Road Marking Detection and Classification Using Machine Learning Algorithms

Tairui Chen*, Zhilu Chen*, Quan Shi+ and Xinming Huang+*
+Nantong University, Nantong, Jiangsu, China 226019
+Worcester Polytechnic Institute, Worcester, MA 01609, USA
Corresponding author email: xhuang@wpi.edu

Abstract—This paper presents a novel approach for road marking detection and classification based on machine learning algorithms. Road marking recognition is an important feature of an intelligent transportation system (ITS). Previous works are mostly developed using image processing and decisions are often made using empirical functions, which makes it difficult to be generalized. Hereby, we propose a general framework for object detection and classification, aimed at video-based intelligent transportation applications. It is a two-step approach. The detection is carried out using binarized normed gradient (BING) method. PCA network (PCANet) is employed for object classification. Both BING and PCANet are among the latest algorithms in the field of machine learning. Practically the proposed method is applied to a road marking dataset with 1,443 road images. We randomly choose 60% images for training and use the remaining 40% images for testing. Upon training, the system can detect 9 classes of road markings with an accuracy better than 96.8%. The proposed approach is readily applicable to other ITS applications.

Index Terms—machine learning, BING, PCANet, road marking

I. INTRODUCTION

Object detection and classification have attracted considerable interests from researchers in recent decades. Various databases are built to evaluate the latest object detection and classification algorithms, such as the Caltech101 [1] and Caltech256 [2], Pascal visual object classes dataset [3], ETHZ shape classes [4], face detection dataset, and etc. These datasets have been broadly used as benchmarks for new algorithm development and performance comparison.

In recent year, the approach of machine learning has become increasingly popular to explore the structures or algorithms that a system can be programmed to learn from data or experience. It has been widely used in computer vision, search engines, gaming, computational finance, robotics and many other fields. Since Hinton et al. proposed an effective method to train the deep belief networks [5] in 2006, deep learning networks have gained lots of attentions in the research community. Deep learning networks are able to discover multiple levels of representations of a target object. Therefore, they are particularly powerful for the tasks of pattern recognition. For instance, the convolution neural network (CNN) has demonstrated superior performance on many benchmarks [6], [7], although CNN requires significant computations. PCA network (PCANet) [8] is a type of deep learning networks that has been introduced recently. When compared to CNN, the structure of PCANet is much simpler, but it has been demonstrated as an effective method for image classification [8]. The PCANet architecture mainly consists of the following components: patch-mean removal, PCA filter convolutions, binary quantization and mapping, block-wise histograms, and an output classifier. More details about the PCANet algorithm will be discussed in Section III.

Advanced Driver Assistance System (ADAS) has become a main stream technology in the auto-industry. Autonomous vehicles, such as Google’s self-driving cars, are evolving and becoming reality. A key component is video-based machine intelligent that can provide information to the system or the driver to maneuver a vehicle properly based on the surrounding and road conditions. There have been lots of research works reported in traffic sign recognition [9], [10], lane departure warning [11], pedestrian detection [12], and etc. Most of these video-based object detection methods are developed using the classic image processing and feature extraction algorithms. For different types of objects, certain features usually works better than others as reported in the literature. Often, object detection is followed by a classification algorithm in these intelligent transportation applications. Typical classifiers, such as Support Vector Machine (SVM), artificial neural network, and boosting, are applied to identify one or multiple classes of the detected objects.

Figure 1. A set of 9 types of road markings are trained for the proposed system

Figure 1. A set of 9 types of road markings are trained for the proposed system
In this paper, we take a different approach focused on machine learning using data for object detection and classification. Instead of using typical features, we can train the object detector to certain types of objects existed on the images using the method of binarized normed gradients, also referred as BING [13]. Given different levels of sensitivity thresholds, BING can provide a number of possible candidates that have some degree of similarity to the interested class of objects. We do not apply additional filtering on these candidates. Subsequently, a PCANet classifier is employed to identify these candidates if they belongs to any class of targeted objects. Using the road marking dataset with 1,443 road images [14], the proposed method is applied to road marking detection. After training with part of the data set, the BING detection algorithm results a total of 30 candidates. These 30 candidates are then classified by the PCANet that is trained to identify 9 different types of road markings plus 1 negative type. Fig. 1 shows a set of 9 road markings that the system has been trained to identify from road scenes.

