Reliable Classification of Vehicle Types Based on Cascade Classifier Ensembles

Bailing Zhang

***Abstract*—Vehicle-type recognition based on images is a challenging task. This paper comparatively studied two feature extraction methods for image description, i.e., the Gabor wavelet transform and the Pyramid Histogram of Oriented Gradients (PHOG). The Gabor transform has been widely adopted to extract image features for various vision tasks. PHOG has the superiority in its description of more discriminating information. A highly reliable classification scheme was proposed by cascade classifier ensembles with reject option to accommodate the situations where no decision should be made if there exists adequate ambiguity. The first ensemble is heterogeneous, consisting of several classifiers, including *k*-nearest neighbors (kNNs), multiple-layer perceptrons (MLPs), support vector machines (SVMs), and random forest. The classification reliability is further enhanced by a second classifier ensemble, which is composed of a set of base MLPs coordinated by an ensemble metalearning method called rotation forest (RF). For both of the ensembles, rejection option is accomplished by relating the consensus degree from majority voting to a confidence measure and by abstaining to classify ambiguous samples if the consensus degree is lower than a threshold. The final class label is assigned by dual majority voting from the two ensembles. Experimental results using more than 600 images from a variety of 21 makes of cars and vans demonstrated the effectiveness of the proposed approach. The cascade ensembles produce consistently reliable results. With a moderate ensemble size of 25 in the second ensemble, the two-stage classification scheme offers 98.65% accuracy with a rejection rate of 2.5%, exhibiting promising potential for real-world applications.**

***Index Terms*—Classification reliability, classification with rejection, Gabor transform, Pyramid Histogram of Oriented Gradients (PHOG), rotation forest (RF) ensemble, vehicle-type classification.**

I. Introduction

VISION-BASED vehicle make-and-model recognition (MMR) is an important part of an intelligent traffic system with a multitude of applications. For example, for vehicle surveillance for high-security areas, conventional number-plate recognition systems have to be augmented as a number plate may be faked or tampered with. An MMR system would offer valuable assistance to the police in identifying blacklisted vehicles at toll stations or in their search for specific suspect vehicles from the traffic surveillance image database [1], [2]. With the increasing demand for security awareness and widespread use of surveillance cameras, the need for vehicle identification and classification technologies has become ever more relevant in the recent years [3]–[5].

Image-based vehicle make recognition is a challenging task with many issues that need to be addressed [3]. For example, the number of vehicle models is constantly expanding, and the similarities between certain models are very high. There is also a large degree of variations in the obtained vehicle images due to the ever changing road environments, backgrounds, illuminations, and motion-caused variability. Although the classification of vehicle types has been a subject of interest for several years, most of the published studies mainly focused on the detection of the presence of a vehicle [6] or the classification of vehicles into broad categories such as cars, buses, heavy goods vehicles, etc [3], [4], [7], [8], [12], [13]. Several papers published recently have applied different feature analysis and used advanced machine learning algorithms to classify vehicles into a predefined and limited number of fine classes [9], [10], For example, certain edge-based features, including Sobel edge response and square mapped gradients, have been experimented [9] together with simple Euclidean measure-based decision module, yielding some interesting results. An oriented-contour point model was proposed to represent a vehicle type, which exploits the edges in the four orientations of vehicle front images as features [10]. Another paper investigated the applicability of some transform-based image features in vehicle MMR, including Fourier transforms, discrete wavelet transforms, and discrete curvelet transforms [11].

Due to the limited progress made, it is still a challenging research problem to devising efficient and robust vehicle MMR methods. Vision-based vehicle-type recognition can be studied in a multiclass classification framework, in which two interweaved issues have to be tackled, i.e., feature representation and classification. An appropriate description of a vehicle image should characterize the differences between distinct makes or models. In recent years, computer vision has seen much progress in the development of various efficient image feature description methods, many of which have become "off-the-shelf" techniques applicable to general image analysis tasks. Among many possible options, we investigated a new edge descriptor, called Pyramid Histogram of Oriented Gradients (PHOG) [14]. PHOG was developed from Histogram of Oriented Gradients (HOG) [15], [16], which calculates a gradient direction and magnitude for every pixel in the image and bins these gradients by their direction weighted with magnitude.

Manuscript received August 17, 2011; revised March 20, 2012 and May 29, 2012; accepted August 9, 2012. Date of publication September 7, 2012; date of current version February 25, 2013. This work was supported by the Suzhou Municipal Science and Technology Foundation under Grant *Key Technologies for Intelligent Video Analysis for Criminal Investigation* (SS201109). The Associate Editor for this paper was K. Wang.

The author is with the Department of Computer Science and Software Engineering, Xi'an Jiaotong–Liverpool University, Suzhou 215123, China (e-mail: bailing.zhang@xjtlu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2012.2213814

The Gabor filter is another widely adopted operator for coding image local properties and has been shown to be very efficient in various image processing tasks [17], [18]. The outstanding characteristics of Gabor filtering lie in its multiresolution and multiorientation properties, which are instrumental to the measure of local spatial frequencies. The distortion tolerance performance from the Gabor transform explains many successful applications [18].

With appropriate feature descriptions, a vehicle-type recognition or identification system can be implemented using certain well-developed classification algorithms. Machine learning offers a plethora of advanced classifier paradigms. Examples include neural networks [19], [20], support vector machines (SVMs) [21], random forest [22], and classifier ensembles [23], [24]. An ensemble of classifiers can integrate multiple component classifiers, such as a decision tree or multiple-layer perceptrons (MLPs), for a single task using the same base learning algorithm [24]. To classify a new case, each member of the ensemble classifies the case independently of the rest, and the results are aggregated via an appropriate decision strategy, such as majority voting, to derive a final classification. The representatives of ensemble learning include bagging [25], AdaBoost [26], [27], and random subspace [28], [29]. Recently, a new technique for generating ensemble classifier, called rotation forest (RF), has been proposed [30], [31] based on the principal component analysis (PCA). The idea is to enhance the diversity among the component classifiers while maintaining their accuracy. Specifically, PCA is applied to each of K randomly split feature subset with all principal components retained for the preservation of the variability information in the data. Thus, K-axis rotations take place to form the new features for a base classifier, which is trained using the entire data set to avoid the sacrifice of accuracy. It has been demonstrated that this method performs much better than several other ensemble methods on some benchmark classification data sets [32].

In the majority of current image classification studies including MMR, the degree of accuracy has been the primary pursuit, aiming at producing a classifier with the smallest error rate (ER) possible. In many scenarios such as surveillance, however, it is more crucial to address the reliability issue in classifier design by introducing reject options, which allow for abstaining from uncertain instance classifications via an extra decision expressing doubt. For instance, if the police needs to identify a blacklisted vehicle at a checkpoint, it is more useful to be able to reject a vehicle image for automatic classification when there is no sufficiently high degree of confidence to make a decision since the cost of misclassification is high, and the rejected instance can be relayed to police manpower for closer examination. Reliable classification permits special domain knowledge and expertise to exert control over the accuracy of the classifier to make determinations. Although classification with rejection option has been an interesting topic in pattern recognition research for many years [33]–[37] and some paradigms have been proposed [38], [39], reliable classification is still a much less studied topic due to the extra difficulty in determining the reliability of a classification assigned to a particular instance.

