

ROAD SIGNS DETECTION AND RECOGNITION UTILIZING IMAGES AND 3D POINT CLOUD ACQUIRED BY MOBILE MAPPING SYSTEM

Y. H. Li^{a,*}, T. Shinohara^a, T. Satoh^a, K. Tachibana^a

^a PASCO Corporation, 2-8-10 Higashiyama, Meguro-ku, Tokyo 153-0043, Japan
(yjiorn_3951, taarkh6651, tuost7017, kainka9209)@pasco.co.jp

Commission I, ICWG I/Va

KEY WORDS: Road Signs Detection and Recognition, Mobile Mapping System (MMS), Object Based Image Analysis, Multi-Scale Image Segmentation, 3D Point Cloud

ABSTRACT:

High-definition and highly accurate road maps are necessary for the realization of automated driving, and road signs are among the most important element in the road map. Therefore, a technique is necessary which can acquire information about all kinds of road signs automatically and efficiently. Due to the continuous technical advancement of Mobile Mapping System (MMS), it has become possible to acquire large number of images and 3d point cloud efficiently with highly precise position information. In this paper, we present an automatic road sign detection and recognition approach utilizing both images and 3D point cloud acquired by MMS. The proposed approach consists of three stages: 1) detection of road signs from images based on their color and shape features using object based image analysis method, 2) filtering out of over detected candidates utilizing size and position information estimated from 3D point cloud, region of candidates and camera information, and 3) road sign recognition using template matching method after shape normalization. The effectiveness of proposed approach was evaluated by testing dataset, acquired from more than 180 km of different types of roads in Japan. The results show a very high success in detection and recognition of road signs, even under the challenging conditions such as discoloration, deformation and in spite of partial occlusions.

1. INTRODUCTION

In recent years, an extensive study has been going on regarding the practicability of automated driving. Among others, high-definition and highly accurate road maps are necessary for the realization of automated driving. Road signs are among the most important element, in the road map, which can be used as landmarks for the correction of the location of a self-driving-car. Therefore, a technique is necessary which can acquire information about all kinds of road signs automatically and efficiently. Due to the continuous technical advancement of MMS, it has become possible to acquire large number of images and 3d point cloud efficiently with highly precise position information.

Road sign recognition algorithms using images usually consist of two stages: road sign detection and road sign classification. In road sign detection, the aim is to extract the regions of interest (ROI) from an image. Then in the classification stage, the detected road sign candidates are assigned to the class that they belong. For the detection stage, there are color based detection (Lopez and Fuentes, 2006), shape based detection (Loy and Barnes, 2004) and the combination of the two approaches (Adam and Ioannidis, 2014). Detection stage is very important because the road signs that are not detected during this step is not available to be recovered later. Classification shall be conducted using traditional template matching (Siogkas and Dermatas, 2006) or via techniques from the field of machine learning, such as Support Vector Machines (Maldonado-Bascon et al., 2007; Adam and Ioannidis, 2014) and deep learning (Seranet and LeCun, 2012; Cirean et al., 2012). Especially, Cirean et al., (2012) presented a state-of-the-art road sign classification approach using deep neural network, which won the German

traffic sign recognition benchmark with a recognition rate of 99.46%, better than the one of humans on this task. But, for the good performance, machine learning based approaches need plenty of sample data which is not easy to collect.

As we know, laser points are not only insensitive to illumination but also have 3D spatial information. Hence in road sign detection, an approach which utilize both of images and point cloud should be more effective, than the approaches based only on images. However, so far, not many researches are conducted about the former, comparing to the later. Soilan et al., (2016) proposed an approach which combines laser point based detection and image based classification. Shi et al., (2008) presented a pipeline to extract and classify road signs using fusion-base processing of multi-sensor data including images and point cloud.

The aim of this research is to create high-definition road maps more efficiently through automating the process of road sign recognition. Since the automatic sign recognition is hard to perform perfectly, manual works such as check and correction are still necessary. Usually, in the manual work, the correction of incomplete detections (false negatives) is more costly, compared with that of false classifications or over detections (false positives). Therefore, the recall of sign detection is the most important element in the evaluation of road sign recognition approach. In this paper, we present a novel automatic road sign recognition technique utilizing both images and 3D point cloud acquired by MMS.

* Corresponding author

2. METHODOLOGY

The proposed approach consists of three stages: 1) detection of road signs from high resolution images based on their color and shape features using object based image analysis method, 2) filtering out of false positives with size and position information estimated from point cloud, ROI of candidates and focal length of lens, and 3) road sign recognition using template matching method after shape normalization.

