Application of Artificial Intelligence in Road Infrastructure Management

Dejan Vasić and Marina Davidović, Republic of Serbia

Key words: education, geoinformation, implementation of plans, laser scanning

SUMMARY

In order to better manage the environment, built objects and their possible deformations, the collection of three-dimensional spatial data is becoming increasingly important. In the last few years, the method of mobile laser scanning has become one of the most important methods of collecting spatial data on the surface of the terrain, as well as structures and objects on it. The data collected in this way contains a wealth of information on buildings, vegetation, streets and other urban infrastructure facilities in digital format. Mobile laser scanning systems, in addition to the laser device, often have high-resolution cameras. The result of data collection is a three-dimensional point cloud, which represents laser reflections from elements in space, as well as photographs, which have proven to be extremely useful in coloring point cloud, detecting road defects, extracting spatial entities, creating different maps and much more. It is important to emphasize here that the higher the quality of the cameras integrated into the system, the wider is the range of possibilities that photos provide.

Nowadays, there is a growing emphasis on automatic detection of objects of interest from the mentioned images. That is due to demand for automatic vehicle identification needed to control traffic, state borders, calculate payment for parking, search for stolen cars or unpaid fees, as well as for reliable identification under different lighting conditions. In addition, these methods prove to be very useful in automating the process of creating maps, ie. inventory of road infrastructure. The last decade has been characterized by an expansion of the field of artificial intelligence and machine learning. Deep learning approaches using objects detected by convolutional neural networks have significantly improved the accuracy of detection compared to traditional methods.

This paper provides an overview of existing methods for detecting road infrastructure facilities, from data obtained by mobile laser scanning. The emphasis is on recognizing objects from images, their classification and localization. At the end of the paper are given discussion and concluding remarks that could serve future development.

Application of Artificial Intelligence in Road Infrastructure Management

Dejan Vasić and Marina Davidović, Republic of Serbia

INTRODUCTION

The necessity for object detection has significantly increased. The reasons for this include a growing demand for automatic vehicle identification required for traffic control, border control, access control, calculation of parking time and payment, and searching for stolen cars or unpaid fees, along with the requirement for reliable identification of other spatial entities, such as street lights, traffic signs, traffic lights, fences, manholes, gullies and many others [1].

During recent years, object detection is addressed as one of the major problems in computer vision. This task refers to the determination of the presence or absence of specific features in image data. Once features are detected, an object can be further classified as belonging to one of a pre-defined set of classes. According to [2], object detection and classification are fundamental building blocks of artificial intelligence. Task of recognizing a specific object in an image is one step further and belongs to very challenging topics as well.

This study focuses on automatic object detection from data obtained by using modern surveying technologies. Objects of interest are all road infrastructure elements: vehicles, traffic signs, traffic lights, lamp posts, road markings etc. The aim of the paper is to present current state of field in detection of these objects by using artificial intelligence.

PREVIOUS WORKING METHODOLOGY

In this chapter is presented one working methodology for the extraction of road infrastructure elements from point cloud, prior usage of artificial intelligence.

For this purposes, Micro Station software package is used.

The process of extraction of characteristic elements contained of the following phases:

- preparation of data for extraction;
- extraction of elements in 2D and 3D;
- quality control.

The first phase is introducing with the data formats and the requested level of detail with which extraction should be carried out. Also, it is necessary to create layers regarding to the entities that are the subject of extraction, and in order to speed the work and reduce the possibility of error, layers should be categorized in groups.

The second phase implies extraction of the required elements. The structural lines are the edge and middle lines of the carriageway, the top and bottom edge of the curbs, the edges defining the channels, etc. After that, the extraction of other elements is carried out: the white lines on the road, the pedestrian crossings, cycle paths, objects, poles, traffic signs, trees, fences, borders of different cultures, etc.