Therefore, we aim to design a classification scheme with rejection option to meet this required functionality. The emphasis is on vehicle-type classification that generates decisions with confidence larger than some prescribed threshold and transfers the decision on cases with lower confidence to an expert. In other words, the objective of reject option is to improve classification reliability and leave the control of classification accuracy to human judgement. With the aforementioned PHOG and Gabor features extracted from frontal images of vehicles, a two-stage cascade classifier ensemble with rejection option was proposed for MMR. Compared with some earlier proposed cascading classifier paradigms [39], our system is distinctive in that it is composed of two different ensembles. The first-stage ensemble consists of eight heterogeneous classifiers (k-nearest neighbors (kNNs), MLPs, random forests and SVMs, using Gabor and PHOG features, respectively), which are employed to obtain high accuracy for easier inputs and reject a subset of class assignments that is harder or ambiguous. Only examples for which the stage 1 ensemble's confidence score in terms of consensus degree is greater than a certain threshold pass through the next stage of the cascade. A second-stage classifier ensemble consists of an RF ensemble of MLPs, which concentrates on the rejected samples from stage 1 during classification. The overall decision for samples below a certain confidence level is made by both stages jointly: If the decisions of the first-stage ensemble and the second-stage MLP ensemble agree, the fusion produces a decision for input examples; otherwise, the fusion rejects it again. Classification with the proposed rejection choice provides an efficient means to reduce the ER by controlling the accuracy-rejection tradeoff.

The rest of this paper is organized as follows. Section II briefly introduces the vehicle image data acquisition and preprocessing. Section III outlines the two feature extraction methods, namely the PHOG and the Gabor transform. The scheme of cascaded classifier ensembles is introduced in Section IV. Section V details the experiments and reports on the classification results, followed by the conclusion in Section VI.

II. Images Collection and Preprocessing

The local police department of Dushu Lake Higher Education Town in Suzhou provided a large collection of vehicle images recorded with their traffic surveillance cameras over a one-week period. The images were captured using charge-coupled device cameras (SP-140N) installed at ten different intersections, between 7:30 A.M. and 9:50 P.M., encompassing a wide range of weather and illumination conditions. From the recorded images ($> 50\,000$), 620 images of different vehicles from 15 brands were selected, accounting for 21 classes, including Audi, Buick (two classes), Changan, Chery (three classes), Chevrolet (two classes), Citroen, Ford, Honda, Hyundai (two classes), Mazda, Nissan, Peugeot, Volkswagen (two classes), Toyota, and Wulin. Some classes describe different models of the same brand of cars, e.g., the Hyundai Sonata and the Hyundai Elantra, when the two models have quite different appearances. If different models of a same brand shared similar appearances, e.g., Audi A7/Audi A8, Buick Excelle/Buick Xt, they would be included into the same class, although the effort to include only one model of each vehicle brand has been made in the data preparation process. All of the images contained

Fig. 1. Samples of the frontal vehicle images.

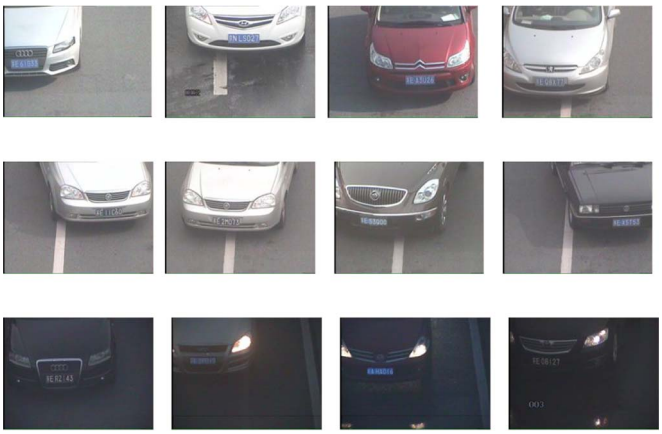
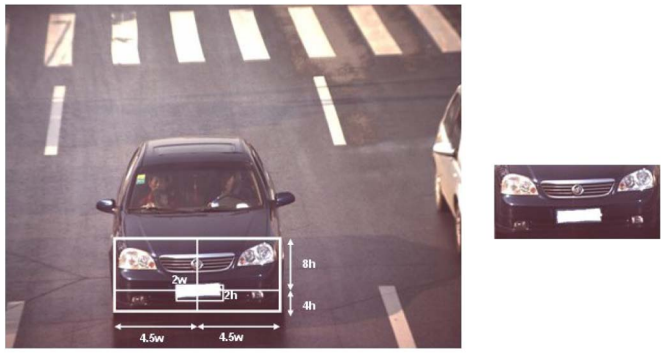


Fig. 2. Different types of image variations in the collected samples. (a) Vehicle position. (b) In-plane rotation. (c) Illumination variations.

frontal views of a single vehicle captured from variable distances. The original images have 1024×1360 color pixels. A sample of the selected images is shown in Fig. 1.

Most vehicles have different appearances, with rich edge information contained in the frontal parts. As demonstrated in Fig. 2, the vehicle images have a number of different variations, including vehicle position, scale, in-plane rotation, and illumination.

The number-plate detection and region-of-interest (ROI) measurement strategies proposed in [9] were followed in this paper to automatically extract and crop the lower part of the frontal vehicle images. Specifically, the number plate is first segmented from an image as the reference by finding a rectangular part in the black-and-white image. Based on the algorithm in [9], rectangular constellation is used to generate hypotheses for the plate location in the image, and the best corner structure fit to each of its corners is chosen from the candidates. After the number plate has been located, the algorithm will find the frontal part of the image as the ROI, which is defined in terms of number-plate width w and height h relative to its center. The segmented ROI covers the headlights, the radiator, the number plate, and the bumper, as shown in Fig. 3.



(a) (b)

Fig. 3. Example of cropping ROI from original images. (a) Original image. (b) cropped ROI.

(a) (b)

Fig. 4. Example of unsuccessful cropping from the original images. (a) Occluded multiple vehicles in an image. (b) Incomplete vehicle(s) in an image.

The ROI segmentation algorithm can successfully crop 98% of the original images. Unsuccessful situations mainly arose due to occluded multiple vehicles and incomplete vehicles, as demonstrated in Fig. 4. The unsuccessful segmentation cases will not be included in classification training/testing sets. In real applications, incomplete vehicles are often extracted. To compensate this defect, several possible schemes are under investigation, including the application of learning-based methods for vehicle detection.

III. Feature Descriptions

Although many types of features could be defined for vehicle images, geometrical features pertinent to edge are more important. Examples of such geometrical features include the square mapped gradients [9] and oriented-contour point [10]. Here, we introduce two well-developed descriptors appropriate for expressing edge-rich information.

A. Gabor Transform

Image representation using Gabor filter responses minimizes the joint space–frequency uncertainty. The filters are orientation- and scale-tunable edge and line detectors. They are widely used in image processing and computer vision. The convolution kernel of the Gabor filter is a product of a Gaussian and a cosine function, which can be characterized by a preferred orientation and a preferred spatial frequency as follows:

$$g_{\lambda,\theta,\varphi}(x,y) = \exp\left(-\frac{x'^2+\gamma y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right) \quad (1)$$

where

$$x' = x\cos\theta + y\sin\theta$$
$$y' = -x\sin\theta + y\cos\theta.$$

The standard deviation σ determines the effective size of the Gaussian signal. The eccentricity of the convolution kernel g is determined by the parameter λ, which is called the spatial aspect ratio. λ determines the frequency (wavelength) of the cosine. θ determines the direction of the cosine function, and finally, φ is the phase offset.