2.1 Detection

In detection stage, an object based image analysis method is utilized to detect road sign candidates from high resolution images based on color and shape information (Fig.1). The object based image analysis method is widely used for classification of object from high resolution aero or satellite images (Blaschke, 2010).

Firstly, multi-scale image segmentation is performed on the images acquired by camera mounted on the MMS. Different from color based segmentation, in the multi scale image segmentation both of color and shape information are considered (Baatz and Schape, 2000) and a segmented region is usually regarded as object. After image segmentation, the objects having specified color (hereinafter referred to as target objects), are extracted according to the hue and saturation value in HSV color space, as HSV is robust to illumination changes.

Red color : (hue > 270 or hue <40) and saturation > 0.15
Blue color : (hue > 190 and hue <290) and saturation > 0.15
Yellow color : (hue > 10 and hue <100) and saturation > 0.15
Green color : (hue > 100 and hue <240) and saturation > 0.15

As shown above, we set the ranges of hue widely for robust extraction even in various illumination conditions e.g. backlight, halation, or shadow.

Secondly, the adjacent or nearer target objects which have similar colors are also merged together separately. For evaluating the similarity between two colors, the Euclidian distance in CIE Lab color space is adopted and when the Euclidian distance falls below the threshold then the colors are considered as similar. However it was found that only one threshold could not perform well due to complex illumination conditions, so we have utilized several thresholds to create various considerable merging results.

Finally, shape recognition is performed to the merged target objects in order to select road sign candidates according the combination of color and shape. A new convex hull based method, which is robust to occlusion and transformation, is proposed to classify 8 types of shapes. The proposed method focused on the length and direction of segments that are parts of convex hull created from target object area. More specifically, curves consist of several short segments with different directions and lines are consisting of one long segment or several short segments with similar directions.

2.2 Filtering

Usually, there are some regulations about size and installation height of road signs, e.g. the possible size of road signs are in range of 40cm~180cm and installation height also have to be within 1m~6m, in Japan. Therefore, the size and installation height of candidates can be used for filtering false positives. In this paper, the size and installation height of candidates are estimated using point cloud, ROI of candidates and focal length

of lens, instead of using only point cloud (Fig. 2). The proposed approach is still valid for low density laser points caused by high travelling speed and existence of displacement between laser point and image.

At first, the point cloud is projected to the images for matching between laser points and image pixels. Then the representative point of candidate is selected among the laser points included in ROI of candidate, via selecting the point which have minimum depth (distance from camera). The depth and the altitude of a representative point are considered as the depth and the altitude of the candidate instead. Therefore the real height of ROI H_{real} can be calculated by Eq. (2).

$$H_{real} = H_{image} / f * d \quad (2)$$

Where H_{image} is the height of ROI (pixel), f is focal length of lens (pixel) and d is depth (m). On the other hand, the installation height can be easily calculated by computing the difference between altitude of candidate and nearby ground. Because the altitude of candidate is calculated on the above process, the ground altitude is the only unknown variable. The ground altitude has to be estimated for each candidate independently as the ground surface is not necessarily flat. At first, the ground altitude is computed at each points of camera trajectory by calculating the median altitude of the laser points, which are closed to the trajectory point. Then the ground altitude of candidate is set with that of the nearest trajectory point from the candidate.

2.3 Classification

The remaining road sign candidates, after filtering, are classified by the traditional template matching approach. In template matching approach, unlike the machine learning base approaches, it is not necessary to collect a lot of sample data. However, the major drawback of template matching approach is the sensitivity to the deformation caused by perspective transformation in images. Hence, the shape normalization process is performed before template matching. In shape normalization, the outlines of candidates are extracted firstly. Then RANSAC based ellipse fitting, and Hough transformation based line extraction are performed to circular and polygonal shape candidates respectively. Furthermore, for the polygonal shape, the vertices are calculated from the extracted lines. According to the above results, the images of candidates are affine transformed for shape normalization.

The normalized image of candidate is converted to several kinds of grayscale images depend on their own color. For example, the candidates of red sign, are converted to grayscale1, 3, 4 and 6.

$$\left\{ \begin{array}{l} \text{grayscale1} = (5 \times R + G + B) / 7 \\ \text{grayscale2} = (R + 5 \times G + B) / 7 \\ \text{grayscale3} = (R + G + 5 \times B) / 7 \\ \text{grayscale4} = R / (R + G + B) \\ \text{grayscale5} = G / (R + G + B) \\ \text{grayscale6} = B / (R + G + B) \end{array} \right. \quad (3)$$

Where R , G and B indicate red, green and blue color respectively.

