

## Traffic Scenes Classification for Self-driving Car

Hongbo Lv<sup>1, a\*</sup>, Xiaolin Zhuang<sup>1, b</sup> and Huifang Cao<sup>1, c</sup>

<sup>1</sup>School of Mechanical and Materials Engineering, North China University of Technology

No.5 Jinyuanzhuang Road, Shijingshan District, Beijing, China

<sup>a</sup>lhb@ncut.edu.cn, <sup>b</sup>qingqingzxl\_108@163.com, <sup>c</sup>xintaiannie@163.com

**Keywords:** Traffic Scenes Classification, Self-driving Car, Road Network Definition

**Abstract.** In this paper, machine learning model is applied to label the attributes of traffic scenes automatically in the RNDF (Road Network Definition File), which is used in the self-driving car. The “gist” features extracted from one image of the video stream are used as the model’s input. In the experiments, the test model is SVM (Support Vector Machine) and the training and test samples are from the FM2 database. The experiments verify the feasibility and effectiveness of this method.

### Introduction

It is always necessary to define the road network for running a self-driving car. And the attributes of the waypoints in the road network is important because different control strategies is needed for different traffic scenes. Usually, the attributes of the waypoints have to be added manually which is boring and time-consuming. In this paper, a method to classify the traffic scenes automatically, which is one of the waypoint’s attributes, is proposed. This method is similar to the scene classification methods based on image [1-3] expect that it is applied in traffic scenes.

Roughly, there are three types methods used for scenes classification, the first is based on objects, the second is based on regions, and the third is based on context [4-7]. On the other hand, in recent years, SIFT (scale-invariant feature transform)[8] and “gist”[9] are two most frequently-used methods to extract features from scene images.

SIFT often works with bag-of-words, which needs more memory to store a large number of key points with different scales. And clustering is needed to build up dictionary from these key points. So the burden of the computer is heavy. Moreover, the number of the features extracted by SIFT for one traffic scene, as shown as arrows in Fig. 1, is often different from another because it is outdoor. Therefore, it is difficult to train the SVM model. So, *spatial envelope* according with “gist” feature [9] was adopted in this paper.

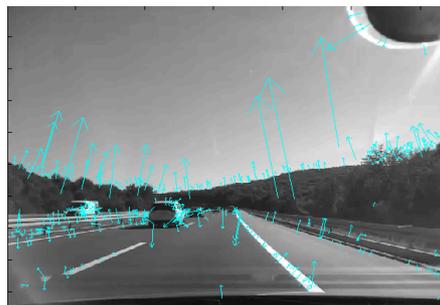


Fig. 1 Features extracted by SIFT of a traffic scene

### Related Work

The method in this paper is proposed to help construct the RNDF file automatically, which is used to create path for self-driving car THU-IV2, as shown in Fig. 2. In the standard of RNDF, one road consists of segments, lanes and waypoints, as shown in Fig. 3. Waypoint is the basic element and its attributes are used to decide the control strategy. For example, in Fig. 3, the position of the first waypoint 1.1.1 is (-12512.5, -14830) relative to the global coordinate system and its attributes are (1, 3, 20, 0). The last attribute “0” could be other number, which represents the type of the traffic scene. The method described in this paper is applied to help decide this attribute.

In normal conditions, when we want to run the self-driving car in a region, the topological map, which is described by RNDF format, has to be built-up first. In order to build up the map, we drive along all the roads in the region and record the path points of the DGPS (Differential Global Positioning System). Then essential points are selected as “waypoints” for the RNDF and the attributes of the waypoints are added by hand. In order to label the attributes automatically, at the same time of recording the DGPS points, the video stream of the color camera in the roof of the car could be saved and the synchronous information is registered also. The images taken from the video stream could be used to classify the traffic scenes using machine learning method.