Typically, an image is filtered with a set of Gabor filters of different preferred orientations and spatial frequencies that cover appropriately the spatial frequency domain, and the responses obtained form a feature vector that can be used for classification. Given image $I(x, y)$, its Gabor transform is defined as

$$W_{mn}(x,y) = \int I(x_1,y_1) g^*_{mn}(x-x_1, y-y_1)\, dx_1\, dy_1 \quad (2)$$

where $*$ indicates the complex conjugate. With the assumption of spatially homogeneous local texture regions, the mean μ_{mn} and standard deviation σ_{mn} of the magnitude of transform coefficients can be used to represent the regions [18] as follows:

$$\mu_{mn} = \int\int |W_{mn}(x,y)|\, dx\, dy \quad (3)$$

$$\sigma_{mn} = \sqrt{\int\int (|W_{mn}(x,y)| - \mu_{mn})^2\, dx\, dy}. \quad (4)$$

A feature vector f (texture representation) is thus created using μ_{mn} and σ_{mn} as the feature components [18].

The Gabor feature vector contains pairs for all the scales and orientations of the transforms. From a number of experiments, we found that a filter bank with six orientations and four scales gave satisfactory classification performance, which means 24×2 component features will be extracted for a given image patch. Therefore, the figuration is applied to 2×4 nonoverlapping resized image (256×512) subregions, each with the size of 128×128, yielding an overall feature vector with length $2 \times 4 \times 48 = 384$ for each image.

B. PHOG

The edge-related attribute is critically important for describing structural information included in an image. Among various techniques to extract edge information, a local edge descriptor, i.e., HOG [15], has been successfully applied to many applications [16]. The notion of HOG is that the shape of a local object can be characterized well by the distribution of local intensity gradients or edge directions, which is similar to that of edge histograms [40] but differs in that it is computed on a dense grid of uniformly spaced cells. For a given subwindow of the image, a HOG accumulates, over each edge point, the angles of the gradients that fall into a particular range. The contribution of an edge point to the HOG is weighted by the gradient magnitude at that point [15], [16].

The HOG feature encodes the gradient orientation of one image patch without consideration of where this orientation is from in the patch. Therefore, it is not discriminative enough when the spatial property of the underlying structure of the image patch is important. The objective of the PHOG [14] is to take the spatial property of the local shape into account while representing an image by HOG, which is accomplished by tiling the image into regions at multiple resolutions and a spatial pyramid matching technique. Each image is divided into a sequence of increasingly finer spatial grids by repeatedly doubling the number of divisions along each axis direction. The number of points in a cell at one level is simply the sum over those contained in the four cells divided at the next level, thus forming a pyramid representation. The cell counts at each level of resolution are the bin counts for the histogram representing that level.

The PHOG descriptors can be compared using a histogram intersection method [14].

IV. Reliable Classification via Cascade Classifier Ensembles

Although plenty of supervised learning algorithms, such as neural networks, kNNs, and SVMs [19]–[21], has been applied to the various pattern recognition problems, few studies have addressed the issue of classification reliability, which is about the confidence in making a decision. Even if good accuracy could be achieved from some of the existed methods, they may still be insufficiently accurate to be used in practical situations such as the recognition of a suspected vehicle during traffic surveillance. High-confidence classifier is a top priority in such scenarios. On the other hand, it is often not realistic to make the assumption that all classes are known beforehand. For example, the number of vehicle types on road is difficult to estimate. Another simple case that highlights the importance of reliability is the presence of missegmented samples from the preprocessing stage (e.g., the segmentation of ROI) for most of image recognition issues. A missegmented sample can represent arbitrary fragments or combinations of proper class patterns. Therefore, in practical classification systems, it is desirable to have a reject option in addition to taking a hard decision for class labels, i.e., an option to withhold a classifier decision and transfer the ambiguity to human judgement.

A. Basic Concept Review

The concept of classification with reject option was introduced by Chow [33]. For two-class classification problems, an example is represented by n-dimensional feature vector $x \in \Re^n$ and label $y \in C$, where $C = \{-1, +1\}$. Following Bayes' theorem, posterior probability is defined as follows:

$$p(C_i|x) = \frac{p(C_i)p(x|C_i)}{p(x)}, \qquad i = 1, 2 \quad (5)$$

where $p(C_i)$ is the prior probability of class C_i, $p(x|C_i)$ is the conditional densities of x given C_i, and $p(x) = \sum_{i=1}^{2} p(x|C_i)p(C_i)$ represents the unconditional density of x.

If we use $\mathrm{risk}(x)$ to denote the risk of making a wrong decision for x, the Bayes' decision rule assigns x to the class of maximum posterior probability (5) by minimizing the error probability, which can be defined by

$$\text{error} = \int_x \text{risk}(x) p(x)\, dx. \tag{6}$$

Bayes' classifier predicts the class having the highest posterior probability and is optimal if true posterior probabilities are available. However, when some classes are not known or when classes overlap in feature space, Bayes' classifier may not be always efficient. An alternative approach is to classify only those samples for which the posterior probability is sufficiently high and reject the remaining samples. Based on this principle, Chow presented an optimal classifier with reject option [33]. A binary decision rule with reject option is optimum if, for a given ER (error probability), it minimizes the reject rate (reject probability). Chow demonstrated that the optimum rule is to reject the pattern if the maximum of the *a posteriori* probabilities is less than a certain threshold. More explicitly, an example x is accepted only if the probability that x belongs to C_i is higher than or equal to a given probability threshold t:

$$f(x) = \begin{cases} \arg\max_{C_i} \left(p(C_i|x)\right), & \text{if } \max_{C_i} p(C_i|x) \geq t \\ \text{reject}, & \text{if } p(C_i|x) \leq t \quad \forall i. \end{cases} \tag{7}$$

where $f : \Re^n \rightarrow C$ represents the classification function, which divides the feature space into two regions R_1 and R_2, one for each predicted class, such that $x \in R_i$ means that $f(x) = C_i$.

The classifier rejects an example if the prediction is not sufficiently reliable. The rejection rate (ReR) is the probability that the classifier rejects the example as follows:

$$p(\text{reject}) = \int\limits_{R_{\text{reject}}} p(x) dx = p\left(\max\left(p(C_i|x)\right) \leq t\right) \tag{8}$$

where R_{reject} denotes the rejection region defined in the feature space, and all examples belonging to this region are rejected by the classifier. Accordingly, the acceptance rate is the probability that the classifier accepts an example, i.e.,

$$p(\text{accept}) = 1 - p(\text{reject}). \tag{9}$$

There is a general relation between the ER and ReR: The ER decreases monotonically, whereas the ReR increases [33]. The following basic properties can be easily verified as follows:

$$\begin{aligned} p(\text{accept}) + p(\text{reject}) &= 1 \\ p\left(f(x) = y\right) + p\left(f(x) \neq y\right) + p(\text{reject}) &= 1 \\ p\left(f(x) = y|\text{accept}\right) + p\left(f(x) \neq y|\text{accept}\right) &= 1. \end{aligned} \tag{10}$$

In Chow's theory, an optimal classifier can be found only if the true posterior probabilities are known. This is rarely the case in practice. Fumera *et al.* showed that Chow's rule does not perform well if a significant error in the probability estimation is present [36]. In this case, Fumera *et al.* [36] claimed that defining different thresholds for each class yields better results.

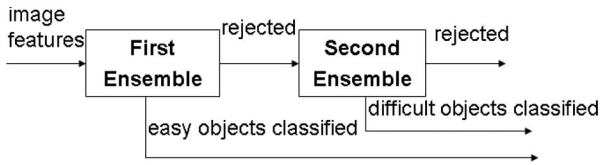


Fig. 5. Cascaded classifier ensembles. Samples rejected at the first stage are passed onto the second stage during classification.

