

AUTOMATIC TRAFFIC SIGN DETECTION AND RECOGNITION USING MOBILE LIDAR DATA WITH DIGITAL IMAGES

Haiyan Guan ¹, Yongtao Yu², Dilong Li³ and Jonathan Li ⁴

¹School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science & Technology, Nanjing, Jiangsu 210044, China

² College of Computer Engineering, Huaiyin Institute of Technology, Huaian, JS 223003, China

³ State Key laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China

⁴ Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada -

KEY WORDS: mobile LiDAR, images, traffic sign, capsule convolutional networks, high-order capsule features

ABSTRACT:

This paper presents a traffic sign detection and recognition method from mobile LiDAR data and digital images for intelligent transportation-related applications. The traffic sign detection and recognition method includes two steps: traffic sign interest regions are first extracted from mobile LiDAR data. Next, traffic signs are identified from digital images simultaneously collected from the multi-sensor mobile LiDAR systems via a convolutional capsule network model. The experimental results demonstrate that the proposed method obtains a promising, reliable, and high performance in both detecting traffic signs in 3-D point clouds and recognizing traffic signs on 2-D images.

1. INTRODUCTION

Traffic signs play an important role in road transportation systems because they provide useful and vital road information and instruction to drivers and road users (Gudigar and Chokkadi, 2014). Therefore, rapidly updating traffic signs is essential for transportation agencies to manage and monitor the status and usability of traffic signs. In addition, accurately locating and recognizing traffic signs contributes to the development of intelligent transportation systems (ITS), including unmanned driving, driver assistance and safety warning systems, and traffic sign maintenance. However, although traffic sign detection and recognition (TSDR) have been developed and employed for recent years (Salti et al., 2014; Jin et al., 2014; Liu et al., 2014), manually investigating traffic signs (by an operator checking the status of each traffic sign on videos and digital images) is still a popular way in traffic sign inventory and monitoring. The manual traffic sign inventory way is labor-intensive and time-consuming, decreasing the reliability of TSDR.

Traffic signs are defined by different colors (e.g. red, blue, green, and yellow in RGB color space) regarding their functions. Besides RGB color space, traffic sign detection works have been conducted based on Y'CBCR (Prieto et al., 2009), HSV (Maldonado et al., 2010), and CIECAM97 (Gao et al., 2006) color models to achieve a reliable detection performance under different lighting situations. Moreover, traffic signs are presented by different shapes to regulate the drivers, and thus developing a variety of shape-based methods. Furthermore, to improve traffic sign detection rate, most methods have been developed based on color and shape, such as SVM (support vector machine) based, machine learning-based, sparse representation based graph embedding (SRGE), Convolutional neural network, feature-based methods, template-

matching-based, eigen-based, supervised low-rank matrix recovery model, and decision fusion and reasoning module based methods. However, video-based and image-based TSDR systems suffer from the following limitations: 1) weather conditions (e.g., fog and rain.), affecting the visibility of traffic signs, 2) shadows, caused by other adjacent objects or different illumination levels, 3) traffic signs with bad placement or disorientation, which is relevant to the usability and viability of traffic signs, and finally affects the road safety of road users, and 4) variable color and shape information of traffic signs.

The first commercial mobile laser scanning (MLS) system was developed in 2003. Although MLS is still in the early process of applying this powerful 3D survey technology, it is being used at an increasing rate for transportation-related surveys because this technology rapidly acquire highly dense and accurate 3-D point clouds. The 3-D point clouds provide accurate geometric and localization information of the objects (Beraldin et al., 2010); whereas the color imagery provides detailed texture and content information of the objects. Therefore, by fusing imagery and 3-D point clouds, MLS systems provide a promising solution to traffic sign detection (based on 3-D point clouds) and recognition (based on imagery).

Currently, most existing methods for traffic sign detection in MLS point clouds are basically based on their prior knowledge, including position, shape, and laser reflectivity. Given the fact that traffic signs are placed close to the boundaries of the road, Chen et al. (2007) proposed a processing chain of cross section analysis, individual object segmentation, and linear structure inference to detect traffic signs. To exploit pole-like structures of traffic signs, Yokoyama et al. (2011) used a combination of Laplacian smoothing and principal component analysis (PCA), where Laplacian smoothing functioned to smooth each point cloud

