Vehicle Type Classification using Hierarchical Classifiers

Jie Lian\textsuperscript{1,2}, Jun Zhang\textsuperscript{1,2}, Tong Gan\textsuperscript{1,2} and Shaobo Jiang\textsuperscript{1,2}
\textsuperscript{1}No.38 Research Institute, China Electronics Technology Group Corporation, Hefei, 230088, China
\textsuperscript{2}Anhui Key Laboratory of Emergency Information Technology of Public Safety, Hefei, 230088, China
Email: 166lianjie@163.com

Abstract. Vehicle type classification based on images is widely applied in traffic surveillance and monitoring system. It has been used in billing at toll collection stations and preventing heavy trucks from entering city viaducts. A hierarchical vehicle detection and classification system is proposed in this paper. A cascade ensemble classifier, accepting Multiple Layer Perceptron (MLP) and K-Nearest Neighbor (K-NN) as base classifiers, is proposed for the vehicle type classification. The experiments are conducted, and the hierarchical classifier with two layers offers a reliability of 97.8\% with a rejection rate of 3.0\%, which show the effectiveness of our proposed hierarchical vehicle detection and classification system.

1. Introduction
As the number of vehicle soaring on the road, Intelligent Transportation Systems (ITS) becomes a significant issue in current’s world. Road traffic monitoring aims at the acquisition and analysis of traffic conditions, such as the presence and numbers of vehicles [1]. More specifically, traffic statistics data from vehicle classification can provide important information for conducting safety and operational evaluation of highway facilities [2]. A vehicle type recognition system is very useful in obtaining safer traffic conditions, such as retrieving blacklisted vehicles from the traffic surveillance image database. In previous works, a semi-supervised convolutional neural network been proposed in [3] to classify vehicles in frontal-view images. Yang et al. in [4] investigated car model classification with Convolutional Neural Networks on a large-scale dataset “Compcars”. Zhang et al.[5] designed a multi-task learning framework to learn fine-grained feature representations by optimizing classification and similarity constraint.

Some researchers have introduced a multi-classifier rejection policy. Zhang et al. [6] proposed a cascade ensemble classifier system for handwritten digits recognition. The ensemble classifier consists of three parallel artificial neural networks (ANNs) and three gating networks (GNs). Similarly, Zhang [7] proposed a cascade ensemble classifier for vehicle make and model classification. Rodriguze et al. [8] dealt with handwritten digits recognition using a three stage classifier with rejection techniques. K-NN and K-nearest centroid neighborhood (k-NCN) were used as base classifiers.

Combing classifiers using weighting or meta-learning [9], we can obtain a reliable classification scheme by exerting control over the accuracy of the classifier to make determinations. Motivated by the above ideas, we design a two stage cascade ensemble classifier for vehicle type classification. Vehicles to be recognized will be recognized by the first stage classifier, and the input of the second stage classifier is the rejected vehicles in the first stage classifier. The classification generalization error can be reduced by enhancing the diversity of base classifiers in an ensemble classifier. We find that k-NN and MLP are both sensitive to different features [7], [10], and different classification outputs can be obtained by modifying the parameter and structure of these two classifiers.
The rest of this paper is organized as follows. Section 2 introduces the vehicle image data acquisition and the vehicle area detection method. Section 3 introduces the ensemble classifier rules and the structure of the cascade ensemble classifier with rejection option. The experimental results are illustrated in Section 4, and the Section 5 is the conclusion of this paper.

2. Images Acquisition and Vehicle Detection
The transportation department of High-tech Zone in Hefei city has installed many surveillance cameras covering the road. The images we use in our experiments were captured by cameras installed at five different testing sites, with a wide range of weather and illumination conditions. More than 1800 images of different vehicles were selected for five vehicle types, including cars, vans, buses, medium-trucks and heavy trucks. All of these images contained front views of one or more vehicles. The original size of every image is $1920 \times 1152$. Fig. 1 shows a part of these images. The method of our earlier work [11] is used to detect vehicle in images.

![Figure 1](image)

**Figure 1.** A part of selected images

3. Classification with Rejection Option and Multi-Stage Classification
In order to improve reliability and reduce error rate at the same time, a cascade classifier is introduced for vehicle classification. Fig. 2 shows the overall classification scheme, where an input testing pattern will firstly be recognized by the first stage (MLP based) ensemble classifier, then the rejected patterns of first stage will be further recognized by the second stage (k-NN based) ensemble classifier.