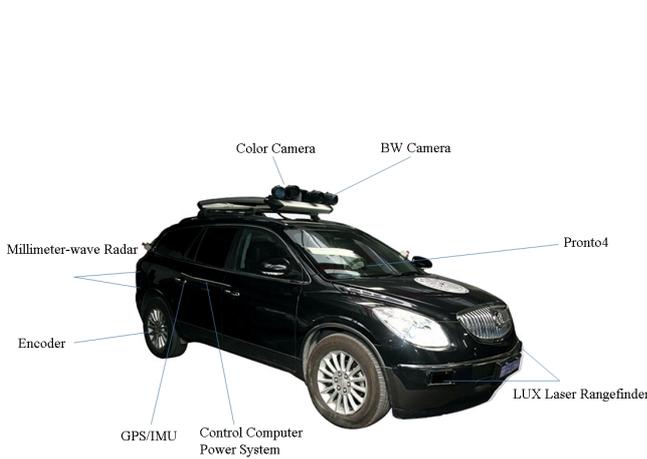


Fig. 2 Self-driving car THU-IV2

```

RNDF.txt
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
RNDF_name sample_RNDF
num_segments 12
num_zones 0
format_version 1.1
creation_date 2014-10-22
segment 1
segment_name 1_road
num_lanes 2
lane 1.1
num_waypoints 2
waypoint
1.1.1 -12512.5 -14830 1 3 20 0
1.1.2 -12454.687629 -14532.99264 2 3 20 0
end_waypoint
end_lane
lane 1.2
num_waypoints 2
waypoint
1.2.1 -12464.5986 -14523.81918 1 3 20 0
1.2.2 -12522.539522 -14821.295856 2 3 20 0
end_waypoint
end_lane

```

Fig. 3 A simple example of RNDF

## Method and Experiment

As mentioned, to gain the type of a traffic scene, the last attribute of the waypoint in the RNDF, is a task of classification just like scene classification. Considering the efficiency and accuracy, SVM is selected as the classification model, which could be described by Eq.(1). The model’s feature is vector  $\mathbf{x}$ , which could be calculated by “spatial envelope” referring to [9]. And the parameters of the SVM classifier could use the open library to solve out, such as LIBLINEAR, LIBSVM, SVM-Light and GPUSVM.

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (1)$$

Binary-classification experiments had been finished in order to verify the feasibility of the method. For the convenience of comparative analysis and saving time, the FM2 dataset [10,11] was used. The database contains 6237 labeled images of traffic scenes which were extracted from videos recorded on European roads by smartphone’s built-in camera. The images’ resolution is 640x480. Just like [11], eight classes were employed for our experiments: highway, open non-highway road, tunnel, tunnel exit, settlement with visible buildings, overpass, toll booth and dense traffic. The Examples of the eight classes from the FM2 dataset are shown in Fig. 4.

Our experiments were finished in Matlab using LIBSVM [12]. A total of 512 gist feature for each image were extracted using the open code with paper [9] by Aude Oliva and Antonio Torralba. Twenty-eight binary-classifiers have been trained for all eight types of traffic scenes. Not all the images in the database were chosen as training samples, but to choose some random so as to improve the training speed. Gaussian RBF kernel function was selected and cross validation method was used to choose parameters. The test result of the models is shown in Table 1. The data in the middle of the table is skew symmetric. Each result consists of the classification precision and “the correct number/the test number”. [11]



Fig. 4 Examples of classes from the FM2 dataset

To each binary-classification, nineteen out of twenty-eight test results are higher than 96% and twelve are higher than 99%. Five of the lowest are 64.44% of road vs. settlement, 67.74% of overpass vs. booth, 71.79% of booth vs. traffic, 77.27% of settlement vs. traffic and 79.47% of highway vs. road. On the other hand, considering of multi-labels, the total precisions is lower. The total precision of the tunnel is highest as 97.39% and the settlement is the lowest as 42.13%. The reason to the low total precision is that some binary-classification is low for the scenes' similarity.