segment to suppress measurement noise and point distribution bias; whereas PCA was performed on the smoothed segments to infer pole-like objects. A 3-D object matching framework was developed by Yu et al., (2015) for detecting traffic signs of varying shapes, completeness, or hidden in trees. In addition, Hough forest methods (Wang et al., 2014) and LiDAR and vision-based real-time traffic sign detection method (Zhou et al., 2014) were also developed for traffic sign detection. To present clear traffic signals, traffic signs are made by highly reflective materials. As a result, traffic signs usually exhibit high retro-reflectivity (in a form of intensity) in the MLS point clouds. Such intensity information becomes an important clue for distinguishing traffic signs from other pole-like objects (Wen et al., 2015). In addition, Chen et al. (2009) detected traffic signs by using a random sampling consensus based method. Similarly, Vu et al. (2013) developed a template-driven method to detect traffic signs with the prior knowledge of symmetric shapes and highly reflective planes perpendicular to the direction of travel.

Generally, there is a two-step procedure - traffic-sign detection using LiDAR point clouds and traffic-sign recognition using digital images. In traffic sign detection, most methods detect traffic signs from point clouds by using topology, intensity, and geometrical dimension, relations, shape, etc. The traffic sign recognition tasks are commonly employed by using machine learning or deep learning algorithms. Comparatively, deep learning methods, such as deep neural networks (DNN) (Arcos-Garcia et al., 2017), and capsule convolutional networks (Sabour et al., 2018) can automatically abstract high-level feature representations from voluminous data samples, which have become attractive in traffic sign recognition. These deep learning methods are proven to generate superior experimental results. Different from classical CNN models that take into account only the probability, capsule networks are more powerful and robust to abstract intrinsic features of objects.

2. MOBILE LASER SCANNING DATA

This research uses data collected by a RIEGL VMX-450 system. The system is composed of the following: two RIEGL VQ-450 laser scanners, four CCD cameras, a set of Applanix POS LV 520 processing systems containing two global navigation satellite system (GNSS) antennas, an initial measurement unit (IMU), and a wheel-mounted distance measurement indicator (DMI). The surveyed area is in Xiamen Island, Xiamen, China. A complete survey, at approximately 50 km/h, was conducted along Huandao Road from Xiamen University to the International Conference and Exhibition Center (ICEC). This is a typical tropical urban environment with high buildings, dense vegetation, and traffic signposts along the surveyed road. To clearly demonstrate the experimental results, we selected a small road-section dataset from the survey dataset (see Fig. 1)

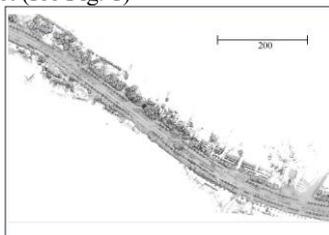


Fig. 1. MLS data.

3. TRAFFIC SIGN DETECTION AND RECOGNITION

3.1. LiDAR-based traffic sign interest region extraction

To locate traffic signs, a filter method is first used to divide mobile LiDAR data into ground and off-ground points. Regarding the motorized mirror scanning mechanism of a RIEGL VMX-450 system, we propose a curb-based filtering method to process mobile LiDAR data scan-line by scan-line. Curbs function to not only contain and direct water flow to as part of the drainage system but also separate road surfaces from sidewalks in an urban environment. We extract curb points using two criteria, namely, height difference and slope. Slopes at the border of pavement and roadway are usually larger than those of consecutive points on the roadway. Moreover, pavement points have larger elevations than road points in the neighborhood. The slope criterion detects non-road points, such as cars and curbs. Then the elevation-difference criterion detects the curbs from the non-road points. After removing road points, it is efficient to use the algorithm of progressive triangulated irregular network (TIN) densification to obtain ground points from non-road points. This TIN densification filtering algorithm is considered to be robust and steady for modelling surfaces with discontinuities such as urban areas. Normalized digital surface model (nDSM), a representation of elevated objects on a flat surface, is generated by subtracting digital terrain model (DTM) from digital surface model (DSM).

To extract pole-like object, a Euclidean clustering algorithm is used to obtain sets of clusters, the covariance matrix's three eigenvalues (λ_1 , λ_2 , and λ_3 , $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$) are decomposed to calculate three members of the eigen-based feature descriptor, $g_{\text{eigen}} = \{a_l, a_p, a_v\}$:

$$a_l = \frac{\sqrt{\lambda_1} - \sqrt{\lambda_2}}{\sqrt{\lambda_1}}, \quad a_p = \frac{\sqrt{\lambda_2} - \sqrt{\lambda_3}}{\sqrt{\lambda_1}}, \quad a_v = \frac{\sqrt{\lambda_3}}{\sqrt{\lambda_1}} \quad (1)$$

where a_l , a_p , and a_v represent linear, planar, and volumetric geometrical features, respectively.