The rest of the paper is organized as follows. Section II describes the related work on road marking detection and classification. The proposed machine learning methods, BING and PCANet, are elaborated in details in Section III. Section IV gives the experimental results, followed by conclusion in Section V.

II. RELATED WORK

Road marking detection is an important topic in Intelligent Transportation System (ITS) and has been researched extensively. As described in [15], many previous works were developed based on various image processing techniques such as edge detection, color segmentation and template matching. Road marking detection can also be integrated as part of a lane estimation and tracking system [16]. The lane borders and arrow markings were detected using scan-lines and template matching methods. The information of the lane types, i.e. forward, left-turn, and right-turn, were sent to the console or the driver. In [17], it presented a method of lane detection. Lines were extracted from the original image through edge detection, following by some rule-based filtering to obtain the candidates of lanes. Additional properties such as brightness and length of the lines were examined to detect the lanes. [18] was able to detect and recognize lanes, crosswalks, arrows and many other markings on the road. The road marking on an image were extracted first using a modified median local threshold method. The road displayed on the image was a trapezoidal area due to the effect of camera angle and 3D space projection. Thus, road markings on the image also had distortions and variations in shape and size. Then perspective transform was applied to convert the trapezoidal road area into a rectangular area, which reduced the distortions and variations of the road marking, making it easier for detection. Similarly, perspective transformation was also applied in [19]. The lanes were detected using Augmented Transition Network (ATN). Subsequently, the detected lanes were used to locate the Region of Interests (ROIs) on an image for detecting other road marking such as arrows. In [14], the Maximally Stable Extremal Regions (MSERs) was employed as an effective way of detecting region of interest. Both Histograms of Oriented Gradients (HOG) [12] features and template matching methods were used for classification.

III. PROPOSED METHOD

We propose a system that is capable of detecting and recognizing different road markings. We use BING feature to find and locate the potential objects on a road image, i.e. road markings. The potential objects are then classified by a PCANet [8] classifier to obtain the final results. Unlike the traditional approach of tuning image processing techniques geared specifically for road marking detection, our system is an extendable framework that can be adopted to other detection and classification tasks.

A. BING Feature for Detection

The BING feature is employed to find the potential objects in an image. It is the binary approximation of the 64D norm of the gradients (NG) feature. Each image window is resized to 8 × 8 pixels for computational convenience, and its norm of the gradients forms the 64D NG feature. It represents the contour of the target object in a very abstracted view with little variation. Thus, the BING features can be used to find objects in an image. It is very efficient in computations compared to some existing feature extract algorithms. The BING feature is suitable for finding road markings, because the road markings have closed boundaries and high gradients around the edges.

In order to locate the target objects in an image using the BING feature, we need to train it with training samples. The positive samples are true objects manually labeled in images and the negative samples are the background in images. The machine learning method inside BING is actually linear SVM. It is observed that some window sizes (e.g. 100 × 100 pixels) are more likely to contain objects than other sizes (e.g. 10 ×
be varied. A typical PCANet has two stages. According to [8], followed by an output stage. The number of PCA stages can consist of a PCANet and a multi-class SVM. The structure to recognize the true road markings. The PCANet classifier produces 30 candidates through object detection. Locations are not precise. Fig. 2 shows an example that BING true objects can still be recognized even if the bounding box locations by multiple runs of BING detection. Therefore, the collected a large number of true objects at various bounding box locations. For the problem of inaccurate bounding box locations, we have similar BING feature as the true objects, and they may still be selected as potential objects. Secondly, as a bounding box based detection algorithm, it has the common problem that a bounding box may not accurately locate the true object. Such inaccuracy may cause failure in the subsequent recognition stage. However, these limitations can be alleviated or overcome. As a fast object detection method, we manually assign an arbitrary number that represents the number of potential objects to be selected by the detector, according the their confidence values. This number is often much large than the number of true objects in an image. For example, BING may provide 30 potential objects from an image, while there are only one or two true objects in it. Therefore, the true objects are unlikely to be missed, but adversely many false objects might also be included in the candidates pool. We deal with the false candidates in the classification step using PCANet. For the problem of inaccurate bounding box locations, we have collected a large number of true objects at various bounding box locations by multiple runs of BING detection. Therefore, the true objects can still be recognized even if the bounding box locations are not precise. Fig. 2 shows an example that BING produces 30 candidates through object detection.