![Diagram](diagram)

**Figure 2.** The cascade ensemble classification scheme.

The two stage cascade classifier satisfies the following relations:
- Overall Recognized Patterns = Recognized Patterns 1 + Recognized Patterns 2
- Overall Misrecognized Patterns = Misrecognized Patterns 1 + Misrecognized Patterns 2
- Overall Rejected Patterns = Rejected Pattern 1 - Recognized Patterns 2 - Misrecognized Patterns 2
More specifically, we have to consider the recognition rule that which pattern should be recognized or rejected in our classification system. In [9], Rokach summarized several recognition rules for ensemble classifiers, including sum rule, product rule and majority voting, etc. For the sake of simplicity, we use majority voting as the decision rule here. Given a testing sample \(x(\in \mathbb{R}^n)\), the recognition result of every classifier is \(h_1(x), h_2(x), \ldots, h_M(x)\), where \(M\) is the number of classifiers in ensemble classifier. The majority voting can be described by:

\[
\text{label}(x) = \begin{cases} 
  \text{argmax}_{c \in H} (\sum_m \delta(h_m(x), c)), & \text{if } \max(\sum_m \delta(h_m(x), c)) \geq \tau \\
  \text{reject}, & \text{otherwise} 
\end{cases}
\]

Where \(h_m(x)(1 \leq m \leq M)\) is the classification result of the \(m\)-th base classifier, \(H\) is the domain of \(h_m(x)\), \(c_i\) is the label for the \(i\)-th class, and \(\delta(h_m(x), c_i)\) is the function that:

\[
\delta(h_m(x), c_i) = \begin{cases} 
  1 & \text{if } h_m(x) = c_i \\
  0 & \text{if } h_m(x) \neq c_i 
\end{cases}
\]

The ensemble classifier accepts the recognition result when there are at least \(\tau\) base classifiers having the same result. Otherwise, the testing sample is rejected. \(\tau\) is the “decision threshold” here, and it is defined as

\[
\tau = \left\lfloor \frac{M}{2} + 1 \right\rfloor, \text{if } M \text{ is even} \\
\left\lfloor \frac{M+1}{2} \right\rfloor, \text{if } M \text{ is odd}
\]

After the introduction of rejection option, the recognition rate (RR), the rejection rate (ReR), the error rate (ER) and the Reliability of classification system can be respectively redefined as:

- RR = # correct recognized samples/ # testing samples
- ReR = # rejected samples/ # testing samples
- Reliability = RR + ReR
- ER = 100% - Reliability

MLP is chosen as the base classifier for the following reasons. First, MLP has the ability to approximate any continuous function if there are enough hidden nodes. Second, the generalization performance of MLP is unstable to initial conditions and activation functions. The existence of such difference makes MLP a good base classifier for ensemble classifiers [14]. In [15], Riedmiller et al. proposed an alternative RPROP algorithm that performs a direct adaptation of the weight step based on local gradient information. Such difference generates different outputs during recognition. The feature description of images that learned by classifiers should give intrinsic and structural information for classification. Here, we introduce an edge feature (HOG) [12] and a texture feature (LBP) [13] for vehicle classification. The first stage ensemble classifier is made up of MLP-BP and MLP-RPROP accepting HOG, LBP and their combined feature HOG-LBP for vehicle classification as their inputs, respectively.

K-NN is arranged at the second stage for its more discriminative capability, which will be testified in the experiment section. Domeniconi[10] pointed out that k-NN techniques are sensitive to different input features and distance functions, as well as the number of K. According to the characteristics of K-NN, Euclidean and Manhattan distance can be used as distance functions to generate an effective ensemble classifier. Two different nearest neighbor number \(k_1\) and \(k_2\) combined with K-NN Euclidean distance (K-NN-E) and K-NN Manhattan distance (K-NN-M) are used as four different classifiers. For the k-NN Manhattan distance, given a data set \(D\) with \(m\) samples, the similarity between an input feature vector \(X = \{x_1, x_2, \ldots, x_n\}\) and the \(i\)th sample in \(D\) is calculated by:

\[
g_i = |X - D_i| = \sum_{j=1}^{n} |(x_i - d_{ij})|, 1 \leq i \leq m
\]

The nearest neighbor \(k_1\) and \(k_2\) number are chosen by grid search to maximize the diversity and accuracy of the ensemble classifier. The similarity of two base classifiers is defined by the number of samples recognized or miss-recognized by \(h_i(x), h_j(x)(1 \leq i, j \leq M)\), which is given in table 1.
Table 1. The classification result of two base classifiers.