Individual elements are assigned only 2D coordinates when drawing, where there are several ways to subsequently join height data. One of them is to directly select elements of interest and set the height. Also, it is possible to create a digital terrain model of points belonging to the ground class. After that, entities that do not have a given height are "dropped" on the created model. Drawing of certain entities is done in 3D, ie, beside the X, Y coordinates, they are also assigned the actual height. Drawing is done in profiles, so it takes more time than 2D extraction, but height is obtained.
In the final stage, the quality of the extraction of the road infrastructure structural elements is controlled, at first by the visual inspection, and then by the verification of the elements whether they are in the appropriate layers, type. Quality control also implies extraction of spatial entities that have been missed out or not digitized appropriately. Figure 1 shows final result, obtained by following these steps. There are two parts of the highway - with and without point cloud [3].

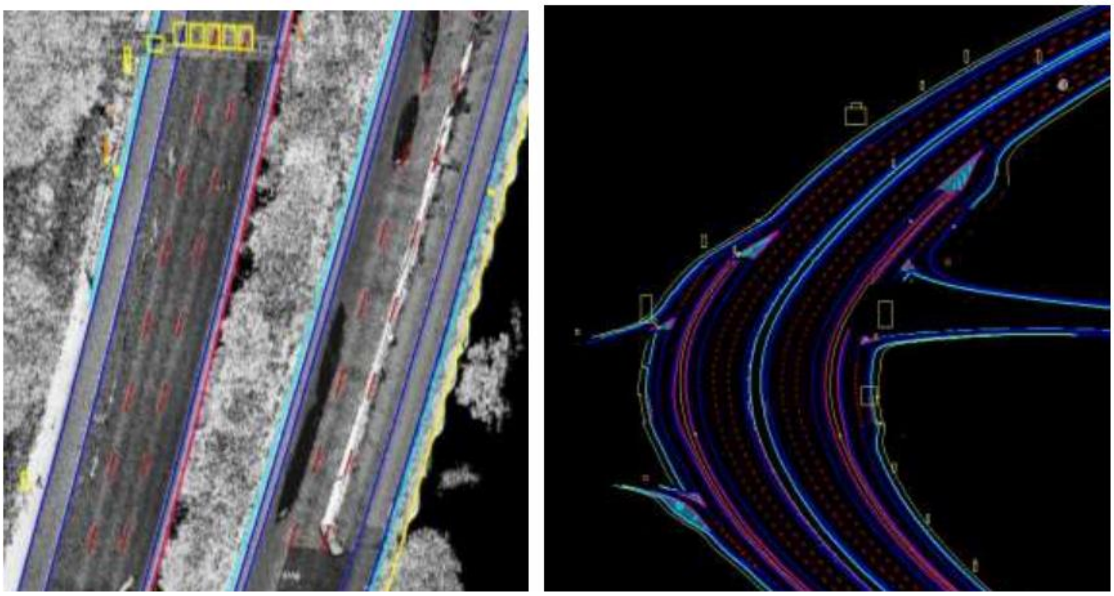

Fig 1. Example of extraction of the road infrastructure – high way

ARTIFICIAL INTELLIGENCE IN ROAD INFRASTRUCTURE OBJECTS DETECTION

Multi-category object detection is an important task for various applications. Previous chapter showed one current methodology for extraction of these objects, stage by stage. It can be concluded that, even this is faster and with more details and better accuracy than traditional methods, it is necessary to work on its improvement and progress. Those processes are still manual, or semi-automatic, at best.

In recent years, deep learning approaches using features extracted by convolutional neural networks (CNN) significantly improved the detection accuracy on detection benchmark datasets compared to traditional approaches based on hand-crafted features as used for object detection images [4].

This chapter focuses on detection of road infrastructure elements from images, with usage of CNN based architectures. A latest research on vehicles detection, road markings detection and pole-like objects detection (such as lamp posts, traffic signs) is presented.
In paper [4] is proposed a deep neural network derived from the Faster R-CNN approach for multi-category object detection in aerial images. It is shown how the detection accuracy can be improved by replacing the network architecture by an architecture especially designed for handling small object sizes. Finally, comparing suggested network to traditional baseline approaches and deep learning based approaches on the publicly available DLR 3K Munich Vehicle Aerial Image Dataset that comprises multiple object classes such as car, van, truck, bus and camper is performed. Figures 2 and 3 present some positive and negative examples of conducted detection.