In the following, we adopted the definitions of recognition rate (RR), the ReR, the ER, and reliability proposed and expounded in [39]:

$$\begin{aligned} \text{RR} &= \frac{\#\text{correctly recognized images}}{\#\text{testing images}} \\ \text{ReR} &= \frac{\#\text{rejected images}}{\#\text{testing images}} \\ \text{Reliability} &= \text{RR} + \text{ReR} \\ \text{ER} &= 100\% - \text{reliability}. \end{aligned}$$

According to the definition of reliability, a high reliability can be achieved with appropriate tradeoff between ER and ReR.

B. Cascade Two-Stage Classifier Ensemble Scheme

It has been proven that classification accuracy can be enhanced by using an ensemble of classifiers. In the past several years, a number of successful ensemble methods have been proposed [23], [24], [28], [29]. The most popular methods for creating a classifier ensemble is to build multiple parallel classifiers and then to combine their outputs according to certain decision fusion strategies, such as majority voting. Alternatively, serial architecture can be adopted with different classifiers arranged in cascade, and the output of each classifier is the input to the classifier of the next stage of the cascade. In the following, we will propose a cascade classification scheme following this line, aiming at enhancing the classification reliability by taking advantages of both parallel and serial ensembles.

The cascade scheme consists of two ensembles. The first is a heterogeneous ensemble made up of four multiclass classifiers (kNN, SVM, MLP, and random forest) accepting Gabor features and PHOG features as their input, respectively. The second ensemble, serially concatenated with the first, is implemented by the RF with MLPs as the components, targeting the rejected patterns from the first stage to further insure the confidence. At all stages, a pattern can be either classified or rejected. Rejected patterns are fed into the next stage, which means the two-stage ensembles can be extended to three stages and so on. The overall system is shown in Fig. 5.

The major issue for designing the above cascaded classification system is to decide under what condition a pattern should be classified by the first ensemble and under what condition it should be treated as an exception (rejected) and then classified by the second ensemble [34]. The standard approach [33], [35] is to estimate the class posteriors and to reject the most unreliable objects (the objects that have the lowest class posterior probabilities). It is often one of the weak links in designing

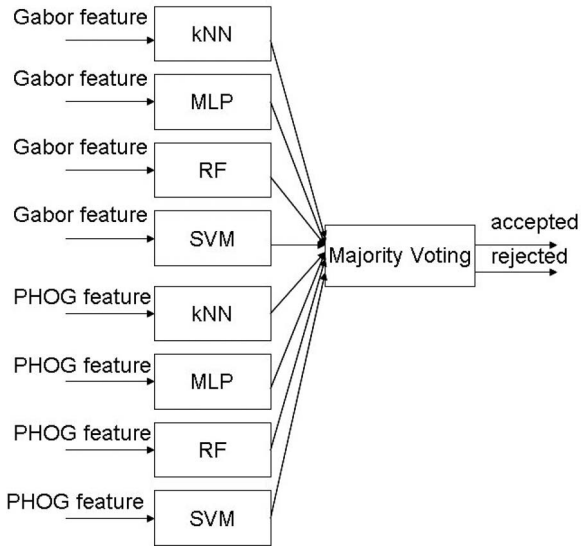


Fig. 6. Classifier ensemble with rejection option in stage 1.

reliable classifiers due to strict demand for a sufficient number of samples for the purpose of probability estimation.

Our innovation to realize classification with a rejection option is based on the simple conception: A rejection is considered neither correct nor wrong; therefore, it is equivalent to a neutral position or an abstention [41]. Toward this, a heterogeneous ensemble of classifiers can be trained. Instead of attempting to heighten the classification accuracy, the objective of the ensemble is to withhold the classification of a pattern if the committee's decision is considered not sufficiently reliable.

Given a test example $x \in \Re^n$, each member of the committee makes a prediction of x's label. With an ensemble of size M_1, labels $l_1, l_2, \ldots, l_{M_1}$ are predicted from the committee. In forming the decision from the M_1 classifiers (committee members), sample x is assigned the class for which there is a predefined consensus degree, or when at least t of the experts are agreed on the label, where

$$t \geq \begin{cases} \frac{M_1}{2} + 1, & \text{if } M_1 \text{ is even} \\ \frac{M_1+1}{2}, & \text{if } M_1 \text{ is odd.} \end{cases} \tag{11}$$

Otherwise, the example is rejected. Since there can be more than two classes, the combined decision is correct when majority of the experts are correct but incorrect when a majority of the decisions are wrong. Obviously, t is a tunable threshold that control the ReR, and we use t to relate the consensus degree from the majority voting to confidence measure and to abstain to classify ambiguous samples in classification. Fig. 6 further explains the principle of the first-stage ensemble with rejection option,

In our design of the first-stage ensemble, diversity of the committee members is the primary concern as it has known to be one of the most important factors in influencing the ensemble's performances. We devised such an ensemble by choosing classifiers differing in feature presentations, architectures, and learning algorithms to achieve their complementary behavior. In the ensemble, the multiclass classification is handled directly by four individual classifiers [19], including kNN, MLP, SVM, and random forest classifiers. These individual classifiers are simultaneously trained with two different features, i.e., PHOG and Gabor.

To make this paper self-contained, a concise description of the four models is summarized in the following.

The kNN is a typical *prototype-based* classifier among the most known classification methods. The idea is to compare pairs of data samples based on a distance function, e.g., the Euclidean distance. It classifies a test sample by first finding the k closest samples in the training set and then by predicting the class by majority voting. Obviously, the kNN classifier needs to access all learning data at the time when a new test case is to be classified.

The MLP is the most popular neural network model, commonly applied in a two-layer feedforward structure with output nodes representing class labels. The activation functions for hidden and output nodes are often logistic sigmoid function and linear function, respectively. An MLP trained with the backpropagation algorithm has been successfully applied to many classification problems.

The SVM is originated from the statistical learning theory [21]. One distinction between the SVM and many other learning systems is that its decision surface is an optimal hyperplane in a higher dimensional feature space. Designing SVM classifiers includes selecting the proper kernel function and the corresponding kernel parameters and choosing proper C value [42]. Recently, an intersection kernel SVMs have been shown to be fast and successful for detection and recognition [43].

Traditional decision tree classifiers are presented in a binary tree structure constructed by repeatedly splitting the data subsets into two descendant subsets. Each terminal subset is assigned with a class label, and the resulting partition of the data set corresponds to the classifier. The random forest classifier [22] consists of many decision trees and outputs the class that is the mode of the classes output by individual trees.

Using different types of classifiers as the constituent members in the "committee of experts" is one of our design strategies in obtaining necessary diversity, thus achieving improved performance in terms of accuracy–rejection tradeoff.

C. RF Ensemble of Neural Network Classifiers

As we aim at constructing a cascade classifier ensembles to produce highly confident classification, the rejected samples from the first ensemble (stage 1) should be further handled to increase the assuredness for their exclusion from classification. From the same initiative for designing first ensemble, the next stage is also implemented by a classifier ensemble. Being different with stage 1, a homogeneous ensemble is selected, which consists of the same classifiers. From many options, the MLP classifier is chosen as the base learners with the following considerations. First of all, it has been proven that a simple neural network, such as MLP, can approximate any continuous function if there is sufficient number of middle-layer units [20]. Second, the generalization performance of MLPs is not very stable in the sense that different settings, such as different network architectures and initial conditions, may all influence the learning outcome. The existence of such differences

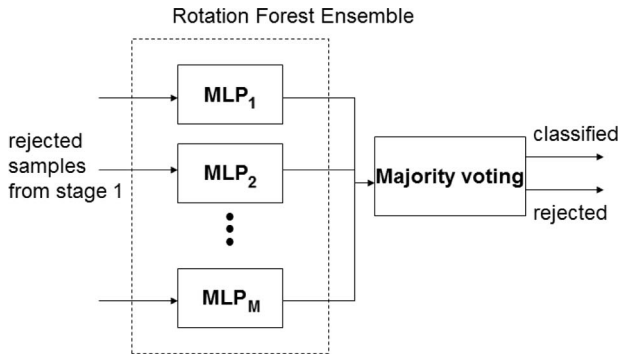


Fig. 7. Stage 2 RF classifier ensemble that consists of a set of MLPs.

between base classifiers is essential for a good classifier ensemble [24], [29].