| Recognized by $h_j(x)$ | Miss-Recognized by $h_i(x)$ |
|------------------------|-----------------------------|
| Recognized by $h_i(x)$ | $N_{11}$                    | $N_{10}$                    |
| Miss-Recognized by     | $N_{01}$                    | $N_{00}$                    |

The con-variance $P_{ij}$ is used to measure the diversity of two base classifiers [16], and it is calculated by:

$$P_{ij} = \frac{N_{11}N_{00} - N_{01}N_{10}}{\sqrt{(N_{11}+N_{10})(N_{01}+N_{00})(N_{11}+N_{01})(N_{10}+N_{00})}}$$ (5)

The overall con-variance $P_{ov}$ of the ensemble classifier is averaged by two-two combination of all base classifiers. A larger con-variance value means higher similarity value, as well as a lower diversity value among base classifiers.

$$P_{ov} = \frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} P_{ij}$$ (6)

4. Experimental Results

The dataset used in vehicle classification is obtained by the detection method introduced in section 2. The numbers of vehicles per vehicle type are 387, 308, 508, 308 and 359 respectively. For each vehicle type, we randomly select 150 samples for training and 100 samples for testing for every classifier, the reported results of the dataset are the averages of 10-holdout experiment results. Fig. 3 shows a part of vehicle samples for vehicle type recognition.

![Figure 3. A part of vehicle samples.](image)

In this work, all input images are resized to the size of $64 \times 64$. For the HOG and LBP feature, a feature vector with a dimension of 1764 and 944 is generated, respectively. For the MLP classifier, we test with a three layer network, which has an input layer of $n_1$ nodes equals to the dimension of feature vectors. The output layer has $n_2$ nodes that are equal to the number of vehicle types. For the k-NN classifier, $k_1$ and $k_2$ are chosen by a grid search to maximize the accuracy and the diversity of the ensemble classifier. We test the performance of k-NN-E and k-NN-M on a range of k (odd from 1
to 19), which accept HOG, LBP and HOG-LBP as inputs, respectively. The experimental result is given in figure 4, which shows a significantly decreasing of recognition rate from k=9. The abbreviations of ‘E’ in hogE corresponding to k-NN Euclidean distance, and ‘M’ in hogM corresponding to K-NN Manhattan distance. As a result, we further examine the diversity of k-NN based classifier on k among 1~9 by formula (5). The overall con-variance $P_{op}$ of every different combination is shown in table 2. Obviously, $k_1 = 1$ and $k_2 = 9$ should be chosen for its lowest con-variance value and a relatively larger averaged recognition rate.

![Figure 4](image)

**Figure 4.** The accuracy rate of k-NN on different number of nearest neighbors.

| k_1 | 1 | 1 | 1 | 1 | 3 | 3 | 3 | 5 | 5 | 7 |
|-----|---|---|---|---|---|---|---|---|---|---|
| k_2 | 3 | 5 | 7 | 9 | 5 | 7 | 9 | 7 | 9 | 9 |
| $P_{op}$ | 0.49346 | 0.48127 | 0.47776 | 0.47463 | 0.51531 | 0.52202 | 0.51955 | 0.53264 | 0.52879 | 0.55976 |

The second experiment was elaborated to compare classification results of every base classifier. From Table 3, we observe that first, for the MLP classifier, the recognition result of HOG is better than LBP, whereas a controversy result is obtained by the k-NN classifier. Second, compared with using single features, we obtain better results by a simple concatenation of HOG and LBP. It demonstrates that MLP and k-NN seem to be able to extract key features from these two features and combine them using a nonlinear combination. Meanwhile, MLP-RPROP gives a better results than MLP-BP, while $k_1$-NN- M gives the best performance among k-NN based classifiers.

|                | MLP-BP     | MLP-RPROP  | k_1-NN-E   | k_2-NN-E   | k_1-NN-M   | k_2-NN-M   |
|----------------|------------|------------|------------|------------|------------|------------|
| HOG            | 87.80%     | 87.60%     | 86.80%     | 84.00%     | 89.20%     | 85.80%     |
| LBP            | 84.60%     | 83.60%     | 87.40%     | 86.00%     | 89.20%     | 86.80%     |
| HOG-LBP        | 88.80%     | 89.80%     | 87.60%     | 85.60%     | 90.00%     | 87.20%     |

A further experiment is elaborated with the proposed cascade ensemble classifier. The ensemble classifier in the first stage has six classification schemes. During classification, k/6 majority voting is applied to decide if an input pattern is received or rejected. k is a threshold value for rejection here, and we chose k = 5, which generate a reliability about 98.60% with a ReR of 19.6%. Rejected patterns of the first stage are further processed by the second stage. Again, k/12 majority voting is applied to make a decision. With k = 8, the reliability of the second classifier is about 96.94% with a ReR
15.31%. The classification rate of the overall system reaches 94.8% with the overall ReR 3.0%. Table 4 shows the classification performance of every stage and the overall cascade ensemble classifier system.

