Research on Vehicle Type Classification Based on Spatial Pyramid Representation and BP Neural Network

Shaoyue Song(✉) and Zhenjiang Miao

Institute of Information Science, Beijing Jiaotong University, Beijing, China
{14112060,zjmiao}@bjtu.edu.cn

Abstract. This paper presents a method of the vehicle type classification based on spatial pyramid representation and BP neural network. We extract feature vectors of each vehicle image by using the spatial pyramid representation method. By this way, we can use different size of pictures instead of changing the picture into a fixed size avoiding the deformation of the target images when cropping or warping and so on. We choose BP neural network to train our classifier and have a good performance on car, bus and truck classification.

Keywords: Vehicle type classification · Spatial pyramid representation · BP neural network

1 Introduction

With the growth in the volume of our country's car ownership, it's important for us to establish and improve an intelligent transportation system. The vehicle type classification is an important part of intelligent transport system and has a wide application prospect in the future. There are many methods on classifying vehicle type. Usually we can divide them into several categories: by contour scanning, by changing magnetic field and by the image-based method. [1] Compared with the other two methods, the method which based on images is simple, fast and effective.

As is known to us, feature extraction is the key task of target classification. Some researchers [2–5] get length, roof length and height of the vehicle through pictures and use them to classify the vehicle. This kind of methods are geometric-based methods that some geometric measurements are needed. Because of the diversity of the vehicle and the change of vehicle's attitude in the image, it's difficult to meet the requirement of accuracy and fast classification.

Another common approach is appearance-based method. In this kind of methods, vehicle images are usually represented as vectors in some high-dimensional space. [6] Weixin Kang et al. [7] combine Harris corner and SIFT feature to classify the vehicle. In the research of Fuqing Zhu et al. [8], the image features they used are based on the

Supported by "The Fundamental Research Funds for the Central Universities".

Y.-J. Zhang (Ed.): ICIG 2015, Part III, LNCS 9219, pp. 188–196, 2015.
DOI: 10.1007/978-3-319-21969-1_17

histogram of the image local feature sparse coding. They use a SVM method to train the classifier and achieve accuracy of more than 90 % over six categories.

The algorithm artificial neural network is one of the most popular research directions in object recognition and classification based on images. [9, 10] BP neural network is the most intensively studied and the most widely applied model in artificial neural network. [11] In some studies of the BP network, the images need to be changed into a fixed size before classification. But after the size normalization, there usually will be some problems like image deletion or geometric distortion. This will reduce the performance of the classification of the vehicle. In [9], the authors present a spatial pyramid representation method which can remove the fixed-size constrain of the method by using the SPP (Spatial pyramid pooling) model. The CNNs networks method in [9] usually needs a period of time to train. We want to find out an easier way to achieve our classification task, so a simple BP neural network is used in our experiment.

In this paper, we extract the feature vector via SPR model presented in [12] and train our classifier using a three-layer BP neural network. Details on the algorithm are discussed in later sections. The images we used in the experiment are captured from traffic video which is collected in the real life scenarios. We train a three-type classifier of vehicle and the classification accuracy rate can reach 91.0 %.

2 Method

We combine spatial pyramid representation and BP neural network to design the vehicle type classifier. Any size of pictures can be directly inputted into our classifier instead of being transformed into the same size.

2.1 BP Neural Network

BP (back-propagation) neural network is a multi-layer feed-forward neural network. Most of the ANN models use BP neural network or its changing forms. [13] A BP neural network consists of an input layer, a number of hidden layers and one output layer and realizes the whole connection between each layer. But there is no connection between each layer's neuron units themselves.

The basic idea of the BP neural network algorithm is the learning process that propagating the signal forward and propagating the error backward. This is one of the reasons for the BP neural networks named. Figure 1 shows a basic structure of the BP neural network. In Fig. 1, w means weight and b represents bias item, and each layer has a threshold. In the forward propagation, data is inputted via input layer and sequentially processed in each hidden layer until reach the output layer. If the output is not expected, the error between the real output and the expected output will be propagated backward as adjustment signal. The network will adjust its weight and threshold according to the error repeatedly until the output is come to the expectancy or the number of iterations or other settings has reach to a limit and so on.