The MLP ensemble gains benefit from the disagreement among the base members. To further promote their diversity, a newly proposed ensemble metalearning algorithm RF can be applied [30], [31]. Similar to random forest, the RF builds base classifiers independently in a rotated feature space [30]. The feature set F is first divided into K subsets, and then PCA is applied to each subset, respectively. A base MLP is trained with the transformed subset. Thus, K-axis rotations take place to form the new attributes for a base MLP, and all the principal components are retained to preserve the variability information in the data.

The stage 2 ensemble is shown in Fig. 7. The MLP was similarly structured as used in stage 1, with one hidden layer and 21 output nodes, each representing a class label. The final decision from these MLP models can follow the same majority voting procedure. In the combination of decisions from the M_2 base MLPs, a sample (which is the rejected from stage 1) is assigned with the class for which there is a predefined consensus degree or when at least t of the MLPs is agreed on the decision, where $t \geq M_2 + 1/2$ if M_2 is odd, and $t \geq M_2/2 + 1$ if M_2 is even. Otherwise, the sample is rejected. Again, t is the threshold that decide the ReR. The consensus degree from the ensemble acts as confidence measure to switch between acceptance and rejection.

To make this paper more self-sufficient, the details of RF is further described as follows.

Consider training set $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, consisting of N samples. $\mathbf{x}_i$ is a vector of the form $[x_{i,1}, \ldots, x_{i,n}]$, where n is the number of features. y_i takes a value from the set of class labels $\{\omega_1, \ldots, \omega_c\}$. Denote $D_1, \ldots, D_L$ as the classifiers in the ensemble, where L is the size of the ensemble, which is a problem-dependent parameter. Designate $\mathbf{F}$ as the feature set. To construct the training set for classifier D_i, the following steps are executed in turn [30].

1) Set parameter value K, and then split $\mathbf{F}$ randomly into K subsets. The subsets may be disjoint or intersecting. The disjoint subsets are chosen with each feature subset that contains $M = n/K$ features for simplicity.
2) For classifier D_i, the training set is the kth subset of features $\mathbf{F}_{i,k}$. Denote X as the matrix of the samples in the training data set. For every feature subset, bootstrap samples from the X data set in a given percentage of the data count. Then, apply PCA using only the M features in $\mathbf{F}_{i,k}$ and the selected subset of X. Store the coefficients of the principal components, $a_{ik}^{(1)}, \ldots, a_{ik}^{(M_k)}$, each of size $M \times 1$. Note that it is possible that some of the eigenvalues are zero; therefore, we may not have all M vectors; therefore, $M_k \leq M$. To avoid identical coefficients if the same feature subset is chosen for different classifiers, PCA is carried out on a subset of X instead of on the whole set.
3) With PCA performed separately on each K subset of $\mathbf{F}$, a new set of linearly n extracted features is constructed by pooling all principal components. Organize the obtained vectors with coefficients in a sparse "rotation" matrix R_i as follows:

$$R_i = \begin{bmatrix} a_{i1}^{(1)}, \ldots, a_{i1}^{(M_1)} & \cdots & \{0\} \\ \{0\} & \cdots & \{0\} \\ \cdots & \cdots & \cdots \\ \{0\} & \cdots & a_{iK}^{(1)}, \cdots, a_{iK}^{(M_K)} \end{bmatrix}.$$

The columns of R_i should be rearranged according to the original feature sequence, and the rearranged rotation matrix is denoted by R_i^a. Then, the transformed training set for classifier D_i is XR_i^a. This way, all classifiers will be trained in a parallel manner.

As mentioned in [39], the following relations are satisfied in the above two-stage cascade classifier system:

1) overall correctly classified samples = Stage 1 correctly classified samples + Stage 2 correctly classified samples;
2) overall misclassified samples = Stage 1 misclassified samples + Stage 2 misclassified samples;
3) overall rejected samples = Stage 1 rejected samples − Stage 2 misclassified samples − Stage 2 correctly classified samples.

D. Comparison With Previous Studies

The advantage of using serially connected classifiers to implement rejection option has been realized for many years [34], [38], and several serial classifier fusion or cascade classifier ensemble schemes have been proposed in the literature [39], [44], [45]. Our work on reliable classification is based on the same motivation, which can be considered as a further effort along the line. To provide an overall perspective, the similarities and differences between our scheme and previous studies are briefly discussed in the following.

The most obvious common point of using a two-stage system can be found in [34] and [38]. For example, a system proposed in [38] consists of a first (global) classifier with rejection followed by a (local) nearest neighbor classifier. The patterns rejected by the first classifier were classified by the second classifier (the nearest neighbor classifier). This notion has been inherited in this paper. In [44], a three-level classifier with rejection techniques was presented for handwritten digit recognition, with kNN and its variants employed in different levels. In [45], a rejection strategy was suggested for the convolutional neural network models, together with a self-organizing map to alter the links between the neural network layers.

Among the previous studies on reliable classification, [39] is closest to the cascade ensemble model proposed in this paper. In fact, the general notion behind our scheme is same as in [39], i.e., the cascade ensemble classifier system can bring a high recognition performance in terms of accuracy–rejection tradeoff. However, our implementation technique is very different with [39]. In [39], gating networks (GNs) were proposed to assemble the confidence values of three parallel MLPs, and the weights of the GNs are trained by the genetic algorithms. A mixture of experts with GNs generally requires sufficiently many training samples, which often poses a serious problem for real-world applications and image classification in particular. Our proposed cascade classifier ensemble essentially meets this rate-limiting challenge.

However, several interesting findings observed in [39] for the cascade ensemble can be extended to the proposed model in this paper.

- Compared with any single-stage classifier, there will be fewer rejected samples after passing the two-stage ensemble classifiers, which means that the ReR of the two-stage ensemble classifiers is lower than that of any single-stage ensemble classifier.
- The correct RR of the two-stage ensemble classifiers is higher than that of any single-stage ensemble classifier.

V. Experimental Results

As elaborated in Section II, we are interested in features that are generally applicable in vehicle images. The first feature is the normalized bin values from PHOG. In our experiment, three levels of pyramids and eight bins in each level were adopted. In forming the pyramid, the grid at level l has 2^l cells along each dimension. Consequently, level 0 is represented by a K-vector corresponding to the K bins of the histogram and level 1 by a $4K$-vector, etc.

The second feature exploits the Gabor filtering introduced in Section II, with six orientations and four scales. Therefore, with a resized image (256×512), the figuration is applied to 2×4 nonoverlapping image subregions, each with the size of 128×128, yielding overall feature vector with length $2 \times 4 \times 48 = 384$ for each image. To normalize for differences in range, each of the Gabor and PHOG feature components is scaled to have a mean of 0 and a standard deviation of 1 across a data set.