Fig 2. Detection examples for category car (red boxes) and category truck (dark blue boxes) and corresponding ground truth for category car (green boxes) and category truck (light blue)

Fig 3. False positive examples for category truck (dark blue boxes) and false negative examples for category truck (light blue boxes)

Another research deals with vehicle detection from images. Namely, after data acquisition and image mosaic creation, authors in [5] conduct detection of objects of interest. The study area (Split, Croatia) had 125 aerial images acquired with UAV. The used system consisted of a ground-based computer, where all processing is done, and an UAV unit, which is used just for taking images. The system was tested with aerial images taken using DJI Phantom 3 Professional, with 4000×3000 pixels in size and JPG compressed.
In the last phase, image mosaic is divided into smaller pieces which were used as inputs to the CNN for detection. For car detection and localization, they used pre-trained Faster R-CNN model. The obtained results are promising and can be used for mapping applications. However, some minor deformations were detected in the study area. Those include facade

visibility, moving objects and still objects. The accuracy measures of the producers and users is considered, where TP are the true positives (i.e., the number of cars correctly identified), FP are the false positives (i.e., the number of cars incorrectly identified), and N is the real number of cars present in the image. Figure 4 and Table 2 present detection results.

Fig 4. Detection and localization results on test images

Table 1. Detection results

Image	Car present	TP	FP	Successfully rate
Image 1	61	49	3	80.32%
Image 2	27	20	2	74.07.%

Authors of [2] adapted and tested a CNN-based software called YOLO, for efficient and automated recognition of particular objects. Validation of the CNN was carried out by testing the classification accuracy on a class of objects labeled as “airplanes”. This class was created by downloading satellite images of airplanes grounded on civil and military airfields across the globe from Google Maps (Figure 5a). These images consisted of a variety of airplane types and a wide range of image scales, resolutions, and compositions. An airplane object category was created using these images for training the network. There were a total of 152 images containing 459 airplane objects in this training dataset.

For testing CNN recognition accuracy, a new dataset of 267 images containing a total of 540 airplanes was used (Figure 5b). Results showed (Table 2) that the CNN was able to recognize “airplane” objects in the data set with 97.5% of accuracy, while only 16 instances were incorrectly categorized.

Figure 5. (a) Training set of object class “Airplane”; (b) Testing set of object class “Airplane”

Table 2. Confusion matrix for the object class “Airplane”

Classification	Class	Detected	
		Airplane	Not Airplane
Actual	Airplane	526	14
	Not Airplane	2	NA

Automatic road identification facilitates a great number of applications such as vehicle navigation [6], urban planning [7], geographical information system upgrading [8], and so on.
Recently, some pioneering work that explores deep learning techniques for road identification has been conducted. For instance, inspired by the deep residual network, Zhang et al. [9] extended U-Net by introducing the short-cut connections between the CNN layers. The model was named as ResUNet and applied to road detection. Cheng et al. [10] developed a cascaded deep CNN approach to road identification. Their method was based on a two-stage deep learning approach, where the first stage dealt with the road detection task, and the second stage focused on centerline extraction by utilizing the domain knowledge learned from the first stage.
Here is elaborated the paper [11], where a novel deep learning model, recurrent CNN U-Net (RCNN-UNet), is proposed. To demonstrate the effectiveness of the proposed models, an extensive experiments based on two publicly available road identification data sets are conducted. Seven well-known methods were adopted as the baselines, for comparative evaluation. Those are: FCN, SegNet, U-Net, DenseUNet, ResUNet, Cascaded CNN and RoadCNN1.
The benchmark data set consists of 224 images [9] which are all collected from Google Earth. Although the sizes of the 224 images vary, the resolution of each image is at least 600×600 pixels. Each pixel denotes 1.2 m. The data set covers different geographical regions including urban, suburban and rural areas. Most of the images contain complex backgrounds and occlusions caused by trees or buildings. This makes the road identification task extremely challenging.