In the template matching, the Zero-mean Normalized Cross-Correlation (ZNCC) are calculated between image of candidate and that of templates which have same type of color and shape with candidate. Then the mean of ZNCC in several different types of grayscales is computed as similarity index. Finally, the template with maximum similarity is chosen as result of

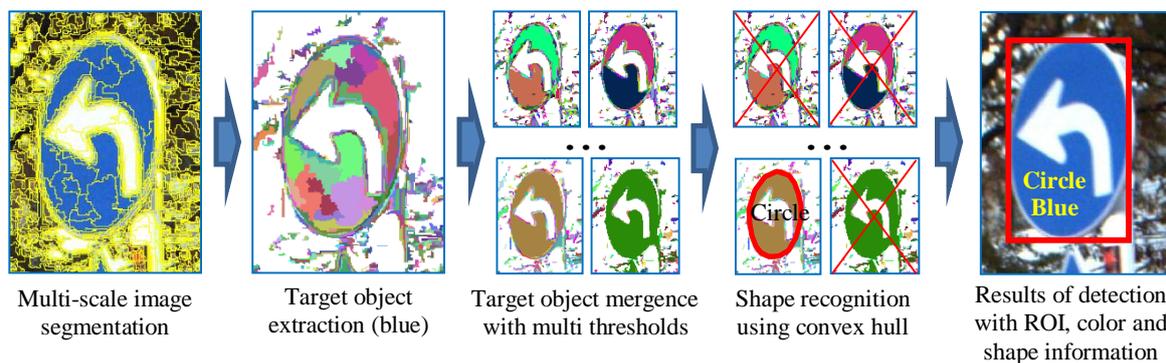


Fig.1 Detection of road sign candidates from MMS image

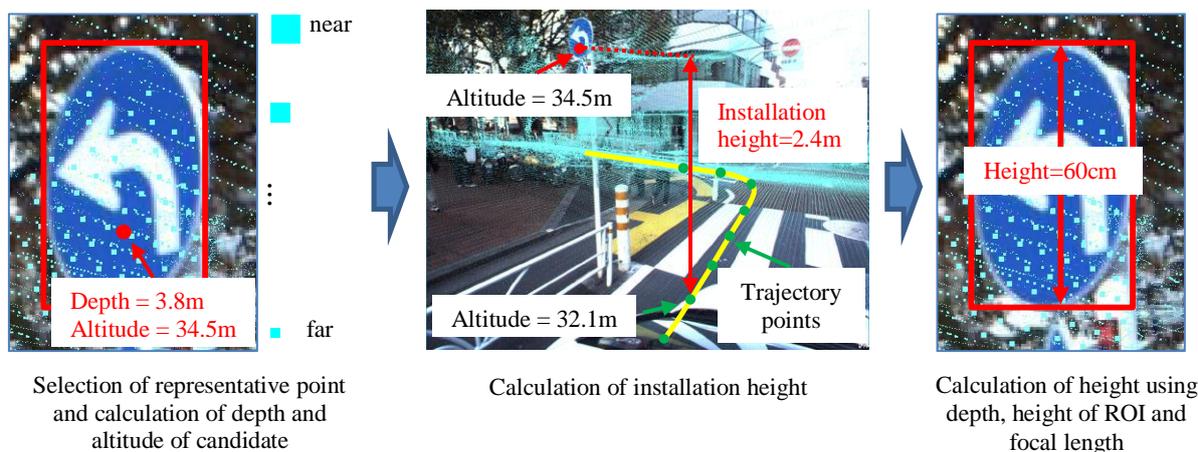


Fig.2 filtering of Detected road sign candidates using point cloud

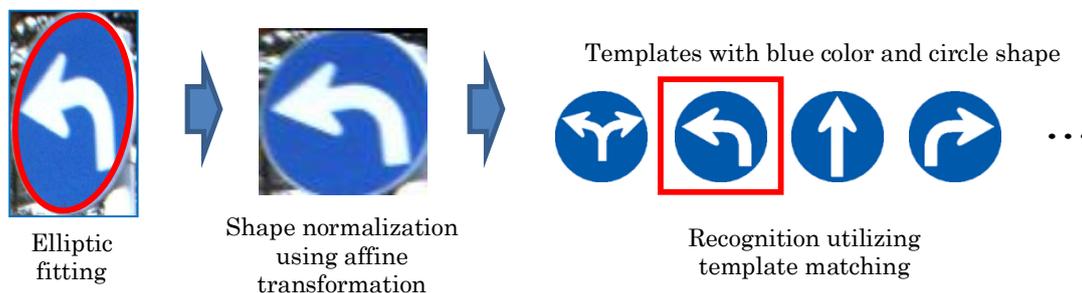


Fig.3 Classification with template matching after shape normalization

classification. But, when the maximum similarity is under 0.5, then the candidate will be filtered out.