B. PCANet for Classification

Taking the detection results from the BING stage, we build a PCANet classifier to filter out the false candidates and to recognize the true road markings. The PCANet classifier consists of a PCANet and a multi-class SVM. The structure of PCANet is simple, which includes a number of PCA stages followed by an output stage. The number of PCA stages can be varied. A typical PCANet has two stages. According to [8], the two-stage PCANet outperforms the single stage PCANet in most cases, but increasing the number of stages does not always improve the classification performance significantly, depending on the applications. In this work, we choose two-stage PCANet.

To a certain extend, the structure of PCANet is to emulate a traditional convolutional neural network [20]. The convolution filter bank is chosen to be PCA filters. The non-linear layer is the binary hashing (quantization). The pooling layer is the block-wise histogram of the decimal values of the binary vectors. There are two parts in the PCA stage: patch mean removal and PCA filters for convolution. For each pixel of the input image, we have a patch of pixels whose size is the same as the filter size. We then remove the mean from each patch, followed by convolutions with PCA filters. The PCA filters are obtained by unsupervised learning during the pre-training process. The number of PCA filters can be variant. The impact of the number of PCA filters is discussed in [8]. Generally speaking, more PCA filters would result better performance. In this paper, we choose the number of filters equals to 8 for both PCA stages. We find that it is sufficient to deliver desirable performance. The PCA stages can be repeated multiple times as mentioned above, and here we choose to repeat it only once.

The output stage consists of binary hashing and block-wise histogram. The output of PCA stages are converted to binary values by a step function, which converts positive values to 1 and else to 0. Thus, we obtain a binary vector for each patch and the length of this vector is fixed. We then convert this binary vector to decimal value through binary hashing. The block-wise histogram of these decimal values forms the final output features. We then feed the SVM with the features from PCANet. Fig. 4 shows the structure of a two-stage PCANet. The number of filters in stage 1 is m and in stage 2 is n. The input images are object candidates from BING.

IV. EXPERIMENTAL RESULT

In our experiments, we evaluate the proposed system using the road marking dataset provided by [14]. The dataset contains 1,443 road images, each with size of 800 × 600 pixels. There are 11 classes road marking in these images. In this paper, we evaluate 9 of them because the data of the other 2 classes are insufficient for machine learning. We train the object detection model by manually labeling the true road markings in the images. The PCANet model is trained iteratively to ensure its accuracy. The initial training samples are manually labeled from a small portion of the dataset, and the trained model along with the object detection model is applied to the whole dataset to detect road markings. The results are examined and corrected by human interference in order to ensure the correctness of the data during the next training iteration. Through the iterative procedure, one road marking on an image can be detected multiple times and generates multiple training samples. Because of the utilization of the BING feature and its object detection model, the true samples may be extracted using various bounding boxes, making the PCANet classifier more robust.
We measure the performance of our PCANet classifier by using 60% images for training and 40% images for test. The 1,443 images are re-ordered randomly and thus the training and test images are selected randomly without overlap. The window-sized training samples and test samples are from the training images and test images respectively. We perform data augmentation over the collected samples by transforming the original images with parameters such as roll, pitch, yaw, blur and noise. Table I shows the evaluation results of the PCANet classifier, which is referred as the confusion matrix. The test samples for each class is 250. The cell at the $i$th row and the $j$th column gives the percentage that the $i$th samples are recognized as the $j$th samples. The “OTHERS” class represents negative samples without road marking. Comparing to the previous results in [14], our classification accuracy is more consistent and significantly better especially for the “FORWARD” sign.

We also test our proposed system upon the 1,443 road images in the original dataset [14]. Visual examination observes desirable output. An example of stop marking is detected in Fig. 5 and an example of right arrow marking is detected in Fig. 6.
V. Conclusion

In this paper, we present a framework for object detection and classification using the latest machine learning algorithms including BING and PCANet. BING can quickly identify the target classes of objects after the system is trained with a set of images with target objects. Subsequently, these detected objects are classified by the PCANet classifier. Similarly, the classifier is also pre-trained using the dataset and is capable of identifying many types of objects simultaneously. As an example, we demonstrate this approach by building a system that can detect and identify 9 classes of road marking at very high accuracy. More importantly, the proposed approach can be employed for many other video-based ITS applications provided that sufficient training datasets are available.

Acknowledgment

This work was partially supported by The MathWorks Fellowship.

References


