

Research Article

Vehicle Type Recognition Combining Global and Local Features via Two-Stage Classification

Wei Sun,^{1,2} Xiaorui Zhang,^{2,3} Shunshun Shi,¹ Jun He,⁴ and Yan Jin¹

¹*School of Information and Control, Nanjing University of Information Science & Technology, Nanjing 210044, China*

²*Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing 210044, China*

³*School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China*

⁴*School of Electronic and Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China*

Correspondence should be addressed to Wei Sun; sunw0125@163.com

Received 22 August 2017; Revised 8 October 2017; Accepted 17 October 2017; Published 13 November 2017

Academic Editor: Yakov Strelniker

Copyright © 2017 Wei Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study proposes a new vehicle type recognition method that combines global and local features via a two-stage classification. To extract the continuous and complete global feature, an improved Canny edge detection algorithm with smooth filtering and non-maxima suppression abilities is proposed. To extract the local feature from four partitioned key patches, a set of Gabor wavelet kernels with five scales and eight orientations is introduced. Different from the single-stage classification, where all features are incorporated into one classifier simultaneously, the proposed two-stage classification strategy leverages two types of features and classifiers. In the first stage, the preliminary recognition of large vehicle or small vehicle is conducted based on the global feature via a k -nearest neighbor probability classifier. Based on the preliminary result, the specific recognition of bus, truck, van, or sedan is achieved based on the local feature via a discriminative sparse representation based classifier. We experiment with the proposed method on the public and established datasets involving various challenging cases, such as partial occlusion, poor illumination, and scale variation. Experimental results show that the proposed method outperforms existing state-of-the-art methods.

1. Introduction

Vehicle type recognition (VTR) is one key component of intelligent transportation systems (ITS) and has a wide range of applications such as traffic flow statistics, intelligent parking systems, electronic toll collection systems, and access control systems [1]. For example, it can be utilized to realize the automatic fare collection (AFC) according to different vehicle types in some paying parking lots or be applied to the nonstop toll collection system to realize automatic toll calculation in highway toll stations. Additionally, it can also be used to find and locate the vehicles that break traffic regulations and are escaping from the accident scene in traffic video monitoring.

With the extensive use of traffic surveillance cameras, image-based methods are attracting more and more attention of researchers in the VTR. The vehicle face image contains precious information for the VTR, and extracting features from the vehicle face image will lead to a better recognition

result. However, illumination change, scale variation, and partial occlusion will badly influence the performance of the VTR in real-world traffic environments. In order to improve the performance of the VTR, researchers have proposed many effective methods. These existing methods mainly consist of two key steps, that is, feature extraction and classifier design, which directly determine how well the VTR method works.

There are many typical features that can be applied to the VTR, such as edge based feature [2, 3], color based feature [4], symmetry based feature [5–7], SIFT descriptor based feature [8, 9], HOG descriptor based feature [10], and Gabor filter based feature [11]. The edge based feature extraction methods extract the edge of vehicle image by a certain edge operator, such as Sobel operator. The symmetry based methods utilize projection or corner detection algorithms according to the geometric symmetry of the vehicle face image in spatial profile to detect and recognize the vehicle. The two kinds of methods are able to extract the geometrical contour of vehicle

image accurately and quickly using small storage space and little computation time. However, these methods are easily influenced by some adverse factors, such as illumination change, scale variation, and partial occlusion; when these factors occur, their performance in feature extraction will degrade. Therefore, these feature extraction methods are commonly used to extract the global contour of vehicle image, and the extracted features also only apply to the preliminary recognition in the VTR.

Unlike the two kinds of methods mentioned above, feature extraction methods, such as SIFT descriptor based, HOG descriptor based, or Gabor filter based, can extract structural details of vehicle image from multiple scales and orientations, and they are insensitive to illumination change or scale variation. Therefore, they are commonly used for precise recognition. However, due to extracting multiple features from multiple scales and orientations, these feature extraction methods always generate a large amount of additional feature information compared with the original image, which will increase the computational complexity of VTR algorithms.

Intuitively, global information means the holistically geometrical configuration of vehicle contour, while structural details are embedded in the local variations of vehicle appearance. Therefore, extracting both global geometrical information and local structural details from vehicle images through certain feature extraction methods and leveraging the extracted feature information via suitable classifiers will help improve the performance of the VTR.

In terms of the classifier design, typical classifiers include KNN [3, 4], SVM [12–14], and ANN [15]. For the KNN classifier, it has a simple principle and does not need training in advance. However, when the number of the samples in training set increases, its computation time will also increase accordingly. The methods based on SVM or ANN classifier can effectively utilize various vehicle features and obtain good classification performance. However, these methods need to train classifier parameters in advance by collecting many samples of different types of vehicles and are easy to fall into a local optimum solution during training the classifier parameters. The classifier based on sparse representation has been successfully applied to the face recognition due to excellent characteristics: without involving complex parameter training and only needing to consider original image samples as a dictionary without any additional transformation [16]. Further research finds that if we can learn a discriminative dictionary from the original dictionary via certain dictionary learning schemes before pattern recognition, then we will achieve more accurate and reliable classification results based on the learned dictionary than based on the original dictionary [17].

Additionally, the above-mentioned classification methods adopt a single-stage classification strategy; that is, all features are incorporated into one classifier together to recognize the vehicle type. When the number of the recognized vehicles types increases, the methods based on the single-stage classification need lots of training samples to train many classifier parameters, which will inevitably increase the difficulty of classifier design for a given recognition performance [18].

To address the aforementioned limitations, this paper proposes a new VTR method combining global and local features via a two-stage classification, whereby the global feature and local feature are jointly applied to the VTR, and their advantages in expressing vehicle geometrical contour and structural details are leveraged by a proposed two-stage classification strategy. The proposed method enables an accurate and reliable VTR. First, the global feature is used to preliminarily recognize the type of a vehicle from the geometrical contour viewpoint, and the local feature is further used to recognize the specific type from the structural details viewpoint. Second, due to exploiting a two-stage classification strategy, the total classification task is appropriately assigned to two different classifiers. Therefore, the design of each classifier is simplified and their design difficulty is also lowered accordingly. This improves the overall classification performance of the VTR in accuracy and reliability compared with the methods based on the single-stage classification strategy.

This paper advances the research on VTR by making the following specific contributions: First, an improved Canny edge detection algorithm with smooth filtering and non-maxima suppression abilities is proposed to extract a continuous and complete global feature of vehicle image. Second, the whole vehicle image is partitioned into four nonoverlapping patches based on the key parts of a vehicle, and the local feature is extracted by a set of Gabor wavelet kernels with five scales and eight orientations based on four partitioned key patches. When the vehicle is partially occluded, it still can be correctly recognized by using the local feature extracted from other nonoccluded patches. Third, a k -nearest neighbor probability classifier (KNNPC) with the Hausdorff distance measure is proposed to improve the reliability of the first stage of classification, where vehicle type is preliminarily recognized as a large or small vehicle from the geometrical contour viewpoint. Fourth, a discriminative sparse representation based classifier (DSRC) that adopts a dictionary learning scheme based on the Fisher discrimination criterion is introduced to the second stage of classification, which enables a more specific classification based on the extracted local feature.

The rest of this paper is organized as follows. Section 2 presents the global and local feature extraction methods as well as the image partition method based on the key parts of a vehicle. Section 3 describes a two-stage classification strategy for the VTR. Experiments and analysis are shown in Section 4 to illustrate the effectiveness of the proposed VTR method. The final section summarizes this study and future research directions.

2. Feature Extraction

As mentioned previously, both the global geometrical contour and local structural details of a vehicle play important roles in the VTR. Therefore, there is a need to extract these features through corresponding feature extraction methods. In this paper, the global geometrical contour is extracted by an improved Canny edge detection algorithm with smooth filtering and non-maxima suppression abilities, and the local

structural details are extracted by a set of Gabor wavelet kernels with multiple scales and orientations.

2.1. Global Feature Extraction. The edge of vehicle image contains rich contour information of the vehicle. Therefore, it is regarded as a global feature to preliminarily recognize the type of a vehicle in this paper.

Commonly, some operators can be used to extract the edge of a vehicle, such as Sobel, Roberts, Prewitt, and Canny. However, these edge detection algorithms based on a certain operator have their own limitations. For example, the Sobel and Prewitt operators have the ability to fast detect the edge of an object but cannot produce a thin edge; therefore, they are unsuitable for accurate location. The Roberts operator is capable of locating the edge accurately but is sensitive to noises; therefore, it cannot effectively suppress the noises existing in the image. The Canny operator has the abilities to smooth a strong edge and suppress noises. It also can extract accurate and complete edge under good illumination; however, when the illumination becomes poor, it cannot detect a weak edge [19].