The four metrics are used to evaluate the performance of the centerline extraction task. These metrics include completeness (COM), correctness (COR), quality (Q), and F1 score. The COM measures the portion of matched areas with respect to the ground truth reference map. The COR denotes the percentage of matched road centerline areas from among the centerline areas detected by a computational method under our evaluation. The Q is a combined metric that takes into account both COM and COR. The F1 score is a harmonic average between COM and COR.
Table 3 summarizes the results evaluated by above explained metrics. It can be seen from the table that RCNN-UNet2 and RCNN-UNet3 perform the best, which is consistent with the visual judgment in the qualitative evaluation. Among all the baseline methods, cascaded CNN is the best one. RCNN-UNet2 and RCNN-UNet3 outperform cascaded CNN by 1.89%, 1.44%, 0.023% and 1.95% in terms of COM, COR, Q, and F1.

Table 3. Quantitative comparative evaluation among different methods for road centerline extraction [11]

Methods Name	COM	COP	Q	F1
FCN	0.7427±0.0390	0.9098±0.0407	0.8516±0.0658	0.8178±0.0392
U-Net	09534±0.0050	0.9285±0.0038	0.9335±0.0034	0.9656±0.0014
SegNet	09641±0.0071	0.9671±0.0165	0.9560±0.0152	0.9408±0.0079
DenseUnet	0.9230±0.0094	0.9686±0.0460	0.8865±0.0522	0.9453±0.0190
ResUNet	0.9262±0.0142	0.9641±0.0211	0.9005±0.0311	0.9448±0.0167
Cascaded CNN	0.9549±0.0030	0.9790±0.0042	0.9362±0.0067	0.9668±0.0035
RoadCNN	0.9687±0.0650	0.9561±0.1219	0.9141±0.1150	0.9624±0.1103
RCNN-UNet1	0.9738±0.0017	0.9800±0.0067	0.9583±0.0044	0.9769±0.0048
RCNN-UNet2	0.9871±0.0034	0.9856±0.0076	0.9732±0.0070	0.9863±0.0055
RCNN-UNet3	0.9794±0.0095	0.9959±0.0012	0.9744±0.0092	0.9876±0.0047

Figure 6 visualizes the centerlines identified by different methods.

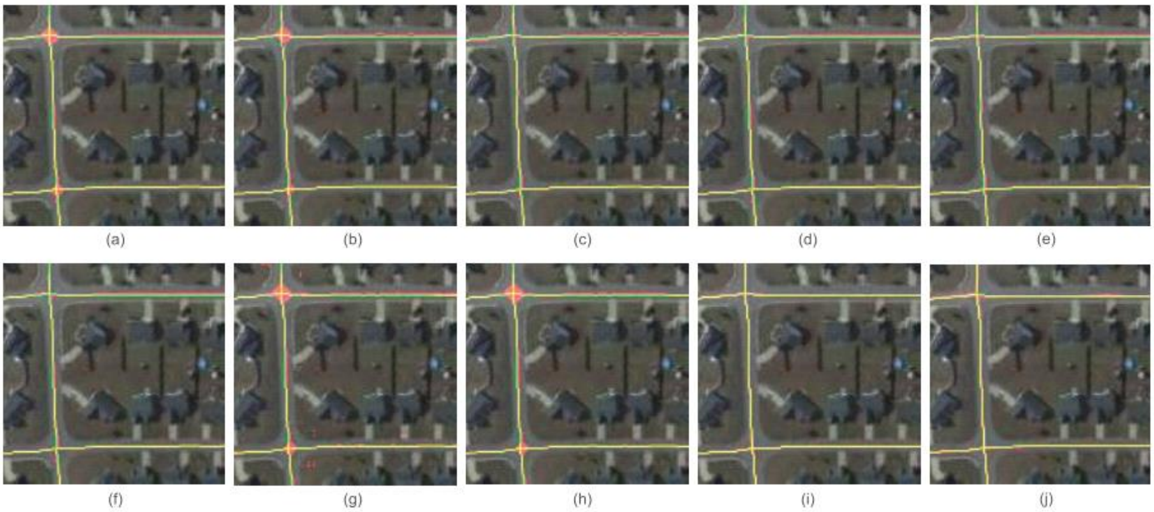

Fig 6. Qualitative comparison of road centerline extraction results. Yellow color: true positive parts. Red color: false positive parts. Green color: false negative parts.

