positive	Alexnet: × Resnet50: × VGG16: × [1]: × [30]: × Ours: ✓

The conclusion that the proposed algorithm has good robustness to light variation has been verified from the results in [Table 4](#).

7. Conclusion

To improve the robustness of vehicle face features to light variation when capturing image, a NMF model with time difference constraint is proposed. The innovation of the thesis is that the stable features which are less affected by light intensity variation can be obtained after model training, i.e., even if the same vehicle is captured twice under different light intensities, we can still conclude accurately that the two images represent the same vehicle. Although we have achieved good recognition results, there are still some problems to be solved, for example, when there is a large difference in the capture angle in two vehicle face images which represent the same vehicle, the recognition accuracy is still not very high. Therefore, the universality of the proposed algorithm needs to be further improved.

References

- [1] X. JIA and F. M. SUN, "Vehicle face re-identification algorithm based on Siamese nonnegative matrix factorization," *Chinese Journal of Scientific Instrument*, vol. 41, no. 6, pp. 132-139, 2020. [Article \(CrossRef Link\)](#)
- [2] K. J. Kim, S. M. Park and Y. J. Choi, "Deciding the number of color histogram bins for vehicle color recognition," in *Proc. of Asia-Pacific Services Computing Conference*, pp. 134-138, December 9-12, 2008. [Article \(CrossRef Link\)](#)
- [3] N. Baek, S. M. Park, K. J. Kim and S. B. Park, "Vehicle color classification based on the support vector machine method," in *Proc. of International Conference on Intelligent Computing*, pp. 1133-1139, August 21-24, 2007. [Article \(CrossRef Link\)](#)
- [4] P. Negri, X. Clady, M. Milgram and R. Poulencard, "An oriented-contour point based voting algorithm for vehicle type classification," in *Proc. of International Conference on Pattern Recognition*, pp. 574-577, August 20-24, 2006. [Article \(CrossRef Link\)](#)
- [5] P. Chen, X. Bai and W. Liu, "Vehicle color recognition on urban road by feature context," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2340-2346, 2014. [Article \(CrossRef Link\)](#)

- [6] B. Zhang, "Reliable classification of vehicle types based on cascade classifier ensembles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 322-332, 2013. [Article \(CrossRef Link\)](#)
- [7] W. W. L. Lam, C. C. C. Pang and N. H. C. Yung, "Vehicle-Component Identification Based on Multiscale Textural Couriers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 4, pp. 681-694, 2007. [Article \(CrossRef Link\)](#)
- [8] A. P. Psyllos, C. N. E. Anagnostopoulos and E. Kayafas, "Vehicle Logo Recognition Using a SIFT-Based Enhanced Matching Scheme," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 322-328, 2010. [Article \(CrossRef Link\)](#)
- [9] M. J. Leotta and J. L. Mundy, "Vehicle surveillance with a generic, adaptive, 3D vehicle model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1457-1469, 2011. [Article \(CrossRef Link\)](#)
- [10] J. Sochor, A. Herout, and J. Havel, "Boxcars: 3d boxes as cnn input for improved fine-grained vehicle recognition," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3006-3015, June 27-30, 2016. [Article \(CrossRef Link\)](#)
- [11] J. Prokaj and G. Medioni, "3-D model based vehicle recognition," in *Proc. of 2009 Workshop on Applications of Computer Vision*, pp. 1-7, December 7-8, 2009. [Article \(CrossRef Link\)](#)
- [12] X. C. Liu, W. Liu, J. K. Zheng, C. G. Yan and T. Mei, "Beyond the Parts: Learning Multi-view Cross-part Correlation for Vehicle Re-identification," in *Proc. of ACM International Conference on Multimedia*, pp. 907-915, October 12-16, 2020. [Article \(CrossRef Link\)](#)
- [13] Y. O. Adu-Gyamfi, S. K. Asare, A. Sharma and T. Tienaah, "Automated vehicle recognition with deep convolutional neural networks," *Transportation Research Record*, vol. 2645, no. 1, pp. 113-122, 2017. [Article \(CrossRef Link\)](#)
- [14] K. Huang and B. Zhang, "Fine-grained vehicle recognition by deep Convolutional Neural Network," in *Proc. of 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, pp. 465-470, October, 15-17, 2016. [Article \(CrossRef Link\)](#)
- [15] Q. Zhang, L. Zhuo, J. F. Li, J. Zhang, H. Zhang and X. G. Li, "Vehicle color recognition using Multiple-Layer Feature Representations of lightweight convolutional neural network," *Signal Processing*, vol. 147, pp. 146-153, 2018. [Article \(CrossRef Link\)](#)
- [16] Y. Y. Wu and C. M. Tsai, "Pedestrian, bike, motorcycle, and vehicle classification via deep learning: deep belief network and small training set," in *Proc. of International Conference on Applied System Innovation*, pp. 1-4, May 26-31, 2016. [Article \(CrossRef Link\)](#)
- [17] W. Hai, Y. Cai and L. Chen, "A Vehicle Detection Algorithm Based on Deep Belief Network," *The Scientific World Journal*, pp. 1-7, 2014. [Article \(CrossRef Link\)](#)
- [18] J. T. Wang, H. Zheng, Y. Huang and X. H. Ding, "Vehicle type recognition in surveillance images from labeled web-nature data using deep transfer learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 9, pp. 2913-2922, 2018. [Article \(CrossRef Link\)](#)
- [19] H. Wang, Y. J. Yu, Y. F. Chen and X. B. Chen, "A vehicle recognition algorithm based on deep transfer learning with a multiple feature subspace distribution," *Sensors*, vol. 18, no. 12, pp. 4109, 2018. [Article \(CrossRef Link\)](#)
- [20] Y. Chen, C. Yang and S. Y. Yang, "A method for special vehicle recognition based on deep-transfer model," in *Proc. of International Conference on Instrumentation & Measurement, Computer, Communication and Control*, pp. 167-170, July 21-23, 2016. [Article \(CrossRef Link\)](#)
- [21] A. Q. Hu, H. Li, F. Zhang and W. Zhang, "Deep Boltzmann machines based vehicle recognition," in *Proc. of Chinese Control and Decision Conference*, pp. 3033-3038, May 31-June 2, 2014. [Article \(CrossRef Link\)](#)
- [22] D. F. S. Santos, G. B. De Souza and A. N. Marana, "A 2D Deep Boltzmann Machine for robust and fast vehicle classification," in *Proc. of SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 155-162, October 17-20, 2017. [Article \(CrossRef Link\)](#)
- [23] C. Gou, K. F. Wang, Y. J. Yao and Z. X. Li, "Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1096-1107, 2015. [Article \(CrossRef Link\)](#)