Table 4. The comparison of RR, ReR and Reliability from stage 1, stage 2 and the overall results.

|                | 1st stage(MLP) | 2nd stage(k-NN) | overall  |
|----------------|----------------|-----------------|----------|
| RR             | 79.00%         | 81.63%          | 94.80%   |
| ReR            | 19.60%         | 15.31%          | 3.00%    |
| Reliability    | 98.60%         | 96.94%          | 97.80%   |

The cascade classifier solves the confusion problems a lot, which rejects hard examples and leaves more discriminative classifier to recognize them. When we use the rejection strategy in this study, the error rate is 2.2% with a rejection rate of 3.0%. When we use single classifiers, the highest recognition accuracy is 91.0%, which is from \( k_p \)-NN-M classifier and yielding an error rate of 10.0%. Obviously, the classifier ensemble technique resolves the classification bias of single classifiers.

5. Conclusion
A cascade ensemble classifier was proposed to promote the reliability of the vehicle type classification system. The multi-classifier is implemented by MLP and k-NN accepts HOG, LBP and the combined feature HOG-LBP as inputs, and a diversity measurement method is discussed to maximize the accuracy and diversity of ensemble classifiers. This classification system solves the confusion issue between car and van, medium truck and heavy truck significantly. It reaches an accuracy rate of 94.8% with a rejection rate of 3.0% after balancing the tradeoff between rejection rate and error rate.

6. Acknowledgments
This work was partially supported by National Key Research and Development Project of China through grant 2017YFC0804900.

7. References
[1] Buch N, Velastin S, Orwell J, et al. A Review of Computer Vision Techniques for the Analysis of Urban Traffic. IEEE Transaction on Intelligent Transportation Systems, 2011, 12 (3), pp. 920-939.
[2] Hasan MM, Saha G, Hoque G, et al. Smart Traffic Control System with Application of Image Processing Techniques. International Conference on Informatics, Electronics &Vision, Dhaka, 2014, pp.1-4.
[3] Dong Z, et al. Vehicle Type Classification Using a Semi supervised Convolutional Neural Network[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16 (4): 2247-2256.
[4] Yang L, Luo P, Loy C, et al. A Large-Scale Car Dataset for Fine-Grained Categorization and Verification. In Computer Vision and Pattern Recognition, 2015.
[5] Zhang X, Zhou F, Lin Y, et al. Embedding Label Structures for Fine-Grained Feature Representation [J]. European Conference on Computer Vision, 2015:1114-1123.
[6] Zhang P, Bui T D, Suen C Y. A novel cascade ensemble classifier system with a high recognition performance on handwritten digits. Pattern Recognition, 2007, 40(12), pp. 3415–3429.
[7] Zhang B. Reliable classification of vehicle types based on cascade classifier ensembles. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(1), pp. 322 - 332.
[8] Rodriguez C, Soraluze I, Muguerza J, et al. Hierarchical classifiers based on neighborhood criteria with adaptive computational cost, The Journal of the Pattern Recognition Society, 2002, pp.2671-2679.
[9] Rokach L. Ensemble-based classifiers. Artificial Intelligence Review, 2010, 33, pp. 1-19.
[10] Domeniconi C, Yan B. Nearest neighbor ensemble. Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, August 23-26, 2004, pp. 228 - 231.
[11] Lian J, Zhao C, Zhang B, et al. Vehicle detection based on information fusion of vehicle symmetrical contour and license plate Position. Journal of Southeast University, 2012, 28 (2), pp. 240 - 244.
[12] Pan H, Zhang B. An integrative approach to accurate vehicle logo detection. Journal of Electrical and Computer Engineering, vol.2013, Article ID 391652, 12 pages, doi:10.1155/2013/391652.
[13] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. On PAMI, 2002, 24 (7), pp. 971-987.
[14] Kuncheva L, Rodriguez J, Plumpton C, et al. Random subspace ensembles for fMRI classification. IEEE Transactions on Medical imaging, 29 (2), 2010, pp. 531 - 542.
[15] Riedmiller B, Braun H. A direct adaptive method for fast back-propagation learning: The RPROP Algorithm. Proc. ICNN, San Francisco, 1993, 1, pp. 586-591.
[16] Everitt B. The Analysis of Contingency Tables, London: Chapman and Hall, 1977.