Beside different vehicle types and road centerlines, another interesting road infrastructure element is road markings. Automated detecting road markings has become an increasing necessity for transportation-related activities, including traffic monitoring, automatic vehicle driving, and autonomous navigation [12]. There are a lot of challenges in road marking segmentation. For example: shadows caused by buildings and trees, partial occlusion caused by other objects such as vehicles, severely destroyed road markings, false road markings caused by illumination, imaging tilt angles, and other road marking materials.

A variety of deep learning networks (e.g. CNN) have drawn the increasing attention of researchers to effectively and highly-accurately detect, extract, and classify road networks and objects above or on road surfaces in complex road scenes, due to their powerful high-order feature representation, characterization, and robustness abilities [13]. However, most CNNs usually fail to extract heterogeneous object regions, and thus generating rough segmentation boundaries. Moreover, CNNs suffer from the issues of representation power and computational efficiency.

The paper [14] proposes a new attentive capsule feature pyramid network (ACapsFPN) that accurately and precisely extracts road markings from UAV images. The proposed method characterizes high-order entity features by leveraging vectorial capsule neurons. The ACapsFPN was trained and performed on a cloud computing platform equipped with ten 16-GB GPUs, a 16-core CPU, and a 128-GB memory. It was trained for 1500 epochs, each of which contained two images per GPU. The model is trained with 0.001 learning rate for the first 1200 epochs and 0.0001 learning rate for the rest 300 epochs. To evaluate the road marking extraction performance of method ACapsFPN, authors applied it to the RMS2020 dataset. The proposed method obtained a precision of 0.7366 and a recall of 0.7513.

Figure 7 illustrates a small group of representative road marking extraction results generated by the proposed architecture. The extracted road markings were colored in red. Although the great variations of the road markings in spatial sizes, intensity appearances, geometric topologies, and complicated scenarios, the ACapsFPN differentiated well the road markings from the surrounding environments with a fairly small proportion of false alarms and missing detections.

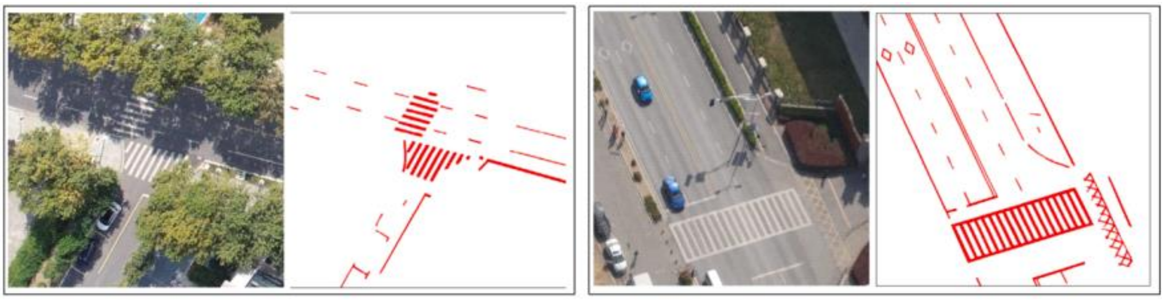

Fig 7. A close view of the road marking extraction results obtained by the ACapsFPN [13]

Even the successfully rate is satisfying, the road markings occluded by objects, such as vehicles and trees, were not continuously delineated due to the fact that road making data did not appear on the images. Additionally, there were some misclassified road markings because of the high spectral similarity of road markings and linear objects, such as the arms of light poles.

Automatic detection of like-pole objects such as traffic signs, traffic lights and lamp posts are topic for many scientific articles. Some studies have been carried out using side-view imagery for electric utilities detection and survey. For example, paper [15] detected utility poles using a template design from video surveillance on car. Song and Li [16] developed a sequential local-to-global algorithm to detect power lines from optical images and tested on 160 pictures taken from the ground with 91.95% and 91.33% true positive rates for detecting straight lines and curved lines respectively. In paper [17], a five-stage detection algorithm (including segmentation, block-oriented quadrilateral extraction, quadrilateral shape determination, orientation based spatial clustering of near-trapeziums, and context-based detection) was developed to detect utility poles in pure side-view images; 70% of poles from 212 frames ground truth were detected.