3. RESULTS AND DISCUSSION

For the experiment, images and point cloud are acquired from cameras and laser scanners respectively mounted on a MMS (Fig.4), by traveling for more than 180km on an express way and a highway (Table 1). The image resolution is 2400×2000 and the frequency of the laser scanner is 54300Hz. Moreover, the image acquisition interval for the cameras is set as 5m on an express way and 2.5m on the highway respectively, according to the complexity of environment and the size of road signs.

The road signs that are bigger than 32×32 pixels are detected and recognized by proposed approach. Fig.5 shows the results of

each process stages, i.e. detection, filtering and classification. In this example, 571 candidates are extracted in the detection stage. The count of candidates is rather many due to the loose restriction on the color and shape for robust detection, because the true positives are first concern in our detection. Then through the filtering stage, only 267 candidates remained, that is more than half of detected candidates are filtered out efficiently. Moreover, in the relatively simple environment (e.g. express way) the filtering ratio (filtered count / candidates count) could reach to 80% or more. Finally, in the classification stage, all of the four road signs included in the image are classified correctly, meanwhile the remained false positives are eliminated thoroughly.

The quantitative results of evaluation are shown in Table 2. Three indices: the recall of detection (detected count / existing count),

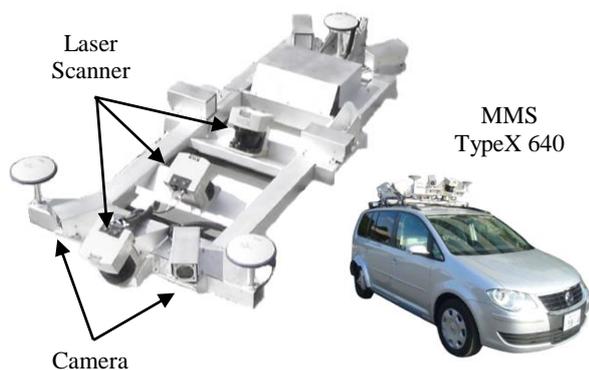


Fig.4 Mobile mapping system

	Express way	High way
Image size	2,400 × 2,000 pixels	
Laser points	54,300 points/second	
Course length	178km	5km
Shooting interval	5m	2.5m
Image count	35,588	2,000
Road sign count	373	185

Table 1. Experiment data



Fig.5 Results of detection, filtering and classification. In the images, box means region of road sign, and color of box means the color of road sign.

the recall of classification (recognized count / detected count) and the average amount of false positives per image (false positive count / image count), are adopted for the evaluation of the results. Since these indices could directly represent the amount of manual works, which are supposed to be performed after automatic road sign recognition process. The average recall of detection is 98.4%, which implies that it is possible to replace the manual extraction by automated one. The recall of detection in express way is slightly better than that in high way, due to the simple environment and large size of road signs on the express way. However, the result is opposite in the recall of classification. The reason is that on the express way there are some animal crossing signs, which have quite similar patterns, and it is hard to identify the slight difference among those patterns. The average recall of classification is 97.1%, and it has room for improvement when comparing to the state-of-the-art. The amounts of false positives per image in express way and high way are different substantially due to the complex environment of the high way. The average amount of false positives per image is 0.11, which implies that the 11% (or less than 11%) of images containing false positives have to be checked and filtered at manual works. However, the filtering task does not take much time. Therefore the amounts of false positives in this results are acceptable.