Moreover, retro-reflectance properties is used to extract traffic signs from pole-like objects. Then, the individual regions of traffic signs are outlined to generate traffic sign interest regions.

3.2. Traffic sign recognition

All individual traffic sign interest regions are projected on images simultaneously collected by a RIEGL VMX-450 system based on image exterior and interior orientation parameters. Then, the traffic sign images are segmented and resized by a pre-defined threshold. To recognize traffic sign, deep learning is used. Recently, deep learning techniques have been attractive for their superior performance in learning hierarchical features from high-dimensional unlabelled data. By learning multi-level feature representations, deep learning models have been proved to be an effective tool for rapid object-oriented classification and recognition problems.

To recognize traffic signs from the segmented image patches, we construct a convolutional capsule network. Capsule network, first proposed in Sabour et al., (2017) for classification tasks, is composed of entity-oriented vectorial capsules, which differs from conventional CNNs that employ scalar neurons to encode the

probabilities of the existence of specific features. A capsule can be viewed as a vectorial combination of a set of neurons (Sabour et al., 2017). For a capsule, its instantiation parameters represent a specific entity type and its length represents the probability of the existence of that entity. Capsule networks have been demonstrated to be powerful and robust in various classification tasks. Thus, to obtain promising traffic sign recognition performance, we extend the original capsule network (containing two conventional convolutional layer, a primary capsule layer, three convolutional capsule layers, a capsule max-pooling layer, and three fully connected capsule layers) to construct a multi-layer convolutional capsule network.

The two conventional convolutional layer uses 3×3 convolution operations to extract low-level features from the input image patches. These features are further encoded into high-order capsules to represent different levels of entities. The two conventional convolutional layers adopt the widely used ReLU as the activation function to nonlinearly transform the outputs.

The primary capsule layer converts the low-level scalar feature representations in the convolutional layer into high-order vectorial capsule representations. This conversion is based on a conventional convolution operation sliding on the convolutional layer with a 3×3 kernel size.

The three convolutional capsule layers extract high-order capsule features from low-order capsules by performing local convolution operations on a group of capsules and representing their features using a new capsule. For the capsules in the convolutional capsule layers, the total input to a capsule j is a weighted sum over all predictions from the capsules within the convolution kernel in the layer below:

$$C_j = \sum_i a_{ij} \cdot \overline{U_{ji}} \quad (1)$$

where C_j is the total input to capsule j ; a_{ij} is the coupling coefficient, indicating the degree of contribution that capsule i in the layer below activates capsule j ; $\overline{U_{ji}}$ is the prediction from capsule i to capsule j and it is defined as follows:

$$\overline{U_{ji}} = \mathbf{W}_{ij} \cdot \mathbf{U}_i \quad (2)$$

where \mathbf{U}_i is the output of capsule i . \mathbf{W}_{ij} is the transformation matrix on the edge connecting capsules i and j . Specifically, the coupling coefficients between capsule i and all its connected capsules in the layer above sum to 1 and are determined by a dynamic routing process [18]. The dynamic routing process considers both the length of a capsule (i.e., the probability of the existence of an entity) and its instantiation parameters (i.e., the orientation of the entity) to activate another capsule. For the convolutional capsule layers, the non-linear ‘‘squashing’’ function [18] is adopted as the activation function, by which the capsules with short vectors result in low probability estimations and capsules with long vectors result in high probability estimations, whereas their orientations remain unchanged. The non-linear squashing function is defined as follows:

$$U_j = \frac{\|C_j\|^2}{1 + \|C_j\|^2} \cdot \frac{C_j}{\|C_j\|} \quad (3)$$

By such a conversion, the capsules with short lengths are narrowed down to a length close to zero and the capsules with long lengths are shrunk to a length close to one.

The three fully connected capsule layers consider all the capsules in the layer below to construct a high-order entity abstraction from a global perspective. The first fully connected capsule layer is obtained using a set of global capsule convolution kernels performing on the capsule max-pooling layer. The last fully connected capsule layer is a softmax layer for classification purposes. We use the capsule length in the softmax layer to represent the probability of a traffic sign image patch being an instance of a specific category (forbidden or warning). The category label of a traffic sign image patch is defined as follows:

$$L^* = \arg \max_k \|U_k\| \quad (4)$$

where U_k is the output of a capsule in the softmax layer.