The experiment settings for the different classifiers involved are summarized as follows. For the kNN classifier, we chose $k = 1$ after testing a range of different values of k with the similar results. For the MLP, we experimented with a three-layer network, with one hidden layer of 50 units and a single linear unit representing the class label. The network is trained using the conjugate gradient learning algorithm for 500 epochs. The popular library for SVM, i.e., LIBSVM (www.csie.ntu.edu.tw/cjlin/libsvm), was used in the experiment. We used the radial basis function kernel for the SVM classifier. Parameter γ that defines the spread of the radial function was set to be 5 and parameter C that defines the tradeoff between the classifier accuracy and the margin (the generation) to be 3. In the comparison experiments, the number of trees for random forest classifier was chosen as 300, and the number of variables to be randomly selected from the available set of variables was selected as 20.

TABLE I
Classification Accuracy From kNN, RF, SVM and MLP, Using Gabor and PHOG Features Together With Their Concatenation

	kNN	RF	MLP	SVM
Gabor	78.00%	82.30%	94.36%	89.05%
PHOG	83.44%	86.32%	94.12%	92.44%
Gabor+PHOG	82.03%	86.60%	95.6%	92.0%

For all of the experiments, we randomly split the vehicle image data set into training and testing sets, each time with 20% of each class' images reserved for testing, whereas the rest for training. The performance was from the average of 100 runs, such that each run used a random split of the data to training and testing sets, The first experiment was designed to provide a comparison of the different classifiers adopted in the stage 1 ensemble. With the settings described earlier and the two features together with their simple concatenation, their accuracy values were summarized in Table I.

Several observations can be made in Table I. First, PHOG offers better results than Gabor from kNN, RF, and SVM and similar results from MLP. This means that the PHOG feature can extract richer information from vehicle images and obtain more discriminative power than the Gabor filtering method. Second, being contrary to the general perception that SVM performs better than the MLP as evidenced from many applications, our earlier experiment results showed that MLP is competitive with SVM. Third, the simple concatenation of the two features together may not perform better or even worse than using single features for the classifiers kNN, SVM, and random forest, a phenomenon that has been reported in previous studies [46].

Further experimentation with the proposed cascade ensembles proceeded as follows. In a typical experiment, the first-stage ensemble has eight base classifiers, with kNN, RF, SVM, and MLP each being exploited twice with PHOG and Gabor features, respectively, as elaborated in Section III. During classification, test examples also presented PHOG and Gabor features to the corresponding classifiers. Then, k-out-of-8 majority voting is applied to the output labels from the eight "committee members" to decide a class label if there is a consensus, or reject if otherwise. Here, the "consensus" criterion k acts like a threshold to split the instances into two partitions. To simplify the discussion, we fixed $k = 7$, which yields a relatively high ReR 35.22%. In other words, the first ensemble collectively labels testing examples as belonging to or not to any of the 21 categories while it rejects them from the remaining categories, i.e., no decision is taken about these latter categories. Using a holdout experiment, the first-stage accuracy approximates 99.84% with a high ReR of 35.22%.

To tackle the more difficult samples rejected by the stage 1 ensemble, our design of the second-stage ensemble is partly based on the observation from the results in Table I. Because different features convey information unequally, with imbalanced element values, simple concatenation of two features with equal weight is obviously suboptimal for most of the classification methods. A better solution should preserve the

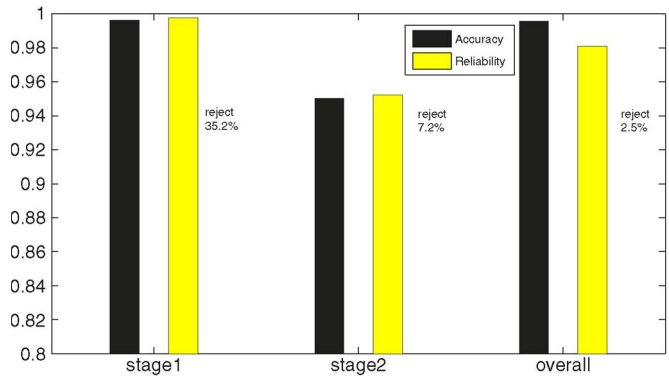


Fig. 8. Bar plots for the comparison of accuracy/reliability pairs from stage 1, stage 2, and the overall results, with corresponding averaged ReRs of 35.2%, 7.2%, and 2.5%. The results were the average of 100 tests from the holdout experiment with 80% of data used for training and the remaining for testing.

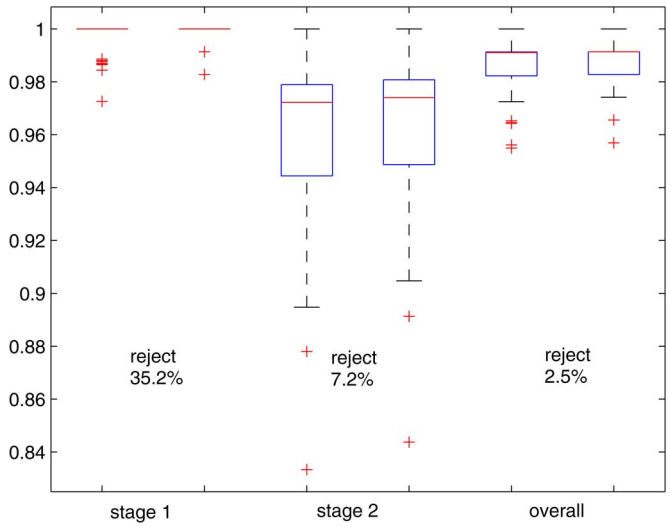


Fig. 9. Box plots for the comparison of accuracy/reliability pairs from stage 1, stage 2, and the overall results, with corresponding averaged ReRs 35.2%, 7.2%,and 2.5%. The results were the average of 100 tests from the holdout experiment with 80% of data used for training and the remaining for testing.

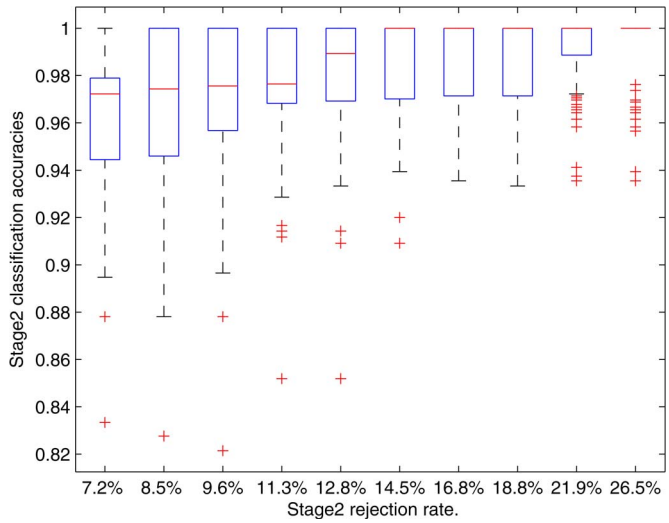


Fig. 10. Averaged stage 2 accuracy with ten varying ReRs.

correlation between different feature pairs. It seems that the MLP has the capability to maximize classification performance by the nonlinear extraction of appropriate key features from the joint PHOG and Gabor attribute values. Compared with the kNN, SVM, and random forest, the MLP has the unique advantage of forming the internal representation from the hidden layer, which may highly relate to the correlation between the PHOG and Gabor feature pairs. It is our hypothesis that such internal feature representation provides the MLP with discriminative features, although a strict proof seems difficult. Therefore, a configuration of second-stage ensemble consists of a number of homogeneous MLPs, trained simultaneously with stage 1 but using a hybrid feature from the concatenation of Gabor and PHOG.

