MAPPING TRAFFIC SIGNBOARD FROM MOBILE MAPPING SYSTEMS USING DEEP LEARNING APPROACH

Pei-Cheng Chen (1), Tee-Ann Teo (2)

[1]Master Student, Dept. of Civil Engineering, National Chiao Tung University, Hsinchu, Taiwan
[2] Professor, Dept. of Civil Engineering, National Chiao Tung University, Hsinchu, Taiwan
Email: a4623217.cv03@nctu.edu.tw; tateo@mail.nctu.edu.tw

KEY WORDS: High definition map, mobile mapping system, traffic sign, deep learning.

ABSTRACT: The high definition map (HD map) requires precise traffic sign information. For example, the OpenDRIVE standard requires the attribute of traffic sign such as type, subtype, location, height, and orientation. The Mobile Mapping System (MMS) collects color images, lidar 3D point cloud, and trajectory from positioning and orientation system (POS). To fulfill the needs of traffic signboard information for HD map, the lidar point clouds may provide geometrical information such as location, height, orientation while the color images may provide semantic meaning about traffic sign. The objective of this study is to develop an automatic framework to generate traffic signboard information for HD map. Most studies utilized either color images or lidar point clouds for traffic signboard detection. As the fusion of images and 3D point clouds may obtain both geometric and semantic information, this study adopts a data fusion approach to fuse the information from color image and lidar point clouds for traffic signboard recognition. The proposed method includes three steps. The first step detects the traffic signboard using lidar intensity. The traffic signboard is covered by a highly reflective surface. Therefore, the lidar intensity for traffic signboard is generally higher than other objects. The geometrical relationship between lidar and color image is based on the exterior orientations from the trajectory. The second step utilizes the exterior orientations of the color image to estimate the location of traffic signboard in image space. The traffic signboard in image space is classified by a modified VGG network to obtain the name of signboard among the 78 classes in Taiwan’s traffic system. The last step extracts the geometrical parameters such as location, height, the orientation of traffic signboard from lidar point. The experimental results indicate that the attributes of traffic signboards can be extracted automatically.

1. INTRODUCTION

Mobile Mapping System (MMS) is an important equipment for collecting geospatial information in a mapping project. The MMS in this study is RIEGL VMX250. This system acquires RGB color images and lidar point cloud. The trajectory of MMS is from position and orientation system (POS). The lidar with 15cm accuracy is ideal geospatial information for high definition map. The standard of HD Map in Taiwan required 20cm accuracy in horizontal direction and 30cm accuracy in vertical direction (TAICS, 2018). The transformation between 3D lidar points and 2D image pixel is established by orientation parameters from direct georeferencing. Therefore, the lidar data can be converted to 2D depth image and intensity image for the detection of traffic signboard.

The previous study about deep learning traffic signboard extraction (Bruno et al., 2018) demonstrated traffic signboard detection by deep learning. Zhou et al., (2014) explained the development of a smart car or self-driving car in the future. They also detected traffic sign by vision-based camera and lidar. The two studies demonstrated the importance of automation for traffic signboard.

This study aims to extract traffic signboard for HD map automatically. The extraction of traffic signboard from MMS includes detection, recognition, and localization. The initial traffic signboard is detected from lidar intensity image. Then, recognize type of signboards using color image. This study adopts convolution neural network (CNN) to recognize type of traffic signboards. It is a standard and powerful method when a large number of training data is available. Since the 3D accuracy from color images is dependent on image scale and intersection geometric, the location of signboard is calculated from lidar data.

2. METHODOLOGY

The methodology includes four steps:(1) generation of intensity and depth images, (2) initial traffic signboard detection and recognition, (3) initial traffic signboard localization using color and depth image, (4) precise traffic signboard localization using lidar points.

2.1 Generation of intensity and depth images

In order to reduce the 3D lidar points into 2D image space, this study uses the Exterior Orientation Parameters (EOP) and Interior Orientation Parameters (IOP) of color to back project 3D lidar points (i.e., E, N, H) into 2D image spaces (i.e., S,L). Then, fill-in and resampling the pixel value into depth image and intensity image. Therefore, the lidar depth, lidar intensity, and color image are all consistence in image space (Figure 1).