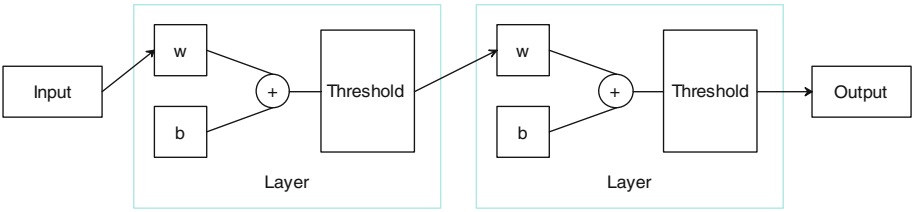

Fig. 1. The structure of BP neural network.

Robert Hecht-Nielsen [14] has proven that a three layers BP neural network which including one hidden layer can approximate any kind of continuous function effectively. So in this paper, we choose a simple three layers BP neural network to train our classifier. The number of input units is the dimensionality of each image's feature vector and the output number is the number of the vehicle type.

The selection of hidden layer units' number is a very important but complicated problem. It's usually directly connected with the number of input and output units. If the number is too small, it would be difficult for the network to get enough information to solve the classification problem; while if the number is too large, it would not only increase the training time, but may not get the best performance and cause the problem of over-fitting which may cause the test error increases and lead to generalization ability of the classifier drop. Therefore the reasonable choice of the number of hidden layer units is very important.

Usually the designers choose the number of the hidden layers by experience and many times experiments. This paper takes the empirical formula presented in [4] as a reference.

$$h = \sqrt{n + m} + \alpha \tag{1}$$

In Eq. (1), h is the number of the hidden layer units, and n is the number of the input units; m is the number of output units and α is a constant between 1 and 10.

We set the number of the hidden layer units mainly according to Eq. (1) and at the same time we also adjust the number of the layer to improve the performance.

2.2 Spatial Pyramid Representation

In this part, we will introduce the Spatial Pyramid Representation (SPR). SPR is a widely used method in image recognition for embedding spatial information into a feature vector. [15] Lazebnik et al. [12] present the SPM method and get a significantly improved result on the recognition tasks. We use the same Spatial Pyramid Representation method as [12].

The images we used in the task of classification usually in different sizes or in different scales. It is important for us to find a standard in term of the size of image to avoid reduce the performance. In some works, the images are simply cut or resized into a fixed size before object classification. But these kinds of size normalization usually result geometric distortion, or change the content of the image. This will reduce the performance of the

classification. Besides, the appearance of vehicle are usually similar to each other, normalizing the size of the images simply may make the feature less obviously, so is not a very suitable choice for the vehicle type classification. The spatial pyramid representation model [9, 12, 15] is an effective way for us to solve the problem.

The structure of the spatial pyramid shows as the Fig. 2. It's a spatial pyramid with three levels. The spatial pyramid can also be viewed as expansion of the BoW (Bag of Words or Bag of Keypoints) [16], and a 0 level spatial pyramid is a simple BoW representation.

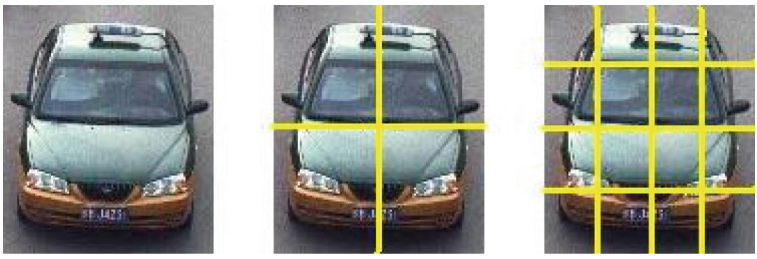

Fig. 2. The three-level spatial pyramid used in this paper.

We divide the images into three different levels of resolution and count dense SIFT feature descriptors in each spatial cell. The SPR defined the number of cells at level L as Eq. (2), and divided each cell into four cells at the next level of the pyramid which shows in Fig. 1. As the number of levels increases, the feature vector becomes large, so the SPR is usually used up to three levels (L = 2) [15], and it is the reason why we choose a three-level spatial pyramid. We calculate the histogram of the number of the dense SIFT features that fall in each spatial bin by the k-means method, and then we normalize all histograms by all features in the image like [12]. For a k cluster centers used in calculate the histogram, the dimensionality D in L level is counted like Eq. (3). We don't use the pyramid match kernel [17], but we put the same weight on our SPR model as [12] and then connect the vector with weight of each level together directly. The whole dimensionality of the SPR vector is come to a total of M (Eq. (4)).