With RF adopted as the metalearning scheme for the stage 2 ensemble, the second stage consists of M MLPs, which are trained simultaneously and independently with stage 1 with the concatenated features of Gabor and PHOG on the same training data set. During classification, the rejected instances from the stage 1 ensemble is passed to stage 2. Similar to stage 1, k-out-of-M majority voting is applied to the output labels from the M MLPs to decide a class label if there is a consensus or reject if otherwise while k is again controlled to yield varying ReR. The overall classification accuracy is defined as the number of correctly classified samples from both stages 1 and 2 over the total number of samples tested. With ensemble size $M = 25$, from the same holdout experiment, the second-stage accuracy is about 96.23% with ReR 7.2% from stage 2, and the overall accuracy is approximately 98.65% with overall ReR 2.5%. Figs. 8 and 9 demonstrate the comparisons of accuracy/reliabilities from stages 1 and 2, together with the overall performance.

To have a closer examination at how the ReR influence the classification accuracy, we fixed the ensemble size as 25 and adjusted the threshold (k-out-of-25) for majority voting from the stage 2 ensemble, whereas the threshold in stage 1 is unchanged, resulting to average ReRs between 7.22% and 26.5% from $k = 13, \ldots, 23$. The corresponding overall ReRs were then in the range of $2.5\%, \ldots, 9.3\%$. The box plots of stage 2 accuracy and corresponding overall accuracy from the varying ReRs were displayed in Figs. 10 and 11, respectively. It is easy to find that higher accuracy can be expected from higher ReR. The tradeoff between accuracy and ReR could be empirically decided in practice.

The confusion matrix is often used in the study of multiclass classification problems and to measure the similarity between classes. The diagonal elements of the confusion matrix that summarizes the detailed situation with ReR 2.5% were displayed in Table II. In the confusion matrix representation, the rows and columns indicate true and predicted classes, respectively. The diagonal entries represent correct classification, whereas the off-diagonal entries represent incorrect classification. It is obvious that, with mere 2.5% ReR, 13 types have been classified, with 100% correct rates among the 21 classes.

With the RF-ensemble metalearning method, an important parameter that need to be considered is the ensemble size L, which is the number of base classifiers in the ensemble. Our next experiment assessed the effect of ensemble size on the

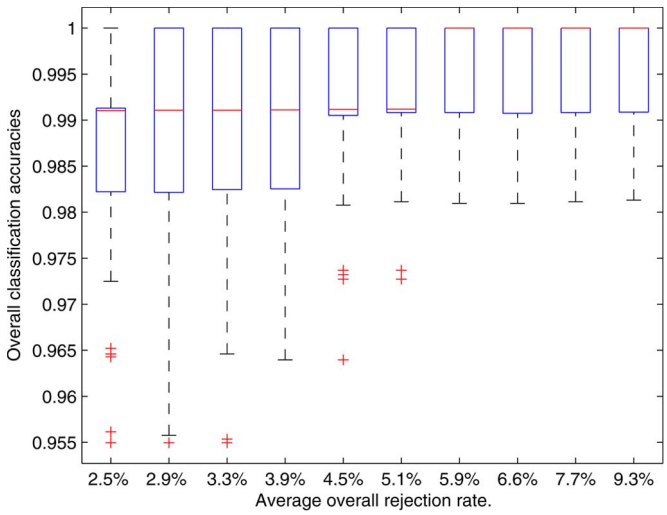


Fig. 11. Box plots of averaged overall classification performances from ten varying overall ReRs.

TABLE II
CLASSIFICATION RATES FOR THE INDIVIDUAL CLASSES

Audi	Buick1	Buick2	Changan	Chery1
100%	90%	100%	98.5%	100%
Chery2	Chery3	Chevrolet1	Chevrolet2	Citroen
97.1%	96.6%	98%	100%	100%
Ford	Honda	Hyundai1	Hyundai2	Mazda
97.5%	100%	100%	100%	98%
Nissan	Peugeot	Toyota	Volkswagen1	Volkswagen2
100%	100%	100%	100%	97%
Wulin				
100%				

TABLE III
CLASSIFICATION ACCURACY (ac) AND RER (re) FROM APPLYING DIFFERENT ENSEMBLE SIZES IN THE PROPOSED CLASSIFICATION SCHEME. THE NUMBERS ARE PERCENTAGE RATES (%)

Size	10		15		20		25	
	ac	re	ac	re	ac	re	ac	re
Stage1	99.7	35	99.8	35	99.6	34.9	99.8	35.2
Stage2	95	3.4	93	5.3	95	3.5	96.2	7.2
Overall	97	1.7	97.5	1.8	98	1.3	98.7	2.5

vehicle-type classifications. We varied the size of the ensembles from ten components to 25. The same experiment procedure described earlier was repeated. The results of the averaged classification accuracy versus the ReRs were summarized in Table III, which shows that there are no real benefits of forming very large ensembles.

Although there has been a number of published works on vehicle-type classification, it is almost impossible to make a fair comparison. First of all, unlike some other image recognition problems such as face recognition, there does not exist any benchmark image sets for vehicle-type classification. However, the results can be roughly contrasted. For example, Petrovic and Cootes reported 87.3% accuracy on an access control car database by exploiting simple gradient features from images [9]. Negri *et al.* in [10] discussed an oriented-contour-point-based voting algorithm and reported a classification performance of 93.1% by using fusion of different classifiers. Simple gradient-based features, however, are generally not robust to traffic field image conditions, such as illumination, scale, in-plane rotation, etc. Accuracy of 89% was presented in [47] by applying the scale-invariant transform features. Several global transforms, including fast Fourier transforms, discrete wavelet transforms, and discrete curvelet transforms, were applied to describe image features in [11], yielding a 92% classification rate for a very small car database, which has only five different vehicle types. In [48], the random-subspace ensemble method was applied for vehicle-type classification based on the combined features of the curvelet transform and PHOG, with accuracy of 96%. None of the published studies considered the classification reliability issue, which is often a serious question in real applications.

VI. CONCLUSION

Accurate and robust classification of vehicle images is still a challenging task in intelligent transportation system and surveillance. To find the best description for vehicle images, many feature extraction methods have been attempted in previous studies, but all encountered with certain aspects of difficulties in dealing with various irregularities in the vehicle images. First, we have strived to demonstrate that the classification can be improved by using highly discriminative image features. This can be achieved by the utilization of two well-known feature description methods, i.e., the PHOG and the Gabor transform. Second, the theme of classification reliability was emphasized, which is distinguishable from the previous research on vehicle-type classification with accuracy as the only subject. In this paper, a new scheme of cascade classifier ensembles has been proposed with rejection strategies. Rather than simply pursuing accuracy, the importance of reject option was stressed to minimize the cost of misclassifications. The cascade ensembles employ a serial approach where the second ensemble is responsible for the patterns rejected by the first, thus labeling some ambiguous samples as undecidable with high confidence. The first-stage ensemble consists of eight classifiers trained in parallel, whereas the second ensemble is composed of an MLP orchestrated by a metalearning scheme RF. For both of the ensembles, rejection option is implemented by relating the consensus degree from majority voting to confidence measure and abstaining to classify doubtful samples if the consensus degree is lower than a predefined threshold. The cascade ensembles prove to secure high classification reliability by balancing a tradeoff between high accuracy and low ReR. As a typical demonstrating result, an accuracy of 98.65% was obtained with ReR 2.5%.