In order to achieve a better edge, we propose an edge detection method based on the improved Canny operator to extract the global feature of vehicle images. It exploits a double-threshold algorithm based on OTSU to self-adaptively determine the edge of a vehicle according to illumination changes. Based on non-maxima suppression and double-threshold judgment, the proposed method can find a continuous and complete edge. The detailed steps are as follows.

Step 1. According to (1), smooth the input image $f(x, y)$ using a Gaussian filter $G(x, y, \sigma)$ to remove Gaussian noise [20].

$$S(x, y) = G(x, y, \sigma) * f(x, y), \quad (1)$$

where σ is variance and $*$ indicates convolution operation. In this paper, when $\sigma = 1$, good smoothing results can be obtained. Therefore, we let $\sigma = 1$, and

$$G(x, y, 1) = \begin{bmatrix} 0.0751 & 0.1238 & 0.0751 \\ 0.1238 & 0.2043 & 0.1238 \\ 0.0751 & 0.1238 & 0.0751 \end{bmatrix} \quad (2)$$

accordingly.

Step 2 (calculate gradient magnitude). The gradient of each pixel in the smoothed image is determined by applying the Sobel operator. The Sobel operators for x and y directions are, respectively,

$$H_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad (3)$$

$$H_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}.$$

In order to improve real-time performance, the gradient magnitude $M(x, y)$ and gradient direction $\theta(x, y)$ are determined by

$$M(x, y) = |M_x(x, y)| + |M_y(x, y)|$$

$$\theta(x, y) = \arctan\left(\frac{M_y(x, y)}{M_x(x, y)}\right), \quad (4)$$

where $M_x(x, y) = H_x * S(x, y)$ and $M_y(x, y) = H_y * S(x, y)$.

Step 3. Implement non-maxima suppression on the gradient magnitude $M(x, y)$ calculated in Step 2 to determine the candidates of edge pixels. We define a 3×3 mask template that can traverse the entire image. In this template, if the gradient magnitude $M(i, j)$ of the central pixel (i, j) is not less than that of two other pixels along the gradient orientation $\theta(i, j)$, then we keep the maximal gradient magnitude and let other gradient magnitude be equal to zero; that is, if $M(i, j)$ is maximum, then let $\tilde{M}(x, y) = M(i, j)$; otherwise, let $\tilde{M}(x, y) = 0$. The specific comparison process is as follows: if $\theta(i, j) \in (-\pi/2, -3\pi/8]$ or $\theta(i, j) \in (3\pi/8, \pi/2)$, then we compare $M(i, j)$ with $M(i+1, j)$ and $M(i-1, j)$; if $\theta(i, j) \in (-3\pi/8, -\pi/8]$, then we compare $M(i, j)$ with $M(i-1, j-1)$ and $M(i+1, j+1)$; if $\theta(i, j) \in (-\pi/8, \pi/8]$, then we compare $M(i, j)$ with $M(i, j-1)$ and $M(i, j+1)$; if $\theta(i, j) \in (\pi/8, 3\pi/8]$, then we compare $M(i, j)$ with $M(i-1, j+1)$ and $M(i+1, j-1)$.

Step 4. Double thresholds are used to determine strong and weak edges. We set two thresholds T_{high} and T_{low} . (i) If $\tilde{M}(x, y) \geq T_{\text{high}}$, then the pixel at (x, y) is determined as an edge pixel and let $\tilde{M}(x, y) = 255$. (ii) If $\tilde{M}(x, y) \leq T_{\text{low}}$, then the pixel at (x, y) is determined as a nonedge pixel and let $\tilde{M}(x, y) = 0$. (iii) If $T_{\text{low}} < \tilde{M}(x, y) < T_{\text{high}}$, then continue to search in a 3×3 neighborhood based on the current central pixel (x, y) to find whether there is a pixel whose gradient magnitude is more than T_{high} . If such a pixel exists, then the pixel is also determined as an edge pixel and let $\tilde{M}(x, y) = 255$; otherwise, the pixel is determined as a nonedge pixel and let $\tilde{M}(x, y) = 0$.

Different from the traditional Otsu algorithm [21] that only determines a single threshold, in this step, we propose a self-adaptive algorithm to determine the two thresholds of T_{high} and T_{low} based on the histogram of the gradient image $M(x, y)$. Assume that the gradient magnitude i ranges from zero to $L-1$ in the $M(x, y)$; that is, $i \in [0, 1, 2, \dots, L-1]$, and we divide the pixels into three categories according to the gradient magnitude, that is, C_0 , C_1 , and C_2 , where C_0 is used to indicate nonedge pixels and their range is defined as $[0, k]$; C_2 is used to indicate edge pixels and their range is defined as $[m+1, L-1]$; and C_1 is used to indicate the pixels that cannot be definitely determined as edge pixels or nonedge pixels and their range is defined as $[k+1, m]$. Let n_i denote the number of the pixels whose gradient magnitude is i , N denotes the total number of the pixels in the gradient image $M(i, j)$, and p_i indicates the percentage of the pixels whose gradient magnitude is i in the gradient image $M(i, j)$; that is,

$p_i = n_i/N$. The expectation of the gradient magnitude in the whole image is $E = \sum_{i=0}^{L-1} i \cdot p_i$.

The expectations of the gradient magnitude of the pixels in C_0 , C_1 , and C_2 are, respectively,

$E_0(k) = \sum_{i=0}^k (i \cdot p_i)/p(k)$, $E_1(k, m) = \sum_{i=k+1}^m (i \cdot p_i)/p(k, m)$, and $E_2(m) = \sum_{i=m+1}^{L-1} (i \cdot p_i)/p(m)$, where $p(k) = \sum_{i=0}^k p_i$, $p(k, m) = \sum_{i=k+1}^m p_i$, and $p(m) = \sum_{i=m+1}^{L-1} p_i$.

In order to determine T_{high} and T_{low} , we define an evaluation function $\sigma^2(k, m)$ inspired by the traditional Otsu algorithm:

$$\sigma^2(k, m) = (E_0(k) - E)^2 \cdot p(k) + (E_1(k, m) - E)^2 \cdot p(k, m) + (E_2(m) - E)^2 \cdot p(m). \quad (5)$$

Calculate and compare every $\sigma^2(k, m)$; let $(\hat{k}, \hat{m}) = \arg \max(\sigma^2(k, m))$, where $k = 0, 1, 2, \dots, L-2$, or $L-1$, and $m = k+1, k+2, \dots, L-2$, or $L-1$. Then, we let $T_{\text{low}} = \hat{k}$ and $T_{\text{high}} = \hat{m}$; the two thresholds T_{low} and T_{high} are determined accordingly.

2.2. Local Feature Extraction. The global feature can be used to recognize the type of a vehicle roughly, such as large or small. In order to further recognize a specific type, such as sedan, van, bus, or truck, other features to represent the local structural details of a vehicle need to be extracted.

2.2.1. Image Partition Based on Key Parts. Not all parts in a vehicle face image are useful for the VTR; only some key parts with salient features (e.g., vehicle roof, windscreen and rear-view mirror, hood, and license plate) are available. Additionally, the partial occlusion always occurs under real-world traffic environments. If we partition the vehicle face image into several key patches, even when the partial occlusion occurs, we can still recognize the vehicle type through other key parts in other nonoccluded patches. Therefore, we averagely partition the vehicle face image into four key patches from the top to the bottom, (i) vehicle roof, (ii) windscreen and rear-view mirror, (iii) hood, and (iv) license plate, as shown in Figure 1.

2.2.2. Local Feature Extraction. Gabor wavelets, whose kernels act very similarly to mammalian visual cortical cells, have strong characteristics of spatial locality and orientation, making them a suitable choice for image feature extraction in the VTR [22]. Therefore, the Gabor wavelet representation of the vehicle image is introduced to extract the local features in every partitioned patch in this paper, which can not only obtain better structural details with multiple scales and multiple orientations but also improve the robustness to illumination change or partial occlusion. The Gabor wavelet kernels can be defined by [22]

$$G_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} \exp \left[-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2} \right] \times \left[\exp(i \cdot k_{u,v} \cdot z) - \exp \left(-\frac{\sigma^2}{2} \right) \right], \quad (6)$$

where u and v define the orientation and scale of the Gabor kernels, respectively, $z = (x, y)$, $\|\cdot\|$ denotes the norm operator, (x, y) represents the pixel coordinates, and the wave vector $k_{u,v}$ is defined as

$$k_{u,v} = k_v \exp(i \cdot \varphi_u), \quad (7)$$

where $k_v = k_{\text{max}}/f^v$, $\varphi_u = u \cdot \pi/8$, k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain.