(a) lidar depth image (b) lidar intensity image (c)corresponding color image

Figure 1. Illustration of the depth image, intensity image, and camera image

2.2 Initial traffic signboard detection and recognition

The traffic signboard is a high reflectivity object. As the traffic signboard usually shows higher intensity than other objects, this study selects higher intensity (e.g.,>4000 digital counts) as the candidate of signboard. Then, a morphological filter removes small object (e.g. 20 pixels x 20 pixels) and generate initial traffic signboard from intensity image. Finally, the intensity image is classified into signboard and non-signboard masks (Figure 2).

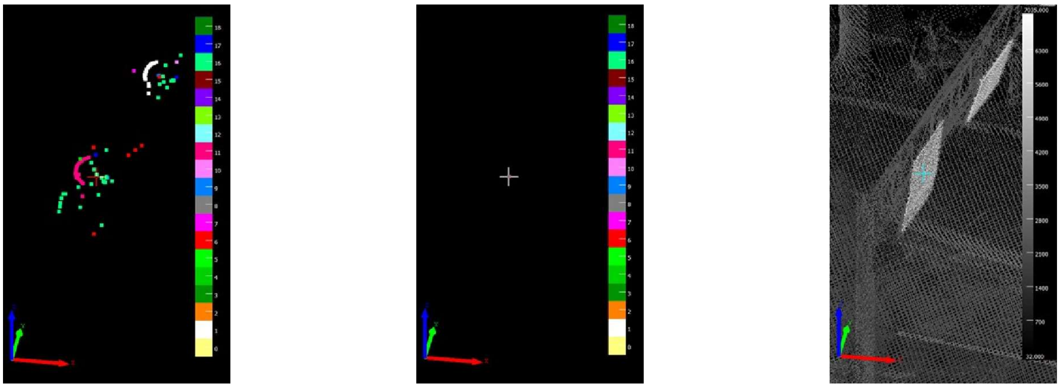
The lidar data is intensity blind and cannot recognize the attribute of traffic signboard. Hence, the color information in signboard mask is adopted to separate different signboards. The recognition stage subsets an image into many signboard image chips, then, perform CNN for signboard recognition. We perform transfer learning by VGG16 (Simonyan and Zisserman, 2014). We reduce the network parameters that the modified model input image size (i.e. 64×64) is smaller than original version (i.e. 227×227).

(a) High-intensity pixels (b)initial traffic signboard and color

Figure 2. An example of a signboard mask from intensity image

2.3 Initial traffic signboard localization using color image

After the signboard detection and recognition, every image chips belongs to each class and the depth information from the corresponding depth image is extracted to determine the 3D points in object space. An image point (i.e., *S,L*) in camera frame (i.e., *x,y,f*) is converted to object point (i.e. *E,N,H*) using scale (i.e., *Scale*) from depth image. The original traffic signboard locations (Figure 3a) are further calculated from these object points. We use centroid to represent the location of traffic signboard (Figure 3b).

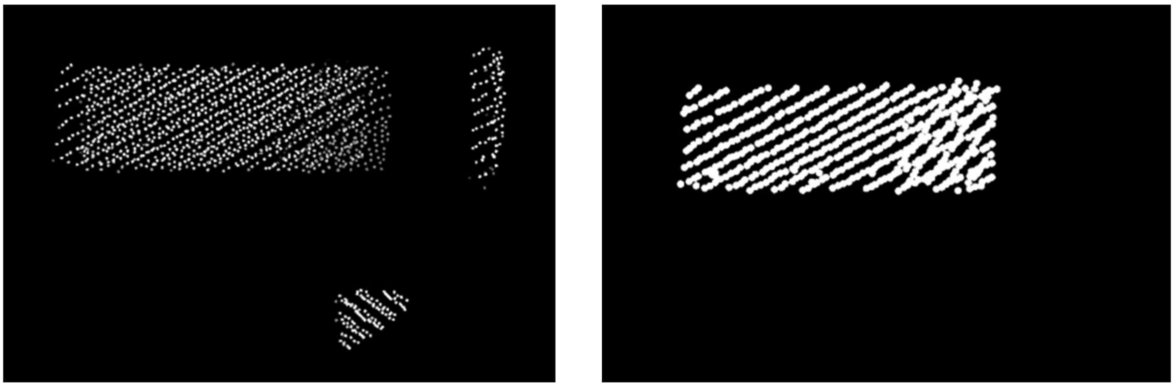


(a) Initial locations from all color images (b) Centroid of all initial locations (c) Corresponding lidar points