	Express way	High way	Total
Existing road signs	373	185	558
Detected road signs (recall of detection)	368 (98.7%)	181 (97.8%)	549 (98.4%)
Recognized road signs (recall of classification)	356 (96.7%)	177 (97.8%)	533 (97.1%)
Image count	35,588	2,000	37,588
Over detected road signs (false positives per image)	3559 (0.10)	420 (0.21)	3979 (0.11)

Table 2. Results of detection and recognition

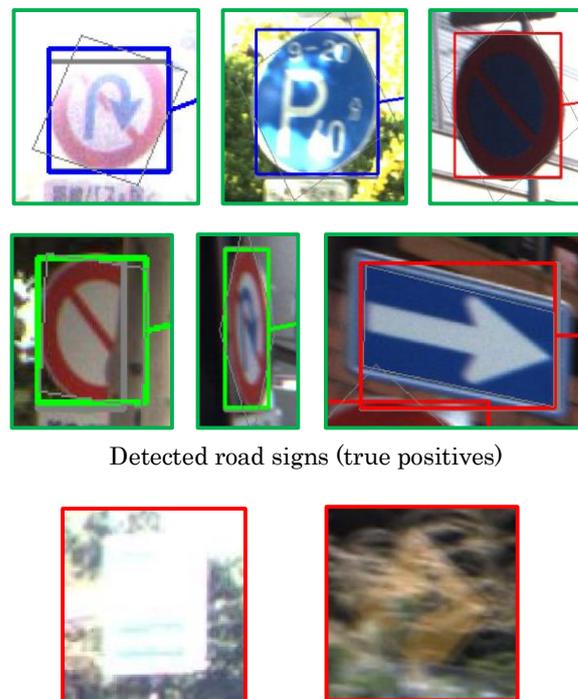


Fig.6 Examples of detected or not detected road signs

Fig.6 shows that the proposed approach is robust to various severe conditions i.e. halation, backlight, shadow, occlusion and deformation etc. On the other hand, the major reasons causing incomplete detection are extreme halation, blur and obstacles in the foreground which have the same color with road sign.

4. CONCLUSION

The objective of this research is to improve the efficiency of the work in creating the high-definition road maps. In this paper, a novel approach of utilizing both of images and point cloud acquired from MMS is proposed for road sign detection and recognition. The approach is evaluated using experimental data which is acquired by MMS traveling more than 180km on an express way and a high way in Japan. The results show that the proposed approach has performed well even under the challenging conditions such as halation, shadows, partial occlusions and large deformations etc. The average recall of sign detection is 98.4%, and it is slightly better than that of manual extraction 98%. Hence, it is possible to improve the efficiency through replacing the manual task with automatic sign recognition.

However, the amount of over detections (false positives) and false classifications (false negatives), which are needed to be eliminated or corrected manually, are still not ideal. Therefore, in the future works, the machine learning based technique such as deep learning (Ciresan et al., 2012) would be considered to be used at road sign classification stage, when there are plenty of sample data for training.

REFERENCES

Adam, A., and Ioannidis, C., 2014. Automatic road-sign detection and classification based on support vector machines and hog descriptors. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Riva del Garda, Italy, Vol. II, Part 5, pp. 1-7.

Baatz M. and Schape A., 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: *Angewandte Geographische Informations-Verarbeitung XII*, pp.12-23.

Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), pp.2-16.

Ciresan, D., Meier, U., Masci, J., and Schmidhuber, J., 2012. Multi-column deep neural network for traffic sign classification. *Neural Networks*, 32, pp.333-338.

Lopez, L. D., and Fuentes, O., 2006. Color-based road sign detection and tracking. *Image Analysis and Recognition*, 4633, pp.1138-1147.

Loy, G., and Barnes, N., 2004. Fast shape-based road sign detection for a driver assistance system. *Intelligent Robots and Systems*, 1, pp.70-75.

Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gomez-Moreno, H., and Lopez-Ferreras, F., 2007. Road-Sign Detection and Recognition Based on Support Vector Machines. *IEEE Transactions on Intelligent Transportation Systems*, 8(2), pp.264-278.

Siogkas, G. K., and Dermatas, E. S., 2006. Detection, tracking and classification of road signs in adverse conditions. In:

Proceedings of Mediterranean Electrotechnical Conference, Benalmádena, Spain, pp.537-540.

Sermanet, P., and LeCun, Y., 2011. Traffic Sign Recognition with Multi-Scale Convolutional Networks. *Neural Networks*, 31, pp.2809-2813.

Shi, Y., Shibasaki, R., and Shi, Z.C., 2008. Towards automatic road mapping by fusing vehicle-borne multi-sensor data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, China, Vol. XXXVII, Part B5, pp.867-872.

Soilan, M., Riveiro, B., Martinez-Sanchez, J., and Arias, P., 2016. Traffic sign detection in MLS acquired point clouds for geometric and image-based semantic inventory. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp.92-101.