It is usual to use the Gabor wavelets at five different scales, $v \in \{0, 1, \dots, 4\}$, and eight orientations, $u \in \{0, 1, \dots, 7\}$, with the following parameters: $\sigma = 2\pi$, $k_{\text{max}} = \pi/2$, and $f = \sqrt{2}$ [23].

For Gabor feature extraction, we convolve the image $I(z)$ with a set of Gabor wavelet kernels defined by (6) at every pixel (x, y) :

$$F_{u,v}(z) = I(z) \otimes G_{u,v}(z), \quad (8)$$

where $z = (x, y)$, $F_{u,v}(z)$ is the convolution result corresponding to the Gabor wavelet kernel at orientation u and scale v , and it also is called Gabor feature image in this paper, $I(z)$ expresses gray level distribution of an image, and \otimes represents the convolution operator. Therefore, the set $S = \{F_{u,v}(z) : u \in \{0, 1, \dots, 7\}, v \in \{0, 1, \dots, 4\}\}$ forms the Gabor wavelet representation of the image $I(z)$.

Applying the convolution theorem, we can derive every $F_{u,v}(z)$ via the fast Fourier transform (FFT) [24].

$$F_{u,v}(z) = \mathcal{F}^{-1} \{ \mathcal{F} \{ I(z) \} \mathcal{F} \{ G_{u,v}(z) \} \}, \quad (9)$$

where \mathcal{F} and \mathcal{F}^{-1} indicate the Fourier transform and inverse Fourier transform, respectively.

To leverage the advantage of Gabor wavelets with five scales and eight orientations, we concatenate all these Gabor feature images $F_{u,v}(z)$ in set S and derive an augmented feature vector χ . Before the concatenation, we first downsample every $F_{u,v}(z)$ into $F_{u,v}^{(\rho)}$ by a factor ρ to reduce the space dimension and normalize it to zero mean and unit variance. We then transform every $F_{u,v}^{(\rho)}$ into a vector by concatenating its columns. Finally, the reduced Gabor feature vector $\chi^{(\rho)}$ is defined as $\chi^{(\rho)} = (F_{0,0}^{(\rho)T} F_{0,1}^{(\rho)T} \dots F_{4,7}^{(\rho)T})^T$, where T is the transpose operator.

3. Recognition

3.1. Two-Stage Classification Strategy. Unlike the single-stage classification based methods that need to design a more complicated classifier, collect more training samples, and spend more computational time on training classifier parameters, we propose a two-stage classification strategy based on two different types of classifiers and features. In the first stage of classification, we firstly recognize the type of the test sample as large vehicle or small vehicle using the KNNPC based on the extracted global feature. Based on this, we further recognize the type of the large vehicle as bus or truck as well as the type of the small vehicle as van or sedan using the DSRC

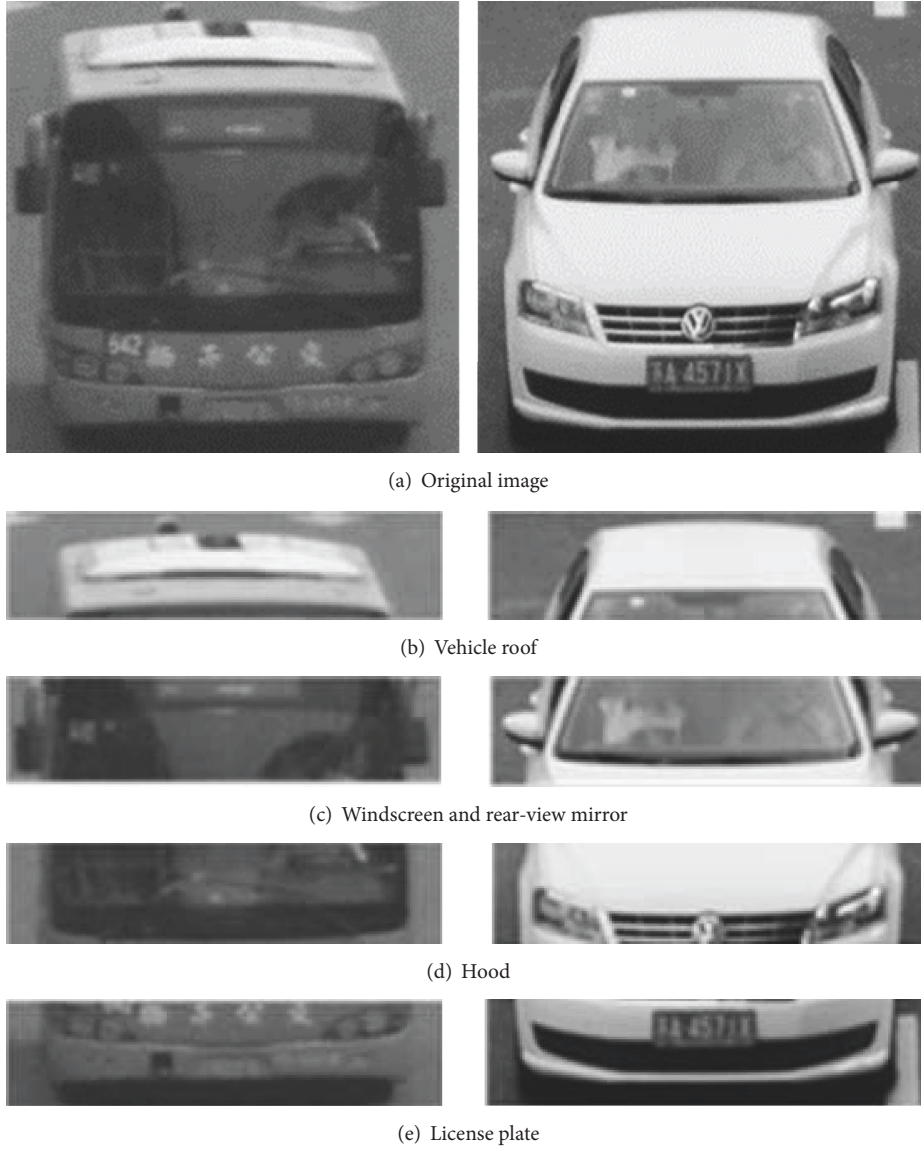


FIGURE 1: Vehicle image partition.

based on the extracted local feature in the second stage of classification. The detailed classification process is illustrated in Figure 2.

3.2. Preliminary Recognition Based on Global Feature and KNNPC. In the first stage of classification, we propose a robust classification method based on the local feature and KNNPC in the first stage of classification. This method first estimates the cumulative probabilities of the test sample on its k -nearest neighbors that may belong to different classes and then selects the maximum weighted class as the classification result. The selection of the k -nearest neighbors is based on an improved Hausdorff distance measure (IHDM), and the cumulative probabilities of the test sample are based on Gaussian kernel density estimation (KDE).

3.2.1. Improved Hausdorff Distance Measure. Hausdorff distance (HD) is one of the commonly used measures for object matching. It calculates the distance between two point sets of the edges in two-dimensional binary images without establishing correspondences. Compared with other methods, such as Euclidean distance, the HD has better robustness to noises and partial occlusion due to not involving point-to-point distance calculation. In order to enhance the first stage of classification of the VTR, we introduce an IHDM based on a statistics scheme to calculate the HD between the test sample and training samples [25].