REFERENCES

[1] D. Munroe and M. Madden, "Multi-class and single-class classification approaches to vehicle model recognition from images," in *Proc. Irish Conf. AICS*, Portstewart, U.K., 2005.
[2] G. Pearce and N. Pears, "Automatic make and model recognition from frontal images of cars," in *Proc. 8th IEEE Int. Conf. AVSS*, Klagenfurt, Austria, 2011, pp. 373–378.
[3] S. Gupte, O. Masoud, R. Martin, and N. Papanikolopoulos, "Detection and classification of vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 1, pp. 37–47, Mar. 2002.

[4] T. Kato, Y. Ninomiya, and I. Masaki, "Preceding vehicle recognition based on learning from sample images," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 4, pp. 252–260, Dec. 2002.
[5] S. Sivaraman and M. Trivedi, "General active-learning framework for on-road vehicle recognition and tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 267–276, Jun. 2010.
[6] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using evolutionary Gabor filter optimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 125–137, Jun. 2005.
[7] W. Wei, Q. Zhang, and M. Wang, "A method of vehicle classification using models and neural networks," in *Proc. IEEE Veh. Technol. Conf.*, Rhodes, Greece, 2001, pp. 3022–3026.
[8] T. Yoshida, S. Mohottala, M. Kagesawa, and K. Ikeuchi, "Vehicle classification systems with local-feature based algorithm using CG model images," *IEICE Trans. Inf. Syst.*, vol. E85-D, no. 11, pp. 1745–1752, Nov. 2002.
[9] V. Petrovic and T. Cootes, "Analysis of features for rigid structure vehicle type recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2004, pp. 587–596.
[10] P. Negri, X. Clady, M. Milgram, and R. Poulenard, "An oriented-contour point based voting algorithm for vehicle type classification," in *Proc. 18th ICPR*, 2006, vol. 1, pp. 574–577.
[11] F. M. Kazemi, S. Samadi, H. R. Poorreza, and M.-R. Akbarzadeh-T, "Vehicle recognition based on Fourier, wavelet and curvelet transforms—A comparative study," in *Proc. 4th Int. Conf. ITNG*, Las Vegas, NV, 2007, pp. 939–940.
[12] S. Wang, L. Cui, D. Liu, R. Huck, P. Verma, J. J. Sluss, and S. Cheng, "Vehicle identification via sparse representation," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 955–962, Jun. 2012.
[13] Z. Chen, T. Ellis, and S. A. Velastin, "Vehicle type categorization: A comparison of classification schemes," in *Proc. 14th Int. IEEE ITSC*, Washington, DC, 2011, pp. 74–79.
[14] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proc. 6th ACM Int. Conf. Image Video Retrieval*, Amsterdam, The Netherlands, 2007, pp. 401–408.
[15] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, New York, 2006, pp. 2169–2178.
[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. CVPR*, San Diego, CA, 2005, pp. 886–893.
[17] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profile," *Vis. Res.*, vol. 20, no. 10, pp. 847–856, 1980.
[18] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of large image data," *IEEE Trans. Pattern Anal. Mach. Intell. (Spec. Issue Digital Libraries)*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
[19] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification,* 2nd ed. New York: Wiley, 2001.
[20] S. Haykin, *Neural Networks: A Comprehensive Foundation,* 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[21] C. Cortes and V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
[22] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[23] L. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990.
[24] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ: Wiley, 2004.
[25] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
[26] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.
[27] H. Schwenk and Y. Bengio, "Boosting neural networks," *Neural Comput.*, vol. 12, no. 8, pp. 1869–1887, Aug. 2000.
[28] T. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.
[29] L. Kuncheva, J. Rodriguez, C. Plumpton, D. Linden, and S. Johnston, "Random subspace ensembles for fMRI classification," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 531–542, Feb. 2010.
[30] J. Rodríguez, L. Kuncheva, and C. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.
[31] L. Kuncheva and J. Rodríguez, "An experimental study on rotation forest ensembles," in *Proc. Multiple Classifier Syst., Lec. Notes Comput. Sci.*, 2007, vol. 4472, pp. 459–468.
[32] K. Liu and D. Huang, "Cancer classification using rotation forest," *Comput. Biol. Med.*, vol. 38, no. 5, pp. 601–610, May 2008.
[33] C. K. Chow, "On optimum recognition error and reject tradeoff," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 1, pp. 41–46, Jan. 1970.
[34] P. Pudil, J. Novovicova, S. Blaha, and J. Kittler, "Multistage pattern recognition with reject option," in *Proc. 11th IAPR Conf.*, 1992, vol. 2, pp. 92–95.
[35] D. M. J. Tax and R. P. W. Duin, "Growing a multi-class classifier with a reject option," *Pattern Recognit. Lett.*, vol. 29, no. 10, pp. 1565–1570, Jul. 2008.
[36] G. Fumera, F. Roli, and G. Giacinto, "Reject option with multiple thresholds," *Pattern Recognit.*, vol. 33, no. 12, pp. 2099–2101, Dec. 2000.
[37] G. Fumera and F. Roli, "Analysis of error-reject trade-off in linearly combined multiple classifiers," *Pattern Recognit.*, vol. 37, no. 6, pp. 1245–1265, Jun. 2004.
[38] N. Giusti, F. Masulli, and A. Sperduti, "A theoretical and experimental analysis of a two-stage system for classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 893–904, Jul. 2002.
[39] P. Zhang, T. D. Bui, and C. Y. Suen, "A novel cascade ensemble classifier system with a high recognition performance on handwritten digits," *Pattern Recognit.*, vol. 40, no. 12, pp. 3415–3429, Dec. 2007.
[40] T. Sikora, "The MPEG-7 visual standard for content description—An overview," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 696–702, Jun. 2001.
[41] L. Lam and C. Y. Suen, "Application of majority voting to pattern recognition: An analysis of its behavior and performance," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 27, no. 5, pp. 553–568, Sep. 1997.
[42] C. W. Hsu and C. J. Lin, "A comparison on methods for multi-class support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.
[43] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *Proc. IEEE Conf. CVPR*, Anchorage, AK, 2008, pp. 1–8.
[44] C. Rodriguez, I. Soraluze, J. Muguerza, J. I. Martin, and G. Alvarez, "Hierarchical classifiers based on neighborhood criteria with adaptive computational cost," *Pattern Recognit.*, vol. 35, no. 12, pp. 2761–2769, Dec. 2002.
[45] H. Cecotti and A. Belaid, "Rejection strategy for convolutional neural network by adaptive topology applied to handwritten digits recognition," in *Proc. 8th ICDAR*, Seoul, Korea, Aug. 2005, pp. 765–769.
[46] Y. Fu, L. Cao, G. Guo, and T. S. Huang, "Multiple feature fusion by subspace learning," in *Proc. ACM Int. CIVR*, Niagara Falls, Canada, 2008, pp. 127–134.
[47] L. Dlagnekov, "Video-based car surveillance: License plate make and model recognition," M.S. thesis, Univ. California, San Diego, CA, 2005.
[48] B. Zhang and C. Zhao, "Classification of vehicle make by combined features and random subspace ensemble," in *Proc. 6th ICIG*, Hefei, China, 2011, pp. 920–925.

Bailing Zhang received the Master's degree in communication and electronic system from the South China University of Technology, Guangzhou, China, in 1987 and the Ph.D. degree in electrical and computer engineering from the University of Newcastle, Callaghan, Australia, in 1999.

He is currently an Accociate Professor with the Department of Computer Science and Software Engineering, Xi'an Jiaotong–Liverpool University, Suzhou, China. His research interests include image/video processing and pattern recognition, with applications in intelligent transportation systems, surveillance, and bioinformatics.