$$c(l) = 4^l \tag{2}$$

$$D(l) = kc(l) = k4^l \tag{3}$$

$$M = \sum_{l=0}^{L} D(l) = \sum_{l=0}^{L} k4^l \tag{4}$$

3 Vehicle Dataset

In this section, we will introduce the vehicle dataset we used in the experiment. In order to imitate the realistic scenario in our life we collect a set of vehicle dataset from the

video of a real life vehicle scene. (Figure 3) We use the same dataset as [18]. The images are divided into three types and tags are put on them. The three types of vehicle are car, bus and truck which are common to see in our life. The vehicle images we used are as the Fig. 4 shows, and we divide them into three types.

Fig. 3. The screenshot of the video we used in the experiment.

(A) Car

(B) Bus

(C) Truck

Fig. 4. Some examples from the dataset we collected. The vehicle images are captured by hand and collected in different time and illumination.

We choose the most common types of vehicle appearing in the video. Figure 3 shows a screenshot in the video. The videos are collected in the same angel and a certain spot but different in time. It is not difficult for us to think of that, the illumination may vary in different time of a day or even in the close time of different days the illumination may be different. So it's important to find a proper method to reduce the influence of the light of the classification and the SIFT descriptor is a good choice. [12, 16]

Another challenge of the classification is the various scales of the vehicle images. We cut out the vehicle images by hand and it's really a tough task to make the vehicle images are in a similar scale.

We use a 900 images dataset which include 300 images per type.

4 Experimental Evaluation

In order to compare our multi-size vehicle classification method with the fix-size method, we divide our experiment into two parts: one is experiment with multi-size images, and the other is the experiment with fixed-size images. In these two kinds of experiment, we use the method we mentioned before. The three-fold cross validation method is used in our experiment.

The number of input layer units is the dimensionality of each image's feature vector and the output unit's number is the number of the vehicle types.

Experiment with Multi-size Images. This time, we use the original size of images directly. In order to get a common vector representation of each images, we extract dense SIFT descriptors of each image and make the feature vectors in the representation of the SPR model.

We choose a different number of k-means method's cluster centers, and set the hidden layer unit's number at 40 to get a compromise between the performance and the training time. We get the performance like the Table 1 shows.

Table 1. Performance with multi-size images.

Cluster center numbers	40	80	120	160	200	300
Performance	79.3 %	86.7 %	88.3 %	89.3 %	90.7 %	91.0 %

We compare our method to the methods which are used in [18]. Table 2 shows the performance in [18].

Table 2. Performance in [18].

Feature types	SIFT	Surf	Eigenface	SIFT + Surf	SIFT + Eigen face	Surf + Eigen-face	SIFT + Surf + Eigenface
performance	73.7 %	70.9 %	67.0 %	77.3 %	81.7 %	80.7 %	89.3 %

From Table 1 and Table 2, we can know that our method achieve better performance on the classification of the vehicle type than the methods in [18].

Experiment with Fixed-Size Images. In this part, we simply resize the vehicle images into a same size, and find the connection between image size and the performance. Obviously, there are some deformations of the vehicle (Fig. 5) after resizing. We take the whole picture as a vector to train the classifier. The performance shows

in the Table 3. The number of BP neural network's input units is the dimensionality of each image. We also set 40 units of the hidden layer to compare with the multi-size experiment.

Table 3. Performance with fixed-size images.

Size	32*32	64*64	128*128
Performance	41.29 %	72.73 %	76.52 %

Fig. 5. A car which are resized into different size. The 32*32 means to resizing the picture into 32×32 pixels.

From Table 3, we can find that the larger the size of the images is the better performance will be. But in the experiment we also find that as the size improved, the training time are increased at the same time.

The dimensionality of feature vector in the fixed-size is 4096 at the 64*64 size. And when the cluster center numbers is 200 the feature vector is 4200. At a similar dimensionality level, the multi-size method gets better performance than the fixed-size one.

5 Conclusion

As we said above, the SPR method is easier to implement than those geometric-based methods. And it gets better performance than the fixed-size image method. By combining the SPR and the BP neural network method, we implement an effective method in the field of vehicle type classification in a real life scenario. The dataset in our experiment is a little small. So we will enlarge our dataset and improve the performance of our algorithm in the future.