The classical HD measure between two point sets $A = \{a_1, a_2, \dots, a_{N_A}\}$ and $B = \{b_1, b_2, \dots, b_{N_B}\}$ with sizes N_A and N_B , respectively, is defined as

$$H(A, B) = \max(h(A, B), h(B, A)), \quad (10)$$

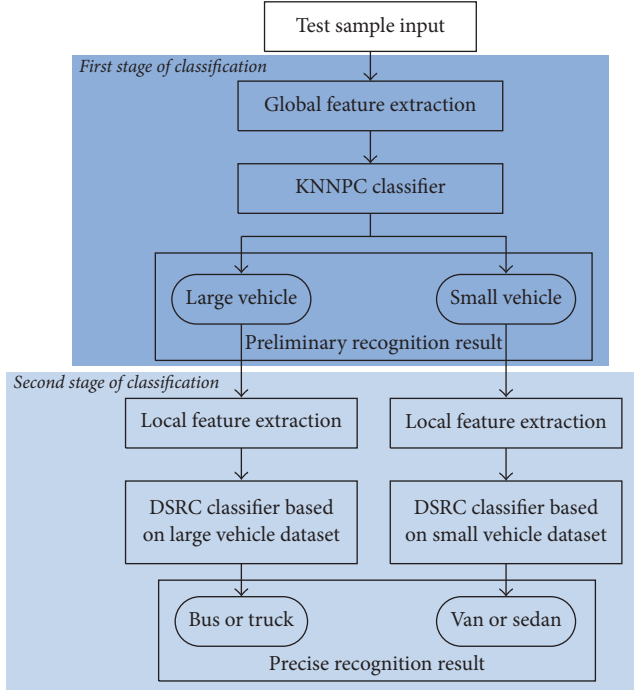


FIGURE 2: Two-stage classification strategy.

where $h(A, B)$ represents the directed distance between two sets A and B . The distance value of point a to the set B is defined as $d_B(a) = \min_{b \in B} \|a - b\|$ and the directed distance $h(A, B)$ is denoted by

$$h(A, B) = \max_{a \in A} d_B(a), \quad (11)$$

where $\|\cdot\|$ represents Euclidean norm.

Because the classical HD measure is sensitive to noises and partial occlusion, the scheme of the least trimmed square (LTS) is introduced. In the IHDM, the directed distance $h_{LTS}(A, B)$ is defined by a linear combination of order statistics:

$$h_{LTS}(A, B) = \frac{1}{K_H} \sum_{i=1}^{K_H} d_B(a)_{(i)}, \quad (12)$$

where $d_B(a)_{(i)}$ represents the i th distance value in the sorted sequence $(d_B(a)_{(1)} \leq d_B(a)_{(2)} \leq \dots \leq d_B(a)_{(N_A)})$; $K_H = f \times N_A$. A parameter f , $0 \leq f \leq 1$, depends on the amount of occlusion. The measure $h_{LTS}(A, B)$ is minimized by keeping the smaller K_H distance values after large distance values are eliminated.

3.2.2. Kernel Density Estimation. Assume that the number of the target classes is M_E , and for each class there are $n_E^{(j)}$ ($j = 1, 2, \dots, M_E$) samples. First, we obtain the K -nearest neighbors to the test sample in training set using the proposed IHDM. Suppose that $a_E(x, y)$ is the point set that consists of the edge points extracted from the test sample by the global feature extraction method proposed in Section 2.1, $b_E^{(i)}(x, y)$ indicates the point set that consists of the edge

points extracted from the i th training sample in the sample set B_E by the global feature extraction method proposed in Section 2.1, and $B_E = \{b_E^{(1)}(x, y), b_E^{(2)}(x, y), \dots, b_E^{(N_E)}(x, y)\}$, $N_E = \sum_{j=1}^{M_E} n_E^{(j)}$. According to (12), we can calculate the Hausdorff distance between $a_E(x, y)$ and every $b_E^{(i)}(x, y)$, defined as $h_{LTS}(i)$, $i \in \{1, 2, \dots, N_E\}$. Compare $h_{LTS}(i)$; we can obtain the smallest K values of $h_{LTS}(i)$, defined as $\tilde{h}_{LTS}(i)$, $i \in \{1, 2, \dots, K\}$. The K training samples corresponding to the smallest K values will be regarded as the K -nearest neighbors $\{\tilde{b}_E^{(i)} \mid i = 1, 2, \dots, K\}$ to the test sample.

Then, the KDE method [26] is used to estimate the cumulative influences on $a_E(x, y)$ from its K -nearest neighbors corresponding to different classes. We use Gaussian kernel function and set window width parameter $w_H = \max_{i \in \{1, 2, \dots, K\}} \tilde{h}_{LTS}(i) / L_H$ in the estimation, where L_H is a coefficient, to narrow (larger L_H) or expand (smaller L_H) the influences of the neighbors with different distances. Finally, we get

$$\begin{aligned} \omega_j(a_E(x, y)) \\ = \frac{1}{\sqrt{2\pi}w_H K} \sum_{l | \tilde{b}_E^{(l)} \in j} \exp\left(-\frac{(\tilde{h}_{LTS}(l))^2}{2(w_H)^2}\right), \end{aligned} \quad (13)$$

where $\omega_j(a_E(x, y))$ is the weight of $a_E(x, y)$ belonging to the j th class and $l | \tilde{b}_E^{(l)} \in j$ indicates that every $\tilde{b}_E^{(l)}$ belongs to the same j th class.

The final classification result is determined by

$$\text{identify}(a_E(x, y)) = \arg \max_j \{\omega_j(a_E(x, y))\}. \quad (14)$$

3.3. Precise Recognition Based on Local Feature and DSRC. To exploit the Gabor feature of vehicle image, before the following precise recognition, we need to firstly express all samples using their reduced Gabor feature vector $\chi^{(\rho)}$ that is computed by the proposed local feature extraction method in Section 2.2. Then, based on the reduced Gabor feature vectors, we set up training set and test set to design the DSRC.

The core idea of the sparse representation based classification (SRC) methods is to represent a test sample using a sparse linear combination of training samples [27]. Suppose that there are C classes of samples, and let $A = [A_1, A_2, \dots, A_C]$ be the set of training samples, called dictionary, where A_i is the subset of training samples from class i . Let y be a test sample. The procedures of the SRC are summarized as follows.

(i) Sparsely represent y on A via l_1 -minimization:

$$\hat{\alpha} = \arg \min_{\alpha} \{\|y - A\alpha\|_2^2 + \gamma \|\alpha\|_1\}, \quad (15)$$

where γ is a scalar constant.

(ii) Implement classification via

$$\text{identity}(y) = \arg \min_i \{e_i\}, \quad (16)$$

where $e_i = \|y - A_i \hat{\alpha}_i\|_2$ and $\hat{\alpha} = [\hat{\alpha}_1; \hat{\alpha}_2; \dots; \hat{\alpha}_C]$ and $\hat{\alpha}_i$ is the coefficient vector associated with the class i . Obviously, the SRC method classifies the test sample as the category to which the smallest representation residual e_i belongs.

Poststudies find that the employed dictionary plays an important role in sparse representation based image classification. While learning a dictionary from the training data has led to state-of-the-art results in image classification, many models of dictionary learning harness only the one-sided discriminative information in either the representation coefficients or the representation residual, which limits their performance. In this paper, we proposed a DSRC that adopts a novel dictionary learning scheme based on Fisher discrimination criterion. Based on this, a structured dictionary, whose atoms have correspondences to the subject class labels, is learned, by which both the representation residual and representation coefficients can be used to distinguish different classes.

3.3.1. Dictionary Learning Based on Fisher Discrimination Criterion. Unlike the method based on the shared dictionary, we adopt a new dictionary learning scheme based on Fisher discrimination criterion [17], which learns a structured dictionary $D = [D_1, D_2, \dots, D_{C_G}]$, where D_i is the subdictionary associated with class i . Let $G = [G_1, G_2, \dots, G_{C_G}]$ express the set of training samples with C_G classes, and let X be the sparse coefficient matrix of G over D ; that is, $G \approx DX$, where G_i is the i th subset of class i . We can write X as $X = [X_1, X_2, \dots, X_{C_G}]$, where X_i is the coefficient matrix of G_i over D . Besides requiring that D should have powerful ability to represent G (i.e., $G \approx DX$), we also require that D should have powerful ability to distinguish the images in D . For this reason, the dictionary learning scheme based on Fisher discrimination criterion is defined as follows:

$$J_{(D,X)} = \arg \min_{(D,X)} \{r(G, D, X) + \lambda_1 \|X\|_1 + \lambda_2 f(X)\} \quad (17)$$

$$\text{s.t. } \|d_n\|_2 = 1, \quad \forall n,$$

where $r(G, D, X)$ is the discriminative data fidelity term; $\|X\|_1$ is the sparsity penalty; $f(X)$ is a discrimination term imposed on the coefficient matrix X ; and λ_1 and λ_2 are scalar parameters. Each atom d_n of D is constrained to have a unit l_2 -norm to avoid that D has arbitrarily large l_2 -norm, resulting in trivial solutions of the coefficient matrix X . Further, by means of the Fisher discrimination criterion, $r(G, D, X)$ and $f(X)$ are defined as $r(G, D, X) = \sum_{i=1}^{C_G} r(G_i, D, X_i)$ and $f(X) = \text{tr}(S_W(X) - S_B(X) + \eta \|X\|_F^2)$, where $\text{tr}(\cdot)$ denotes the trace of a matrix, $S_W(X)$ and $S_B(X)$ indicate the within-class scatter and between-class scatter of X , respectively, $S_W(X) = \sum_{i=1}^{C_G} \sum_{x_k \in X_i} (x_k - m_i)(x_k - m_i)^T$, $S_B(X) = \sum_{i=1}^{C_G} n_i (m_i - m)(m_i - m)^T$, where m_i and m are the mean vectors of X_i and X , respectively, and n_i is the number of samples in class G_i ; η is a parameter.