Figure 3. Illustration of initial locations, centroid, and lidar points

2.4 Precise traffic signboard localization using lidar

The depth error induced by interpolation error and pixel spacing may exaggerate the position error (Figure 3a). Consequently, the accuracy of the signboard location from the previous section cannot meet the requirement of HD Map. To overcome this issue, an additional single-linkage hierarchical clustering (Johnson, 1967) is used to refine the location of traffic signboard using original lidar points (i.e. precise coordinate without depth error) in object space (Figure 3c). The properties of a traffic signboard include signboard type, location, height, width, thickness, and plane orientation. The plane orientation represents the look direction of signboard for the driver. In this study, we use principal component analysis (PCA) (Pauly et al., 2002) to reduce the impact of noisily points for signboard direction (Figure 4).



(a) Mixed lidar points (b) Refined lidar points

Figure 4. Illustration of lidar points for traffic signboard

3. RESULTS

3.1 Training dataset for signboard recognition

The training dataset includes Taiwan dataset and German Traffic Sign Detection Benchmark (GTSRB, Stallkamp et al., 2012). The Taiwan dataset was manually edited from mobile mapping color images in Taipei City. There were 78 classes of traffic signboards with 43791 image chips (Figure 5).

(a)Taiwan dataset (b)GTSRB dataset

Figure 5. The training dataset for signboard recognition

3.2 Mobile mapping data

The mobile mapping system was Riegl VMX-250 (Rigel, 2012). This system includes two laser scanner and four cameras. We selected a crossroad in Taipei city (Figure 6) to evaluate the proposed scheme. The test data includes 462 images taken from 3 cameras and lidar points from two laser scanner. This intersection has many traffic signboards can be an ideal experiment area.

Figure 6. Aerial orthoimage of Sec. 3, Minquan E. Rd and Dunhua N. Rd Intersection

3.3 Accuracy of traffic signboard recognition

The correctness was evaluated by human interpretation in image space (Table 1). The traffic signboards were detected from lidar intensity and further recognized from color image using deep learning automatically. In these 462 images, 2167 objects were detected, and 1450 objects were detected correctly. The correctness was 66.91% and the right-view camera showed higher correctness than left-view camera. The accuracy of recognition was evaluated in object space (Table 2), 94.11%(=32/34) of signboards were detected, and 79.41% (=27/34) of signboards were classified correctly.

Table 1. The correctness of traffic signboard detection in image space

	Camera 1 (Back-Left)	Camera 2 (Back-right)	Camera 3 (Front-right)	Total
Number of correct object	504	417	529	1450
Number of detected object	804	600	763	2167
Correctness(%)	62.69%	69.50%	69.33%	66.91%

Table 2. The correctness of traffic signboard recognition in object space

	Number of objects
False negative	3
False positive	1
True positive	27
Total detected	32
Ground truth	34
Correctness(%)	27/34 = 79%

3.4 Accuracy of traffic signboard localization and orientation

Figures 7 and 8 show the traffic signboard location and orientation. The mean error for initial localization was 19.5cm. With lidar points, the mean error was reduced to 5.07cm and improved to 4.63cm after refinement (Figure 9). Most signboard direction can be extracted successfully from lidar points. These signboard’s orientations were parallel to road direction. However, half of them are the opposite direction