Table 1 Binary Classification Results

|            | highway           | road              | tunnel            | exit              | settlement        | overpass          | booth             | traffic           | total  |
|------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------|
| highway    |                   | 79.47%<br>360/453 | 100%<br>438/438   | 100%<br>351/351   | 97.26%<br>390/401 | 99.43%<br>351/353 | 99.43%<br>350/352 | 98.06%<br>354/361 | 74.93% |
| road       | 79.47%<br>360/453 |                   | 99.54%<br>216/217 | 100%<br>130/130   | 64.44%<br>116/180 | 96.21%<br>127/132 | 96.95%<br>127/131 | 89.29%<br>125/140 | 42.45% |
| tunnel     | 100%<br>438/438   | 99.54%<br>216/217 |                   | 97.39%<br>112/115 | 93.94%<br>155/165 | 100%<br>117/117   | 100%<br>116/116   | 98.40%<br>123/125 | 89.61% |
| exit       | 100%<br>351/351   | 100%<br>130/130   | 97.39%<br>112/115 |                   | 100%<br>78/78     | 100%<br>30/30     | 100%<br>29/29     | 100%<br>38/38     | 97.39% |
| settlement | 97.26%<br>390/401 | 64.44%<br>116/180 | 93.94%<br>155/165 | 100%<br>78/78     |                   | 97.50%<br>78/80   | 89.87%<br>71/79   | 77.27%<br>68/88   | 39.87% |
| overpass   | 99.43%<br>351/353 | 96.21%<br>127/132 | 100%<br>117/117   | 100%<br>30/30     | 97.50%<br>78/80   |                   | 67.74%<br>21/31   | 92.50%<br>37/40   | 58.45% |
| booth      | 99.43%<br>350/352 | 96.95%<br>127/131 | 100%<br>116/116   | 100%<br>29/29     | 89.87%<br>71/79   | 67.74%<br>21/31   |                   | 71.79%<br>28/39   | 42.13% |
| traffic    | 98.06%<br>354/361 | 89.29%<br>125/140 | 98.40%<br>123/125 | 100%<br>38/38     | 77.27%<br>68/88   | 92.50%<br>37/40   | 71.79%<br>28/39   |                   | 44.21% |

It can be found to analyze the images of the database FM2 that even with the eyes of people it is hard to make a distinction between some images in the five lowest groups without context information. From another point of view, because our goal is to label the attribute of the waypoints, so sequential images nearby the waypoint can be used for classification to improve the precision and higher precision is completely possible. Moreover, the ultimate purpose is to decide the control strategies according to the traffic scene type attributes. For example, one of the control strategies is to modify navigation method in the area lacking of GPS signal. So according to this strategy, tunnel,

booth and overpass could be a group or a big type. Similarly, in accordance with the speed strategy and the obstacle detection strategy, open non-highway road and highway should be a big type, and settlement, booth and traffic should be another. So similar images are in a type and the precision would be improved.

## Conclusion and Future Work

Learning from common scenes classification, method of machine learning is engaged to label the waypoints' attributes in RNDF for self-driving car. Taken in account of the storage space and efficiency, "gist" feature is applied as model input. In order to verify the method, SVM is used as model and the database FM2 as samples, and binary-classification experiments have been finished. According to the results of the experiments and the reason behind low precision, the method is effective and able to help decide the attribute of the waypoints automatically in the RNDF for the self-driving car. Off course, more work is needed, such as how to make use of the sequential images, how to arrange the types according to the control strategies, and experiments in our self-driving car THU-IV2.

## Acknowledgements

This work was financially supported by the General Program of Science and Technology Development Project of Beijing Municipal Education Commission of China.

## References

- [1] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, Learning multi-label scene classification, *Pattern recognition*, vol. 37, no. 9, pp. 1757-1771, 2004.
- [2] A. Bosch, A. Zisserman, and X. Muñoz, Scene classification via pLSA, *Computer Vision—ECCV 2006*, pp. 517-530: Springer, 2006.
- [3] A. M. Cheriyyadat, Unsupervised feature learning for aerial scene classification, *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 1, pp. 439-451, 2014.
- [4] W. Feng, *Research on Key Technologies for scene classification*, Beijing Jiaotong University, 2008.
- [5] W. Xie, *Research on Middle Semantic Representation Based image scene classification*, Beijing Jiaotong University, 2011.
- [6] J. Liu, *Research on the fast scene classification based on gist of a scene*, Jilin University, 2013.
- [7] Y. Zhao, G. Jun, X. Zhao, and W. Kewei, Scene categorization of local Gist feature match kernel, *Journal of Image and Graphics*, vol. 18, no. 3, pp. 264-270, 2013.
- [8] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [9] A. Oliva, and A. Torralba, Modeling the shape of the scene: A holistic representation of the spatial envelope, *International journal of computer vision*, vol. 42, no. 3, pp. 145-175, 2001.
- [10] Information on <http://www.zemris.fer.hr/~ssegvic/datasets/unizg-fer-fm2.zip>.
- [11] I. Sikiric, K. Brkic, J. Krapac, and S. Segvic, Image representations on a budget: Traffic scene classification in a restricted bandwidth scenario. pp. 845-852.
- [12] C.-C. Chang, and C.-J. Lin, LIBSVM: a library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.