However, the previously used methods are complicated because they involved utilizing a variety of models and algorithms, such as feature segmentation, extractions, filters, detection, and template match, among others. That is why research goes further and result is usage of deep learning (DL) algorithms. DL has shown its powerful ability in computer vision, natural language processing, and many other fields. However, there are very few published studies on using DL to map or inspect power line components. Here is elaborated study [18].

In this study [18], authors proposed using a DL-based method for automatically mapping roadside utility poles with cross arms from Google Street View (GSV) images. The method combines the state-of-the-art DL object detection algorithm and a modified brute-force-based line-of-bearing measurement method. This method estimates the locations of detected utility poles from GSV.

The town of Mansfield (CT, USA) was selected as the study area. The GSV images were acquired between 2011 and 2016. The validation data included 1039 poles, which were located within a 20 m buffer zone around selected roads. The authors created 3500 ground-truth data points by manually labelling utility poles in GSV images. The experiment is

conducted on a customized server, which is equipped with an Intel i5 CPU, 16 GB RAM, a GeForce GTX 970 graphic card and a GeForce GTX 1080ti graphic card.
Figure 8 presents examples of accuracy assessment of utility poles detection with intersection-over-union (IoU) parameter. IoU defines the extent of overlap of two boxes.

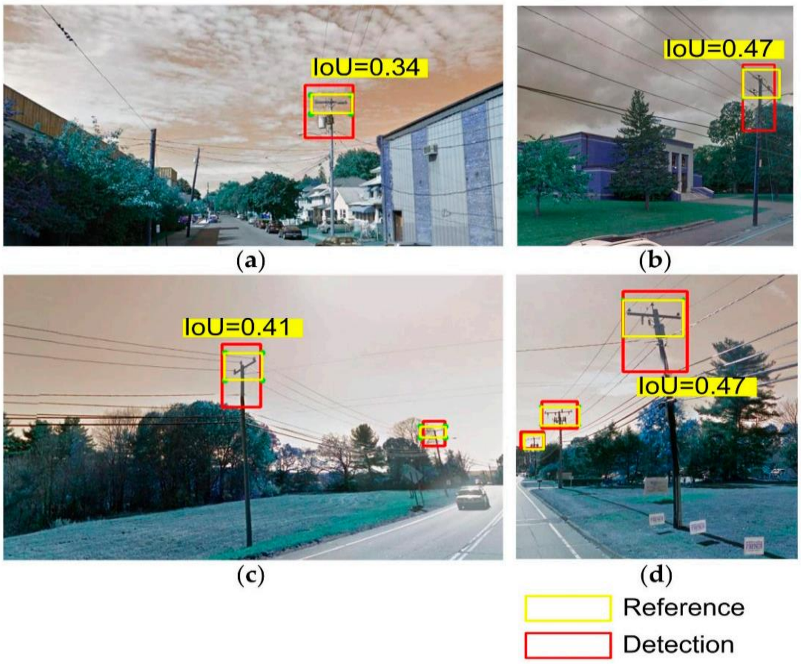

Figure 8. Positive detected utility poles [18]

Examples of errors in estimating locations of utility poles from GSV, caused by the difference between ground locations (black solid lines) and their corresponding orthographic projected location (green dash lines) is presented in Figure 9.