- [24] X. B. Liu, S. L. Zhang, X. Y. Wang, R. C. Hong and Q. Tian, "Group-Group Loss-Based Global-Regional Feature Learning for Vehicle Re-Identification," *IEEE Transactions on Image Processing*, vol. 29, pp. 2638-2652, 2019. [Article \(CrossRef Link\)](#)
- [25] X. Jin, C. L. Lan, W. J. Zeng and Z. B. Chen, "Uncertainty-Aware Multi-Shot Knowledge Distillation for Image-Based Object Re-Identification," in *Proc. of the AAAI Conference on Artificial Intelligence*, pp. 11165-11172, February 7-12, 2020.
- [26] D. C. Meng, L. Li, X. J. Liu, Y. D. Li, S. J. Yang, Z. J. Zha, X. Y. Gao, S. H. Wang and Q. M. Huang, "Parsing-based View-aware Embedding Network for Vehicle Re-Identification," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7103-7112, June 13-19, 2020.
- [27] T. S. Chen, C. T. Liu, C. W. Wu and S. Y. Chien, "Orientation-aware Vehicle Re-identification with Semantics-guided Part Attention Network," in *Proc. of European Conference on Computer Vision*, pp. 330-346, August 23-28, 2020. [Article \(CrossRef Link\)](#)
- [28] Z. D. Zheng, T. Ruan, Y. C. Wei, Y. Yang and T. Mei, "VehicleNet: Learning Robust Visual Representation for Vehicle Re-identification," *IEEE Transactions on Multimedia, Early Access*, 2020. [Article \(CrossRef Link\)](#)
- [29] P. Khorramshahi, N. Peri, J. C. Chen and R. Chellappa, "The Devil is in the Details: Self-Supervised Attention for Vehicle Re-Identification," in *Proc. of European Conference on Computer Vision*, pp. 369-386, August 23-28, 2020. [Article \(CrossRef Link\)](#)
- [30] C. H. Shi and C. D. Wu, "Vehicle Face Recognition Algorithm Based on Weighted Nonnegative Matrix Factorization with Double Regularization Terms," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 5, pp. 2171-2185, 2020.
- [31] X. Jia, F. M. Sun, H. J. Li, Y. D. Cao and X. Zhang, "Image Multi-Label Annotation Based on Supervised Nonnegative Matrix Factorization with New Matching Measurement," *Neurocomputing*, vol. 219, pp. 518-525, 2017. [Article \(CrossRef Link\)](#)
- [32] X. Jia, F. M. Sun, H. J. Li and Y. D. Cao, "Hand Vein Recognition Algorithm Based on NMF with Sparsity and Clustering Property Constraints in Feature Mapping Space," *Chinese Journal of Electronics*, vol. 28, no. 6, pp. 1184-1190, 2019. [Article \(CrossRef Link\)](#)
- [33] Z. Dong, Y. W. Wu, M. T. Pei and Y. D. Jia, "Vehicle Type Classification Using a Semisupervised Convolutional Neural Network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2247-2256, 2015. [Article \(CrossRef Link\)](#)
- [34] X. C. Liu, W. Liu, H. D. Ma and H. Y. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *Proc. of IEEE International Conference on Multimedia and Expo*, pp. 1-6, July 11-15, 2016. [Article \(CrossRef Link\)](#)
- [35] Y. H. Lou, Y. Bai, J. Liu, S. Q. Wang and L. Y. Duan, "VERI-Wild: A Large Dataset and a New Method for Vehicle Re-Identification in the Wild," in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3230-3238, June 15-20, 2019. [Article \(CrossRef Link\)](#)
- [36] H. Y. Liu, Y. H. Tian, Y. W. Wang, L. Pang and T. J. Huang, "Deep Relative Distance Learning: Tell the Difference between Similar Vehicles," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2167-2175, June 27-30, 2016. [Article \(CrossRef Link\)](#)



Na Ma. She received her B.Sc. degree in 2008 from Shenyang University, received her M.Sc. degree in 2010 from Northeastern University, now she is a doctoral student in Liaoning Technical University. Her main research interests include machine learning and intelligent decision.



Tingxin Wen. He received his B.Sc. degree in 1997 from Liaoning Technical University, received his M.Sc. degree in 2006 from Liaoning Technical University, received his Ph.D. degree in 2016 from Liaoning Technical University, now he is a professor in Liaoning Technical University. His main research interests include information system and intelligent decision.