Although the objective function $J_{(D,X)}$ in (17) is not jointly convex to (D, X) , we will find that it is convex with respect to each of D and X when the other is fixed. Therefore, the objective function $J_{(D,X)}$ can be divided into two subproblems

by optimizing D and X alternatively: updating X with D fixed and updating D with X fixed. The alternative optimization is iteratively implemented to find the desired dictionary D and coefficient matrix X .

Suppose that the dictionary D is fixed, and then the objective function in (17) is reduced to a sparse representation problem to compute $X = [X_1, X_2, \dots, X_{C_G}]$. We can compute X_i class by class. When computing X_i , all X_j , $j \neq i$, are fixed. The objective function in (17) is further simplified into

$$\min_{X_i} \{r(G_i, D, X_i) + \lambda_1 \|X_i\|_1 + \lambda_2 f_i(X_i)\}, \quad (18)$$

where $f_i(X_i) = \|X_i - M_i\|_F^2 - \sum_{k=1}^{C_G} \|M_k - M\|_F^2 + \eta \|X_i\|_F^2$; M_k and M are the mean vector matrices (by taking the mean vector m_k or m as all the column vectors) of class k and all classes, respectively. We can solve (18) to obtain X_i using the improved iterative projection method (IPM) [28].

Then we will discuss how to update $D = [D_1, D_2, \dots, D_{C_G}]$, when X is fixed. We also update $D_i = [d_1, d_2, \dots, d_{p_i}]$ class by class. That is, when every D_i is updated, all D_j , $j \neq i$, are fixed. The objective function in (17) is reduced to

$$\min_{D_i} \left\{ \|\widehat{G} - D_i X_i\|_F^2 + \|G_i - D_i X_i\|_F^2 + \sum_{j=1, j \neq i}^{C_G} \|D_i X_j^i\|_F^2 \right\} \quad (19)$$

$$\text{s.t. } \|d_l\|_2 = 1, \quad l = 1, 2, \dots, p_i,$$

where $\widehat{G} = G - \sum_{j=1, j \neq i}^{C_G} D_j X_j^j$, X_i^i is the representation matrix of G over D_i , and X_j^i is the representation of G_j over subdictionary D_j . Equation (19) can be efficiently solved to obtain every D_i via the algorithm like [29].

3.3.2. Classification Scheme. Using the dictionary D obtained by the proposed dictionary learning scheme based on Fisher discrimination criterion to represent the test sample, both the representation residual and the representation coefficients will be discriminative, and hence we can make use of both of them to achieve more accurate classification results.

Let $g = \chi(y)$ express the reduced Gabor feature vector $\chi^{(p)}$ of the test sample y ; then sparsely represent g on D via l_1 -minimization:

$$\hat{\alpha} = \arg \min_{\alpha} \{\|g - D\alpha\|_2^2 + \gamma \|\alpha\|_1\}, \quad (20)$$

where γ is a constant, $\hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_{C_G}]$, and $\hat{\alpha}_i$ is the coefficient subvector associated with subdictionary D_i .

By considering the discrimination capability of both representation residual and representation vector, we define the following metric for classification:

$$e_i = \|g - D_i \hat{\alpha}_i\|_2^2 + \omega \cdot \|\hat{\alpha} - m_i\|_2^2, \quad (21)$$

where ω is a preset weight to balance the contribution of the two terms to classification. The classification rule is defined as

$$\text{identity}(g) = \arg \min_i \{e_i\}. \quad (22)$$

4. Experiments

4.1. Experiment Setup. To validate the proposed method, we constructed a dataset including 6,000 vehicle images. The vehicle images are captured by a camera fixed on an overpass with 640×480 pixels and 256 gray scale levels. The proportion of the challenging vehicle images that are partially occluded by other vehicles or captured in a bad illumination condition is about 10% in the whole dataset. The location of each vehicle is adjusted to the center of the whole image and the size is cropped into 96×96 pixels by manual operations in advance. Figure 3 shows the example images of the dataset under various conditions.

To facilitate the VTR, all vehicle images in the whole dataset are firstly divided into two datasets: large vehicle and small vehicle. The large vehicle dataset consists of two subdatasets: bus and truck. The small vehicle dataset consists of two subdatasets: van and sedan. The numbers of the images in every subdataset are all 1,500.

All the experiments are conducted on the computer with 3 GHz CPU and 16 Gb memory, and all program codes are compiled and run on Matlab 2014b.

4.2. Results of Global Feature Extraction. In order to verify the advantage of the improved Canny operator, the edge detection results based on other three operators such as Sobel, Roberts, and Prewitt are compared in Figure 4. As can be seen from Figure 4, the proposed method based on the improved Canny operator in Section 2.1 can obtain a more accurate and complete edge compared to the methods based on three other operators.

In addition, we compare the global feature extraction method based on the improved Canny operator with the method based on traditional Canny operator. Comparative results are shown in Figure 5, where original gray images are in the first column, the detection results based on traditional Canny operator are in the second column, and the detection results based on the improved Canny operator are in the third column. Additionally, in order to verify the performance of the proposed global feature extraction method under various illumination, Figure 5(a) is captured in the morning in a fine day with good illumination, Figures 5(b) and 5(c) are captured at dusk in a cloudy day, and Figure 5(d) is captured in the afternoon in a fine day, but the bus is partially covered by shadow for the lighting is shielded by a building nearby. As can be seen from Figure 5, we can find that the method based on the improved Canny operator can obtain a more continuous and complete edge with respect to different kinds of vehicles compared to the method based on traditional Canny operator, even though the illumination condition was poor.

4.3. Results of Local Feature Extraction. Based on the method proposed in Section 2.2.2, we use the Gabor wavelet kernels with five different scales and eight different orientations to extract the Gabor feature of every local patch of the detected vehicle image. Take the patch of the hood as an example, the extracted Gabor feature image by a set of Gabor wavelet kernels with five different scales and eight different orientations is shown in Figure 6.

As can be seen from Figure 6, the feature extraction method based on the Gabor wavelet kernels can extract many structural details of local patch of vehicle image from multiple scales and multiple orientations, and the extracted Gabor feature images can be regarded as local feature for the VTR.

In the paper, the resolution of every patch is defined as 96×24 pixels. After implementing the convolution operation, the dimension of augmented feature vector χ will reach 92160 ($40 \times 96 \times 24$). The increased dimension will result in slow computation speed and large memory occupation, which will be adverse to the following recognition and classification. Therefore, before implementing the VTR, we need to down-sample χ using an appropriate sample factor ρ . In order to select an appropriate sample factor, we experiment on the augmented Gabor feature vector $\chi^{(\rho)}$ defined in Section 2.2.2 with five different downsampling factors, respectively: $\rho = 16, 32, 64, 128$, or 256 . Experimental results show that the average accuracy rates based on the DSRC proposed in Section 3.3 are 95.8%, 95.9%, 95.9%, 96.8%, 73%, and 34%, respectively, when $\rho = 1, 16, 32, 64, 128$, or 256 . It is very clear that when $\rho = 64$, the DSRC has the highest accuracy rate. Therefore, in this paper, we let $\rho = 64$, and the dimension of the augmented Gabor feature vector is reduced to 1440 ($40 \times 12 \times 3$) accordingly, which will reduce the computational complexity of VTR on the premise to assure a high recognition accuracy.