Figure 7. Traffic signboards detection

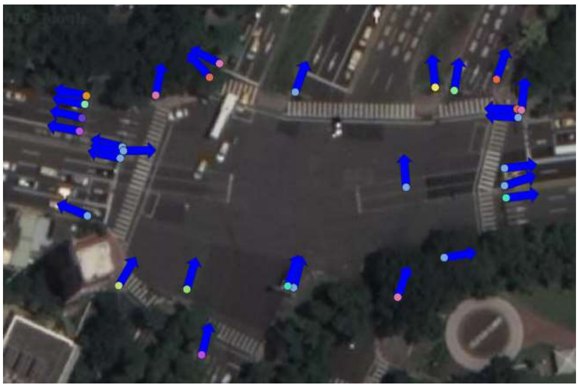


Figure 8. Traffic signboard orientation

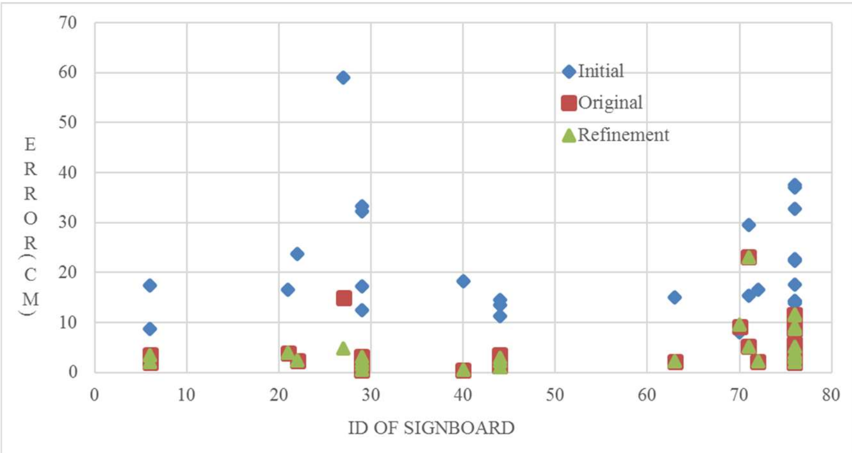


Figure 9. Improvement after using lidar points and refined lidar points

4. CONCLUSIONS AND FUTURE WORKS

The traffic signboard for HD map can be extracted by three major products from mobile mapping systems: The color images for object recognition; the lidar point clouds for object detection and localization; the direct georeferencing of positioning and orientation system for the conversion between image and object space. The localization accuracies of this study were better than 20cm in the horizontal direction and 30cm in vertical direction. However, not all traffic signboard orientation is correct when some of them in opposite direction. The problem can be solved by importing trajectory information again in order to know which image is front view or back view.

ACKNOWLEDGMENTS
This research was partially supported by the Ministry of Science and Technology of Taiwan. The author would like to thank the Department of Land Administration, Ministry of the Interior of Taiwan for providing the test datasets.

REFERENCES

1. Bruno, D. R., Sales, D. O., Amaro, J., & Osório, F. S., 2018. Analysis and fusion of 2D and 3D images applied for detection and recognition of traffic signs using a new method of features extraction in conjunction with Deep Learning. In 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
2. Johnson, S. C., 1967. Hierarchical clustering schemes. Psychometrika, 32(3), 241-254. doi.org/10.1007/BF02289588
3. OpenDRIVE, 2019. OpenDRIVE V 1.5 Format Specification, Reversion M, March 31, 2019,

Available online: http://www.opendrive.org/download.html (Accessed on September 8, 2019)

4. Pauly, M., Gross, M. and Kobbelt, L. P., 2002. Efficient simplification of point-sampled surfaces. In: VIS ’02: Proceedings of the conference on Visualization ’02, IEEE Computer Society, Boston, Massachusetts, pp. 163–170.
5. Rigel, 2012. Riegl VMX-250 Datasheet, Available online:: http://www.riegl.com/uploads/tx_pxpriegldownloads/10_DataSheet_VMX-250_20-09-2012.pdf (Accessed on September 8, 2019)
6. Stallkamp, J., Schlipsing, M., Salmen, J., & Igel, C., 2012. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. Neural networks, 32, 323-332. Available online: http://www.sciencedirect.com/science/article/pii/S0893608012000457 (Accessed on September 8, 2019)
7. TAICS, 2018. HD maps operation guidelines, Available online:: https://www.taics.org.tw/userfiles/file/20181226/20181226174628_35411.pdf (in Chinese) (Accessed on September 8, 2019)
8. Zhou, L., & Deng, Z., 2014. LIDAR and vision-based real-time traffic sign detection and recognition algorithm for intelligent vehicle. In 17th international IEEE conference on intelligent transportation systems (ITSC) (pp. 578-583). IEEE.