To effectively train the convolutional capsule network towards classification tasks, the margin loss (Sabour et al., 2017) is used as the objective function to direct the error backpropagation process.

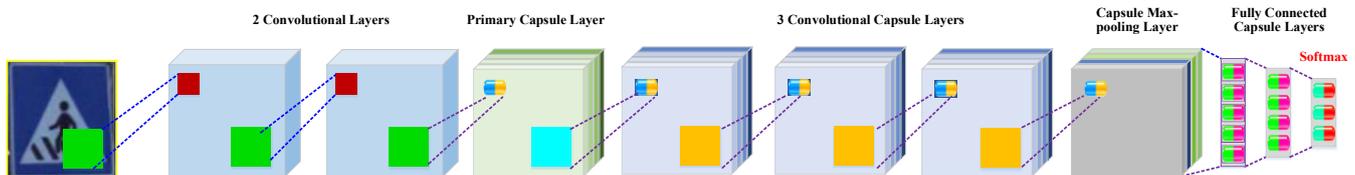


Fig. 2 Illustration of the proposed model.

4. RESULTS AND DISCUSSION

(1) Traffic-sign Detection

Fig. 3 shows the results of traffic-sign detection (red points). Visual inspection shows that the detected traffic-sign results were satisfactory and hung on the poles. According to the detected traffic signs, their poles are then determined from clusters in the neighborhoods. The two data sets in this study contain a total of

1,268 traffic signs. We detected 1,162 traffic signs, including 1,101 correctly-detected traffic signs and 61 non-traffic signs. The detection accuracy is 86.8%. Some incompletely scanned traffic signs, caused by occlusions, were also undetected because of insufficient salient features. Although some advertising boards attached to light poles were misclassified as traffic signs due to strong reflectance, the majority of traffic signs were correctly detected.

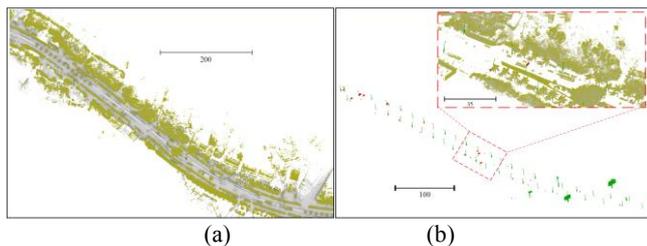


Fig. 3 (a) Filtered results and (b) traffic-sign detection results

After the traffic-sign interest regions were extracted from the mobile LiDAR points, to obtain their images, we projected them onto the digital images (see Fig. 3). Then, we performed the proposed GB-DBM classifier to classify the resized traffic-sign images into specific categories.

(3) Traffic-sign Recognition Performance

This test set contained 1,162 traffic sign image patches covering 35 different categories of traffic signs and the background. At the test stage, the test images were fed into the convolutional capsule network to recognize traffic signs. For the output of the softmax layer of the convolutional capsule network, the capsule with the longest length corresponded to the category of an image patch. For an image patch labeled as a traffic sign, the length of the capsule encoded the probability of the image patch belonging to an instance of that traffic sign type. The proposed framework was capable of processing eighteen traffic sign patches per second. To quantitatively evaluate the traffic sign recognition accuracy, we used the recognition rate as the evaluation metric, which is defined as the proportion of correctly classified traffic signs. On average, our proposed framework achieved a traffic sign recognition rate of 0.965 on the test set.

The misclassification of the traffic signs might be caused by: 1) distorted traffic-sign images due to a very large viewpoint; 2) poor traffic-sign image quality due to extremely strong or poor illumination; and 3) incomplete traffic-sign images due to serious occlusion.

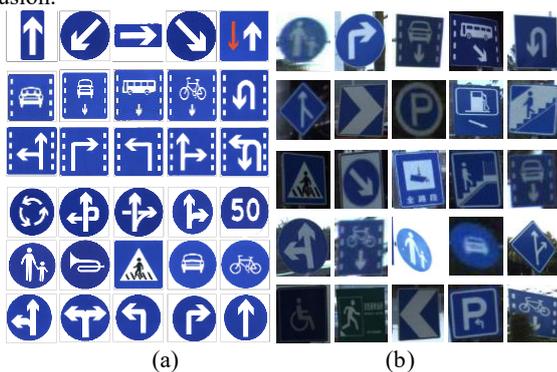


Fig. 4 (a) standard traffic signs downloaded from the Ministry of Transport of the People's Republic of China, and (b) the segmented traffic sign images from LiDAR data.

ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 41671454 and Grant 419. Contact corresponding author via guanhy.nj@nuist.edu.cn for more information about the software tools used in this study.

REFERENCES

- Arcos-García, Á. Soilán, M. Álvarez-García, J. A. and Riveiro, B., 2017. Exploiting synergies of mobile mapping sensors and deep learning for traffic sign recognition system., *Expert Syst. Appl.* 89, 286–295.
- Beraldin, J., Blais, F., Lohr, U., 2010. Laser scanning technology. In: Vosselman, G. and Mass, H. (eds.), *Airborne and Terrestrial Laser Scanning*, pp. 1-42, 2010.
- Chen, Y. Zhao, H. and Shibasaki, R., 2007. A mobile system combining laser scanners and cameras for urban spatial objects extraction, in *Proc. IEEE Conf. Mach. Learn. Cybern.*, Pokfulam, Hong Kong, 2007, vol. 3, pp. 1729–1733, 2007.
- Chen, X. Kohlmeyer, B. Stroila, M. Alwar, N. Wang, R. and Bach, J., 2009. Next generation map making: Geo-referenced ground-level LiDAR point clouds for automatic retro-reflective road feature extraction, in *Proc. ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, Seattle, WA, USA, 2009, pp. 488–491, 2009.
- Gao, X. W., Podladchikova, L., Shaposhnikov, D., Hong, K., Shevtsova, N., 2006. Recognition of traffic signs based on their colour and shape features extracted using human vision models, *J. Vis. Commun. Image Represen.*, 17(4), 675-685.
- Gudigar, A., Chokkadi, S., 2014. A review on automatic detection and recognition of traffic sign,” *Multimedia Tools and Applications*, 71,1363-1380.
- Jin, J., Fu, K., Zhang, C., 2014. Traffic sign recognition with hinge loss trained convolutional neural networks, *IEEE Transactions on Intelligent Transportation Systems*, 15(5), 1991-1998.
- Liu, C., Chang, F., Chen, Z., 2014, Rapid multiclass traffic sign detection in high-resolution images, *IEEE Transactions on Intelligent Transportation Systems*, 15(6), 2394-2404.
- Maldonado Bascón, S., Acevedo Rodríguez, J., Lafuente Arroyo, S., Fernández Caballero, A. López-Ferreras, F., 2010. An optimization on pictogram identification for the road-sign recognition task using SVMs, *Comput. Vis. Image Understand.*, 114(3), 373-383.
- Prieto, M. S., Allen, A. R., 2009. Using self-organizing maps in the detection and recognition of road signs, *Image Vis. Comput.*, 27(6), 673-683, 2009.
- Sabour, S. Frosst, N. and Hinton, G.E., 2017. Dynamic routing between capsules,” in *Proc. Conf. Neural Inform. Process. Syst.*, Long Beach, USA, 2017, 1-11.

Salti, S., Petrlli, A., Tombari, F., Fioraio, N., Stefano, L., 2014. Traffic sign detection via interest region extraction, *Pattern Recognition*, 48, 1039-1049.

Vu, A., Yang, Q., Farrell, J., and Barth. M., 2013. Traffic sign detection, state estimation, and identification using onboard sensors,” in *Proc. Int. IEEE Annu. Conf. Intell. Transp. Syst.*, The Hague, The Netherlands, 2013, pp. 875–880, 2013.

Wang, H. Wang, C. Luo, H. Li, P. Cheng, M. Wen, C. , and Li, J., 2014. Object detection in terrestrial laser scanning point clouds based on Hough forest, *IEEE Geosci. Remote Sens. Lett.* 11(10), 807–1811.

Wen, C. Li, J. Luo, H. Yu, Y. Cai, Z. Wang, H. and Wang C., 2015. Spatial-related traffic sign inspection for inventory purposes using mobile laser scanning data,” *IEEE Trans. Intell. Transp. Syst.*.

Yokoyama, H., Date, H., Kanai, S., Takeda, H., 2011. Pole-like objects recognition from mobile laser scanning data using smoothing and principal component analysis, *ISPRS Archives*, 38-5/W12, 1–6, 2011.

Yu, Y., Li, J., Guan, H., Wang, C., 2015. Automated extraction of urban road facilities using mobile laser scanning data, *IEEE Trans. Intell. Transp. Syst.* 16(4), 2167–2181, 2015.

Zhou, L. and Deng. Z., 2014. LiDAR and vision-based real-time traffic sign detection and recognition algorithm for intelligent vehicle, in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Qingdao, China, 2014, pp. 578–583, 2014.