4.4. Results of Two-Stage Classification. In order to demonstrate the performance of the proposed two-stage classification strategy, we introduce three evaluation criteria: precision, recall, and accuracy [30]. Their definitions are as follows: precision = $TP/(TP + FP)$, recall = $TP/(TP + FN)$, and accuracy = $(TP + TN)/(TP + FN + FP + TN)$, where TP, FP, FN, and TN are abbreviations for true positives, false positives, false negatives, and true negatives, respectively.

We randomly select 400 samples as training samples and 400 samples as test samples from four vehicle type datasets, bus, truck, van, and sedan, respectively.

4.4.1. Results of the First Stage of Classification. For the first stage of classification, we experiment on the whole dataset. We randomly select 1200 samples as training samples and 400 samples as test samples. If the type of the test sample is recognized as bus or truck, then the test sample is determined as a large vehicle. Similarly, if the type of the test sample is recognized as van or sedan, then the test sample is determined as a small vehicle. Table 1 shows the experimental results where the test samples are captured under good illumination and no occlusion. Further, Table 2 gives the results under bad illumination or partial occlusion.

As can be seen from Tables 1 and 2, the first stage of classification still has high accuracy and reliability, even though the test samples are captured under bad illumination or partial occlusion.

4.4.2. Results of the Second Stage of Classification. Based on the result of the first stage of classification, if the test sample is recognized as a large vehicle, the large vehicle dataset



(a) Fine day



(b) Partially occluded



(c) Rainy day



(d) Dusk and night

FIGURE 3: Example images under various conditions.

including the bus and truck images needs to be used in the following second stage of classification. Similarly, if the test sample is recognized as a small vehicle, the small vehicle dataset including the van and sedan images needs to be used.

We still randomly select 1200 samples as training samples and 400 samples as test samples from the large vehicle dataset or small vehicle dataset in the second stage of classification. Table 3 shows the experimental results where the test samples

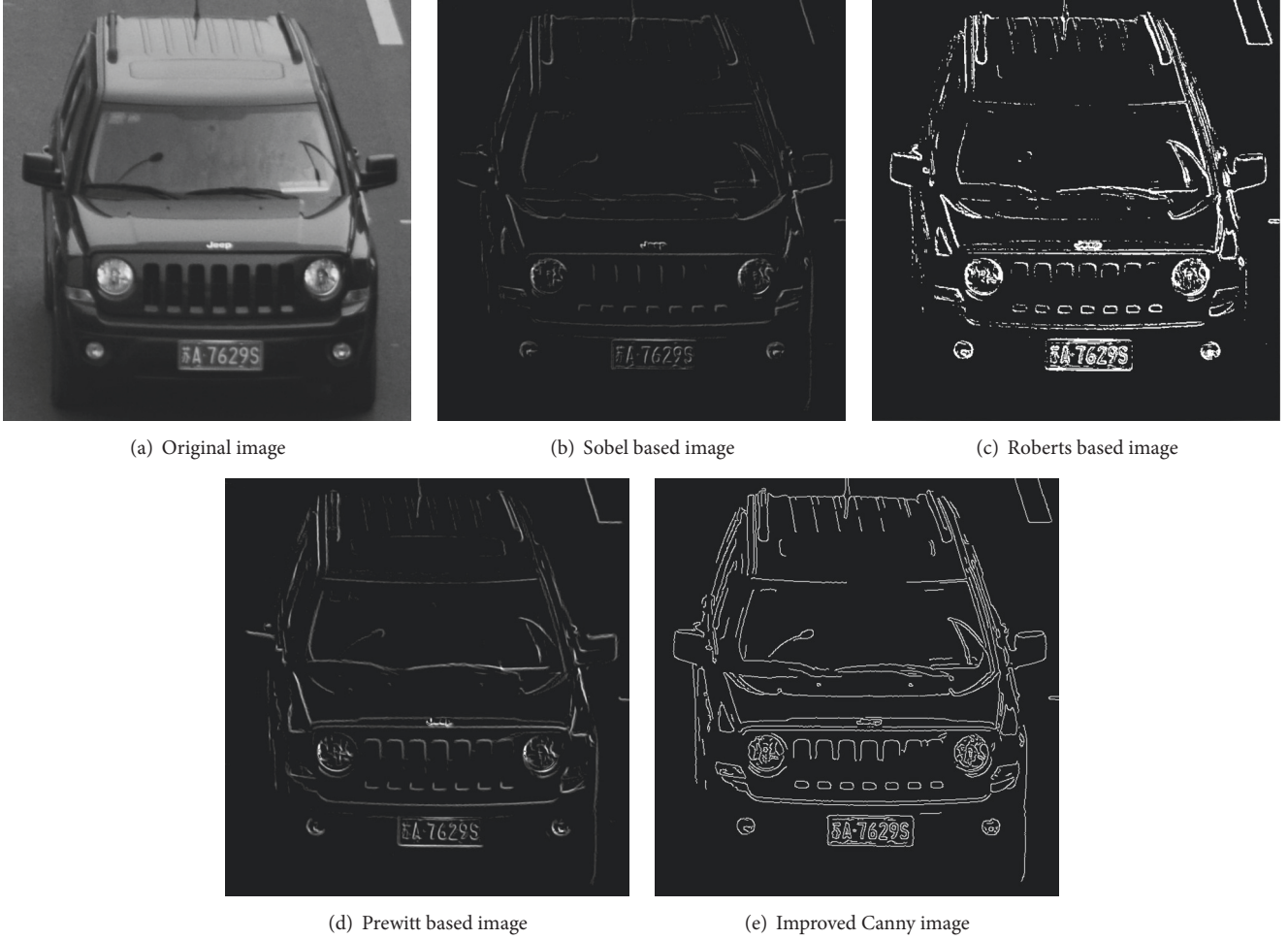


FIGURE 4: Edge detection results based on improved Canny operator and other operators.

are captured under good illumination or no occlusion. Table 4 gives the results under bad illumination or partial occlusion.

As can be seen from Tables 3 and 4, although the performance of the second stage of classification slightly degrades compared with the first stage of classification, it still has very good reliability.

To verify that the proposed method exploiting the dictionary learning scheme based on Fisher discrimination criterion is effective, after implementing the first stage of classification, we use the traditional SRC method that does not exploit the dictionary learning scheme based on Fisher discrimination criterion to implement the second stage of classification. The classification results under good illumination and no occlusion are shown in Table 5.

As can be seen from Tables 3 and 5, the proposed classification method that exploits the dictionary learning scheme based on Fisher discrimination criterion is superior to the traditional method in terms of precision, recall, and accuracy. Therefore, exploiting the dictionary learning scheme based on Fisher discrimination criterion in the second stage of classification is very effective for improving recognition performance of classifier for the VTR.

In order to demonstrate the efficacy of the two-stage classification strategy, the proposed KNNPC in Section 3.2 and the DSRC in Section 3.3 are regarded as single-stage classifiers to implement the classification task of four types of vehicles, respectively. We also randomly select 1200 samples as training samples and 400 samples as test samples from the whole dataset. The results of single-stage classification based on the KNNPC and global feature and those based on the DSRC and local feature are shown in Tables 6 and 7, respectively. It is clearly noted that the proposed two-stage classification strategy overpasses the single-stage classification strategy in terms of precision, recall, and accuracy. Further analysis finds out that the extracted global feature has an excellent ability to distinguish the large vehicles from small vehicles or to distinguish the small vehicles from large vehicles based on the KNNPC. When the four types of vehicles are mixed together, it becomes difficult for the global feature to distinguish the buses or trucks in the large vehicle dataset or distinguish the vans or sedans in the small vehicle dataset. Moreover, when the four types of vehicles are mixed together, the single-stage classification based on the DSRC and local feature needs to train more classifier parameters simultaneously using more