References

1. Ma, B.: Vehicle Identification Technology In Video surveillance. D. Xidian University, Xi'an (2010) (in Chinese)(马蓓: 车型识别技术在视频监控中的应用. 硕士学位论文. 西安电子科技大学, 西安 (2010))
2. Zhou, X.: A recognition of automobile types method based on the BP neural network. J. Microelectron. Comput. **20**(4), 39–41 (2003). (in Chinese) 周红晓: 基于 BP 神经网络的汽车车型识别方法. J. 微电子学与计算机. 20(4), 39-41 (2003))
3. Hu, F., Jian, Q., Zhang, X.: The classifier of car type using BP neural networks. J. Xianan Univ. **32**(3), 439–442 (2005). (in Chinese) (胡方明, 简琴, 张秀君: 基于 BP 神经网络的车型分类器. 西安电子科技大学学报. J. 32(3), 439-442 (2005))
4. Wu, Z.: Research on vehicle type recognition based on BP neural network. J. Mod. Comput. **2**, 38–41 (2013). (in Chinese) (吴志攀: 基于 BP 神经网络的车型识别研究. J. 现代计算机 2, 38-41 (2013))
5. Du, H.: Implementation of BP algorithm using matlab based on vehicle type recognition. J. Comput. Modernization. **5**, 20–22 (2012). (in Chinese) (杜华英: 基于车型识别的 BP 算法 Matlab 实现. J. 计算机与现代化.5, 20-22 (2012))
6. Ghada, S.M.: Vehicle type classification with geometric and appearance attributes. Int. J. Civil Architectural Sci. Eng. **8**(3), 273–278 (2014)
7. Kang, W., Cao, Y., Sheng, Z., Li, P., Jiang, P.: Harris corner and SI FT Feature of vehicle and type recognition. J. Harbin Univ. Sci. Technol. **17**(3), 69–73 (2012). (in Chinese) (康维新, 曹宇亭, 盛卓, 李鹏, 姜澎: 车辆的 Harris 与 SIFT 特征及车型识别. J. 哈尔滨理工大学学报. 17(3), 69-73 (2012))
8. Zhu, F., Jia, J., Mi, X.: Vehicle Image Classification Based on Sparse Coding. J. Video Eng. **37**(11), 198–202 (2013). (in Chinese) (朱福庆, 贾世杰, 米晓莉: 基于稀疏编码的车型图像分类研究. J. 电视技术.37(11), 198-202 (2013))
9. He, K. et al.: Spatial pyramid pooling in deep convolutional networks for visual recognition. arXiv preprint arXiv:1406.4729 (2014)
10. Gan, J., Youwei, Z.: Face recognition based on BP neural network systems. Eng. Electron. **25**(1), 113–115 (2003). (in Chinese) (甘俊英, 张有为: 基于 BP 神经网络的人脸识别. 系统工程与电子技术.25(1), 113-115 (2003))
11. Shao, H., Xu, Q., Cui, C.: Research of Human Face Recognition Method Based on BP Neural Network. J. Shenyang Univ. Technol. **22**(4), 346–348 (2000). (in Chinese) (邵虹, 徐全生, 崔文成: 基于 BP 神经网络的人脸图像识别方法的研究. J. 沈阳工业大学学报. 22 (4), 346-348 (2000))
12. Lazebnik, S., Cordelia S., Jean P.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2. IEEE (2006)
13. Huichao, Q., Hongping, H., Yanping, B.: BP neural network classification on passenger vehicle type based on GA of feature selection. J. Meas. Sci. Instrum. **3**(3), 251–254 (2012)
14. Hecht-Nielsen, R.: Theory of the backpropagation neural network. In: International Joint Conference on Neural Networks, 1989 IJCNN. IEEE (1989)
15. Tatsuya, H., et al.: Discriminative spatial pyramid. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2011)
16. Gabriella, C., et al.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV. Vol. 1, pp. 1–22 (2004)

17. Kristen, G., Darrell, T.: The pyramid match kernel: discriminative classification with sets of image features. In: Tenth IEEE International Conference on Computer Vision, 2005, ICCV 2005, vol. 2, IEEE (2005)
18. Ma, W.: Vehicle Classification Methods Research Based on Multi-feature Fusion. D. Beijing Jiaotong University, Beijing (2014). (in Chinese) (马文华: 基于多特征融合的车型分类方法研究. 硕士学位论文. 北京交通大学, 北京 (2014))