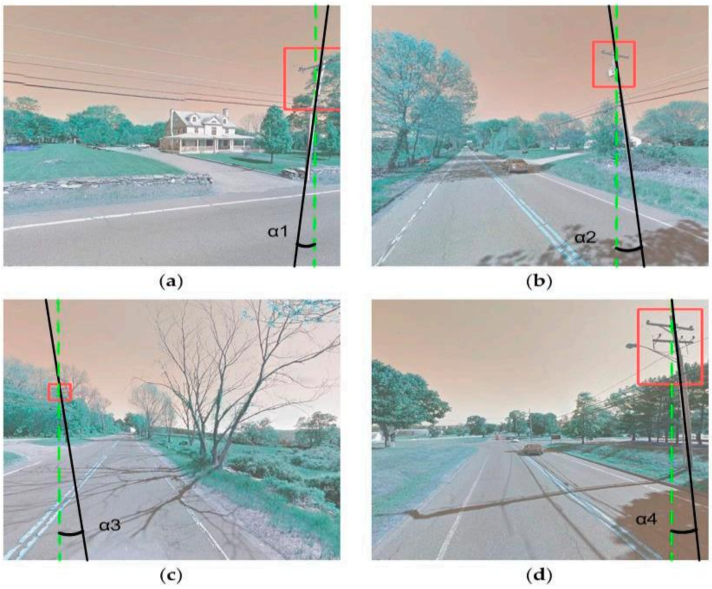

Figure 9. Wrongly positioned detected utility poles [18]

Even though this study presents a great potential of using DL to map utility poles, there are some limitations worth a mention. For example, a large amount of training dataset is needed to achieve an acceptable accuracy with this method. Therefore, a comprehensive study about the minimum requirement of the training dataset for using DL to map pole-like or other geographic objects is needed.

CONCLUSION

Despite the wide range of possibilities that mobile laser scanning technology provides, it is still working on its further improvement. One way to improve this method is to integrate high-resolution cameras into an existing system. In this way, beside point cloud as main product, there will be obtained images taken from different angles, depending on the position and number of cameras mounted on the system. Images facilitate visual inspection of spatial objects and provide a simpler interpretation of scan results. They can be extremely useful in coloring point cloud, detecting road defects, extracting spatial entities, creating different maps and much more. It is important to emphasize here that the higher the quality of the cameras integrated into the system, the wider is the range of possibilities that photos provide.

Due to the large amount of recorded data that requires post-processing, the focus has shifted from fieldwork to the office work. Accordingly, the trend of improving post-processing data is on the rise, which reveals new opportunities for products obtained by MMS technology. This further requires networking of different scientific disciplines in order to exploit the full potential of this technology.

This paper presents the use of artificial intelligence in the detection and extraction of road infrastructure elements, from images obtained by MMS technology. These images consisted of a variety of road elements and a wide range of image scales, resolutions and compositions. Most of the images contain complex backgrounds and occlusions caused by trees or buildings. This made the road identification task extremely challenging.

It is shown that neural networks, such as CNN, significantly improved the detection accuracy compared to traditional approaches. The reliability of neural networks depends on the network's training and the selection of operational parameters. Here are elaborated some newest researches with their results, future work, parameters and usage.

Automated detection of road infrastructure elements has become an increasing necessity for transportation-related activities, including traffic monitoring, automatic vehicle driving, and autonomous navigation. Even mentioned methodologies improved object detection from images and positioning based on MMS data, there is still a lot of space for further development.

REFERENCES

[1] Vasić D., Davidović M., Radosavljević I., Obradović Đ. (2021), "Architecture of solution for panoramic image blurring in GIS project application", Geoscientific Instrumentation Methods and Data Systems, vol. 10, no. 2, pp. 287-296.

[2] Radovic M., Aderkwa O. and Wang Q. (2017), "Object Recognition in Aerial Images Using Convolutional Neural Networks", Journal of imaging, vol. 3 (2), no. 21

[3] Vasić D., Batilović M., Davidović M., Kuzmić T. (2019), "Modern Methods of Processing and Extracting Data from Point Cloud", FIG Working Week 2019, Hanoi, Vietnam, April 2019

[4] Sommer L., Schuchert T. and Beyerer J. (2017), "Deep learning based multi-category object detection in aerial images" , Automatic Target Recognition XXVII Proceedings, vol. 10202

[5] Božić- Štulić D., Krzić S., Gotovac S. and Papić V. (2018), "Complete Model for Automatic Object Detection and Localisation on Aerial Images using Convolutional Neural Networks", Journal of communications software and systems, vol. 14, no. 1

[6] Li Q., Chen L., Li M., Shaw S.-L. and Nüchter A. (2014), "A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios," IEEE Trans. Veh. Technol., vol. 63, no. 2, pp. 540–555.