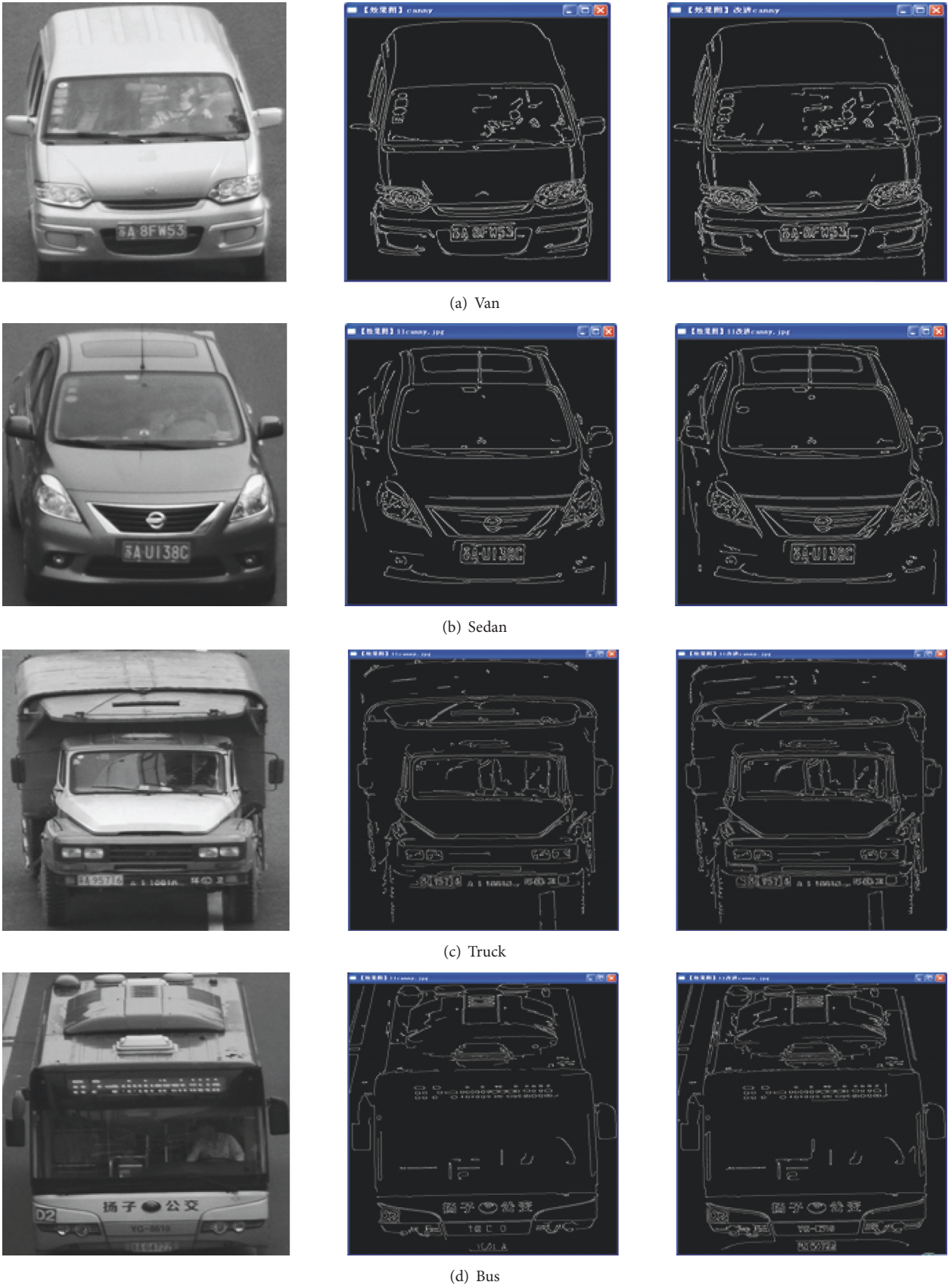


FIGURE 5: Global feature extraction of four types of vehicles based on traditional and improved Canny operators under various illumination.

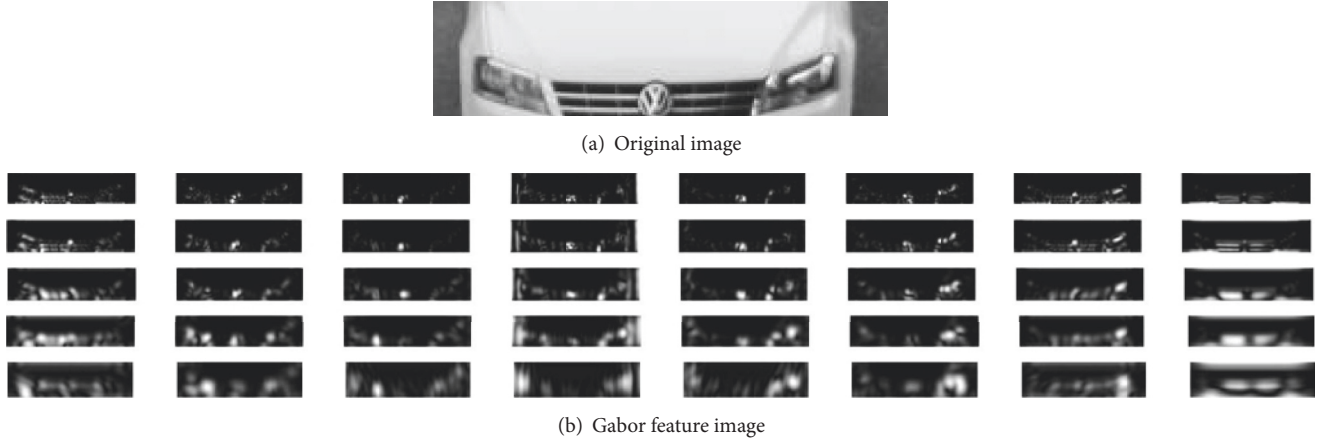


FIGURE 6: Extracted Gabor feature image.

TABLE 1: Results of first stage of classification under good illumination and no occlusion.

| Vehicle type | Precision | Recall | Accuracy |
|---------------|-----------|--------|----------|
| Large vehicle | 98.2% | 96.9% | 98.7% |
| Small vehicle | 98.1% | 97.2% | 98.5% |

TABLE 2: Results of first stage of classification under bad illumination or partial occlusion.

| Vehicle type | Precision | Recall | Accuracy |
|---------------|-----------|--------|----------|
| Large vehicle | 91.6% | 90.8% | 91.7% |
| Small vehicle | 91.3% | 90.6% | 91.1% |

TABLE 3: Results of second stage of classification under good illumination and no occlusion.

| Vehicle type | Precision | Recall | Accuracy |
|--------------|-----------|--------|----------|
| Bus | 96.1% | 96.2% | 96.4% |
| Truck | 96.7% | 95.9% | 96.6% |
| Van | 96.1% | 95.8% | 96.3% |
| Sedan | 95.6% | 96.3% | 96.2% |

TABLE 4: Results of second stage of classification under bad illumination or partial occlusion.

| Vehicle type | Precision | Recall | Accuracy |
|--------------|-----------|--------|----------|
| Bus | 88.3% | 87.7% | 87.3% |
| Truck | 91.2% | 89.3% | 90.9% |
| Van | 89.1% | 90.1% | 89.5% |
| Sedan | 88.2% | 87.3% | 87.6% |

training samples than when two types of vehicles are mixed together for a given recognition performance. Therefore, the performance of the single-stage classification based on the DSRC and local feature will degrade compared with the proposed two-stage classification strategy.

TABLE 5: Results of second stage of classification without the dictionary learning scheme based on Fisher discrimination criterion.

| Vehicle type | Precision | Recall | Accuracy |
|--------------|-----------|--------|----------|
| Bus | 90.8% | 91.6% | 91.1% |
| Truck | 91.3% | 90.8% | 91.7% |
| Van | 90.7% | 91.5% | 91.3% |
| Sedan | 90.8% | 91.1% | 90.6% |

TABLE 6: Results of single-stage classification based on the KNNPC and global feature.

| Vehicle type | Precision | Recall | Accuracy |
|--------------|-----------|--------|----------|
| Bus | 88.8% | 88.3% | 88.9% |
| Truck | 88.8% | 88.7% | 88.2% |
| Van | 88.1% | 87.8% | 87.9% |
| Sedan | 88.0% | 87.7% | 87.6% |

TABLE 7: Results of single-stage classification based on the DSRC and local feature.

| Vehicle type | Precision | Recall | Accuracy |
|--------------|-----------|--------|----------|
| Bus | 92.1% | 93.2% | 92.8% |
| Truck | 92.3% | 92.8% | 92.5% |
| Van | 91.8% | 91.6% | 92.1% |
| Sedan | 91.3% | 90.8% | 91.2% |

4.5. Comparison of Results with Other Methods. In order to compare our method with other popular methods, we test our method on the dataset used in [31]. Similar to [31], the experiments on daylight images and nighttime images are performed, respectively. Before implementing the classification, we firstly divide the dataset in [31] into two categories: large vehicle dataset and small vehicle dataset, where large vehicle dataset consists of two types of vehicles, bus and truck, and small vehicle dataset consists of three types of vehicles, passenger car, minivan, and sedan. Our method averagely achieves 96.3% classification accuracy on daylight images and 89.5% on nighttime images, better than the results of previous

TABLE 8: Comparison between our method's results and other methods' results.

| Methods | Accuracy | |
|--------------------------|----------|-----------|
| | Daylight | Nighttime |
| Psylos et al. [32] | 78.3% | 73.3% |
| Petrovic and Cootes [33] | 84.3% | 82.7% |
| Peng et al. [31] | 90.0% | 87.6% |
| Dong and Jia [8] | 91.3% | — |
| Dong et al. [1] | 96.1 | 89.4 |
| Ours | 96.3% | 89.7% |

methods, as demonstrated in Table 8. Additionally, we also test our method on the BIT-Vehicle dataset provided in [1]; our method achieves 90.1% classification accuracy, yet the accuracy of the method used in [1] reaches 88.11%.

The underlying reasons are as follows: the proposed Canny edge operator and Gabor wavelet kernels are able to extract discriminative global and local features for VTR. The proposed two-stage classification strategy can leverage the advantages of the extracted global and local features according to their characteristics; that is, the extracted global feature that can represent the geometrical contour of a vehicle is just applied to the first stage of classification to determine whether the test sample belongs to large vehicle or small vehicle, and then the local feature that can represent the structural details of a vehicle is just applied to the second stage of classification to determine whether the sample belongs to bus or truck in the large vehicle dataset as well as van or sedan in the small vehicle dataset. The dictionary learning scheme based on Fisher discrimination criterion is able to learn a discriminative classifier for precision recognition in the second stage of classification. Extracting local feature from the four partitioned patches enables strong robustness to partial occlusion.

5. Conclusions

The two key steps of improving the VTR are the feature extraction and classifier design. Based on the need to recognize the vehicle type accurately and reliably, we propose a VTR method combining global and local features via two-stage classification. The improved Canny edge detection algorithm is capable of extracting the continuous and complete global feature. The employed Gabor wavelet kernels with five scales and eight orientations are able to successfully extract the local feature. The proposed KNNPC is able to realize the preliminary recognition of a large vehicle or small vehicle based on the global feature. Further, the DSRC has a stronger ability in recognizing bus, truck, van, or sedan based on the local feature. As demonstrated by the experiments on the challenging dataset and a compared dataset, the proposed method can solve the VTR problem much more efficiently and outperforms existing state-of-the-art methods.

The study offers the possibility of developing more sophisticated VTR methods. First, this method can be extended to the VTR context involving more vehicle types.

Second, more effective features and corresponding feature extraction algorithms can be adopted. Third, more discriminative classifiers can be incorporated into the two-stage classification.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (nos. 61304205, 61502240, 61203273, and 41301037), Natural Science Foundation of Jiangsu Province (no. BK20141002), and Innovation and Entrepreneurship Training Project of College Students (nos. 201710300051 and 201710300050).

References

- [1] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2247–2256, 2015.
- [2] B. Lin, Y. Lin, L. Fu et al., "Integrating appearance and edge features for sedan vehicle detection in the blind-spot area," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 737–747, 2012.
- [3] F. M. D. S. Matos and R. M. C. R. De Souza, "An image vehicle classification method based on edge and PCA applied to blocks," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC '12)*, pp. 1688–1693, Seoul, South Korea, October 2012.
- [4] H.-Z. Gu and S.-Y. Lee, "A view-invariant and anti-reflection algorithm for car body extraction and color classification," *Multimedia Tools and Applications*, vol. 65, no. 3, pp. 387–418, 2013.
- [5] M. Rezaei, M. Terauchi, and R. Klette, "Robust vehicle detection and distance estimation under challenging lighting conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2723–2743, 2015.
- [6] S. Kamkar and R. Safabakhsh, "Vehicle detection, counting and classification in various conditions," *IET Intelligent Transport Systems*, vol. 10, no. 6, pp. 406–413, 2016.
- [7] R. K. Satzoda and M. M. Trivedi, "Multipart vehicle detection using symmetry-derived analysis and active learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 926–937, 2016.
- [8] Z. Dong and Y. Jia, "Vehicle type classification using distributions of structural and appearance-based features," in *Proceedings of the 20th IEEE International Conference on Image Processing (ICIP '13)*, pp. 4321–4324, Melbourne, VIC, Australia, September 2013.
- [9] A. Ambardekar, M. Nicolescu, G. Bebis, and M. Nicolescu, "Vehicle classification framework: a comparative study," *Eurasip Journal on Image and Video Processing*, vol. 2014, no. 1, article 29, 2014.
- [10] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "A hybrid vehicle detection method based on viola-jones and HOG + SVM from UAV images," *Sensors*, vol. 16, no. 8, article 1325, 2016.

- [11] A. Nurhadiyatna, A. L. Latifah, and D. Fryantoni, "Gabor filtering for feature extraction in real time vehicle classification system," in *Proceedings of the 9th International Symposium on Image and Signal Processing and Analysis (ISPA '15)*, pp. 19–24, Zagreb, Croatia, September 2015.
- [12] J. Kim, J. Baek, Y. Park, and E. Kim, "New vehicle detection method with aspect ratio estimation for hypothesized windows," *Sensors*, vol. 15, no. 12, pp. 30927–30941, 2015.
- [13] W. Zhang, Q. Wang, and C. Suo, "A novel vehicle classification using embedded strain gauge sensors," *Sensors*, vol. 8, no. 11, pp. 6952–6971, 2008.
- [14] J. Fang, Y. Zhou, Y. Yu, and S. D. Du, "Fine-grained vehicle model recognition using a coarse-to-fine convolutional neural network architecture," *IEEE Intelligent Transportation Systems Society*, vol. 99, pp. 1–11, 2016.
- [15] X. Chen, R.-X. Gong, L.-L. Xie, S. Xiang, C.-L. Liu, and C.-H. Pan, "Building regional covariance descriptors for vehicle detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 4, pp. 524–528, 2017.
- [16] Y. Gao, J. Ma, and A. L. Yuille, "Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples," *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2545–2560, 2017.
- [17] R. Jiang, H. Qiao, and B. Zhang, "Efficient fisher discrimination dictionary learning," *Signal Processing*, vol. 128, pp. 28–39, 2016.
- [18] C. Mi, Z. Zhang, X. He, Y. Huang, and W. Mi, "Two-stage classification approach for human detection in camera video in bulk ports," *Polish Maritime Research*, vol. 22, no. 1, pp. 163–170, 2015.
- [19] G. Abdel-Azim, S. Abdel-Khalek, and A. S. Obada, "A novel edge detection algorithm for image based on non-parametric Fisher information measure," *Applied and Computational Mathematics*, vol. 14, no. 3, pp. 316–327, 2015.
- [20] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [21] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [22] G. Donate, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–989, 1999.
- [23] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A*, vol. 4, no. 12, p. 2379, 1987.
- [24] T. Acharya and A. K. Ray, *Image Processing: Principles and Applications*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2005.
- [25] A. A. Taha and A. Hanbury, "An efficient algorithm for calculating the exact hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 11, pp. 2153–2163, 2015.
- [26] X. Tang and A. Xu, "Multi-class classification using kernel density estimation on K-nearest neighbours," *IEEE Electronics Letters*, vol. 52, no. 8, pp. 600–602, 2016.
- [27] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [28] M. Sadeghi and M. Babaie-Zadeh, "Iterative sparsification-projection: fast and robust sparse signal approximation," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5536–5548, 2016.
- [29] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Metaface learning for sparse representation based face recognition," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 1601–1604, Hong Kong, China, September 2010.
- [30] D. Olson and D. Delen, *Advanced Data Mining Techniques*, Springer, Berlin, Germany, 2008.
- [31] Y. Peng, J. S. Jin, S. Luo, M. Xu, and Y. Cui, "Vehicle type classification using PCA with self-clustering," in *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW '12)*, pp. 384–389, Melbourne, VIC, Australia, July 2012.
- [32] A. Psyllos, C. N. Anagnostopoulos, and E. Kayafas, "Vehicle model recognition from frontal view image measurements," *Computer Standards & Interfaces*, vol. 33, no. 2, pp. 142–151, 2011.
- [33] V. S. Petrovic and T. Cootes, "Analysis of features for rigid structure vehicle type recognition," in *Proceedings of the British Machine Vision Conference*, 10 pages, London, UK, September 2004.