[7] Du B., Zhang M., Zhang L., Hu R. and Tao D. (2017), "PLTD: Patch-based low-rank tensor decomposition for hyperspectral images," IEEE Trans. Multimedia, vol. 19, no. 1, pp. 67–79.

[8] Bonnefon R., Dhérété P. and Desachy J. (2002), "Geographic information system updating using remote sensing images," Pattern Recognit. Lett., vol. 23, no. 9, pp. 1073–1083.

[9] Zhang Z., Liu Q. and Wang Y. (2018), "Road extraction by deep residual U-net," IEEE Geosci. Remote Sens. Lett., vol. 15, no. 5, pp. 749–753.

[10] Cheng G., Wang Y., Xu S., Wang H., Xiang S. and Pan C. (2017), "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," IEEE Trans. Geosci. Remote Sens., vol. 55, no. 6, pp. 3322–3337.

[11] Yang X., Li X., Ye Y., Lau R. Y. K., Zhang X. and Huang X. (2019), "Road Detection and Centerline Extraction Via Deep Recurrent Convolutional Neural Network U-Net", IEEE Transactions on geoscience and remote sensing, vol. 57, no. 9

[12] Tian, Y., Gelernter, J., Wang, X., Chen, W., Gao, J., Zhang, Y. and Li, X. (2018), "Lane marking detection via deep convolutional neural network," Neuro computing, vol. 280, pp. 46–55.

[13] Xiao, D., Yang, X., Li, J. and Islam, M. (2020), "Attention deep neural network for lane marking detection", Knowl-Based Syst., vol. 194, no. 105584, pp. 1–10.

[14] Guan H., Lei X., Yu Y., Zhao H., Peng D., Marcato J. and Li J. (2022), "Road marking extraction in UAV imagery using attentive capsule feature pyramid network", International Journal of Applied Earth Observations and Geoinformation, vol. 107, no. 102677

[15] Murthy V.S., Gupta S. and Mohanta, D.K. (2011), "Digital image processing approach using combined wavelet hidden Markov model for well-being analysis of insulators", IET Image Process., vol. 5, pp. 171–183.

[16] Song B. and Li X. (2014), "Power line detection from optical images", Neuro computing vol. 129, pp. 350–361.

[17] Sharma H., Adithya V., Dutta T. and Balamuralidhar, P. (2015), "Image Analysis-Based Automatic Utility Pole Detection for Remote Surveillance", Proceedings of the 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Adelaide, Australia

[18] Zhang W., Witharana C., Li W., Zhang C., Li X. and Parent J. (2018), "Using Deep Learning to Identify Utility Poles with Cross arms and Estimate Their Locations from Google Street View Images", Sensors, vol. 18, no. 2484

BIOGRAPHICAL NOTES

Dejan Vasić was born in Sarajevo, Bosnia and Herzegovina, in 1980. He received the Ph.D. degree in geodesy and geomatics from FTS, UNS, Novi Sad, Serbia in 2018. Currently, he is an Assistant Professor at the FTS, UNS. His areas of interest are 3D terrestrial and airborne laser scanning, BIM modelling and Engineering Geodesy.

Marina Davidović was born in Foča, Bosnia and Herzegovina, in 1992. She received the B.Sc. and M.Sc. degrees in geodesy and geomatics from the FTS, UNS, Novi Sad, Serbia in 2015 and
2016, respectively. Currently, she is an Associate Researcher at the Faculty of Technical Sciences FTS, University of Novi Sad (UNS) and works at DataDEV. Her areas of interest are modern acquisition technologies, especially mobile mapping systems.

CONTACTS

Assistant Professor Dejan Vasić
University of Novi Sad, Faculty of Technical Sciences
Dositej Obradović Square 6
Novi Sad
Serbia
Tel. +381658097787
Email: dvasic@uns.ac.rs
Web site: http://www.ftn.uns.ac.rs/691618389/fakultet-tehnickih-nauka

Associate Researcher Marina Davidović
University of Novi Sad, Faculty of Technical Sciences
Dositej Obradović Square 6
Novi Sad
Serbia
Tel. +381655707104
Email: marina.davidovic@uns.ac.rs
Web site: http://www.ftn.uns.ac.rs/691618389/fakultet-tehnickih-nauka