IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9554811

INTEGRATION OF PHOTOGRAMMETRY AND DEEP LEARNING IN EARTH OBSERVATION APPLICATIONS

José Marcato Junior[1]; Pedro Zamboni[1], Mariana Campos[2], Ana Ramos[3], Lucas Osco[3], Jonathan Silva[1], Wesley Gonçalves[1], Jonathan Li[4]

[1] Federal University of Mato Grosso do Sul, UFMS, Campo Grande, MS, Brazil
[2] Finnish Geospatial Research Institute, FGI, Finland
[3] University of Western São Paulo, Presidente Prudente, SP, Brazil
[4] University of Waterloo, Waterloo, Canada

ABSTRACT

The integration of photogrammetry and deep learning methods can be powerful for Earth observation applications. Photogrammetry techniques allow the achievement of detailed geospatial products with cm-level positional accuracy. Deep learning enables automatic image classification, segmentation, and object detection. For instance, when dealing with a large data set, photogrammetric processing steps, such as image orientation and dense point cloud generation, results in high computational costs. In contrast, deep learning methods are fast in the inference step. Here, we explore the complementarity of deep learning and photogrammetry, aiming to generate accurate and fast geospatial information. The main aim is to discuss the possibilities of using deep learning in the photogrammetric process. We conduct experiments to present the potential of the Mask R-CNN method trained on the COCO dataset to generate masks, essential to remove image observations from moving objects during the orientation (alignment) step.

Index Terms— remote sensing, structure-from-motion, computer vision, machine learning

1. INTRODUCTION

Images from low-cost RGB cameras attached to unmanned aerial vehicles (UAV) or mobile mapping systems have been widely used. Photogrammetric techniques can provide high-detailed and accurate geospatial data [1], even using image datasets from these image sensors.

The photogrammetric process can be summarized in three steps: (a) sensor (inner) and platform (exterior) orientation (alignment); (b) dense cloud generation; and (c) orthoimage generation. This process is currently based on the computer vision techniques structure from motion (SfM) [2] and multi-view stereo (MVS) [3], which can be computationally expensive when dealing with a large dataset.

In another perspective, deep learning has been achieved outstanding results in several remote sensing applications [4][5]. When proper data is available, deep learning generally outperforms, in terms of accuracy, traditional remote sensing, and machine learning techniques in image classification tasks. Methods based on deep learning are expensive to train, requiring powerful graphics processing units (GPU); however, they are fast in the inference process after training the model [6].

Deep learning methods have been applied directly in orthoimages or point clouds to generate geospatial information. In summary, first, the orthoimages and point cloud are generated in the photogrammetric process, and after the objects are mapped from them based on deep learning techniques. Therefore, there is no integration between photogrammetry and deep learning.

Here, we explore the complementarity of deep learning and photogrammetry, aiming to generate accurate and fast geospatial information. The main purpose of this work is to discuss the possibilities of integrating deep learning into the photogrammetric process. In the experimental section, we especially focus on identifying moving objects in the mobile mapping images using a pre-trained deep learning method, aiming to avoid unstable observations in the photogrammetric orientation (alignment) procedure. The removal of these observations can provide a more robust solution for the orientation procedure, generating, as a consequence, more reliable point clouds and orthoimages. Even not considered in the experiments, other integration possibilities are discussed.

2. BACKGROUND

This section presents the photogrammetric processing steps (Section 2.1) and the deep learning methods for object detection and segmentation (Section 2.2).

2.1. Photogrammetric process

The first step in the photogrammetric process is the sensor (inner) and platform (exterior) orientation, also known as alignment. This process is currently based on the SfM [2] method, in which the sensor interior and exterior orientation parameters are estimated simultaneously. The success of this estimation is directly related to the quality of image observations obtained automatically in the image matching procedure. In urban scenarios, moving objects can be tracked in the images and outputted as image observations from the image matching procedure, including errors in the alignment process and degrading the next photogrammetric processes, such as dense cloud generation. To cope with this problem, here we propose the integration of deep learning-based methods (Section 2.2) in the photogrammetric process to identify and mask moving objects from the images, avoiding the detection of these objects by image matching operators.

We believe that the integration of deep learning in the photogrammetric process can provide a more reliable alignment, which is underlying for the second photogrammetric process step, which is the generation of the dense cloud (digital surface model – DSM). The DSM generation is based on the MVS method [3]. And as a final step, the orthoimages are generated.

2.2. Deep learning-based methods

Deep learning methods enable automatic image classification, segmentation, and object detection, which can be applied to optimize the photogrammetric process. In our work, we focus on the use of object detection and segmentation.

Object detection methods generate bounding boxes on objects of interest. Applications in remote sensing with these methods have been increasing [7]. Several novel benchmarks and methods were proposed recently. In remote sensing applications, Faster R-CNN [8] and RetinaNet [9], and are the most used methods. Recently, new methods, such as ATSS [10], had been developed, providing accurate results for pole detection in aerial imagery [11] and apple detection in terrestrial images [12].

Segmentation methods aim to establish a class for each pixel on the image. Regarding semantic segmentation, SegNet [13], U-Net [14], Deeplab [15] and others have been mostly explored in remote sensing [16][17]. In our work, an instance segmentation method (Mask R-CNN) to define each object's bounding box and classify the pixels inside the bounding box was assessed. The Mask R-CNN [18] is a well-known instance segmentation method, which was adopted in the current work as a mask generator for undesirable moving objects in the image.

3. METHODOLOGY

3.1. Dataset

The dataset is composed of 287 frames with a resolution of 1280 × 720 pixels, acquired with a GoPro HERO6 Black RGB camera. The frames were generated based on a video considering one fps (frame per second) in a street of the municipality of Campo Grande, Mato Grosso do Sul, Brazil. This dataset was also used in our previous work [19], aiming to detect manholes and storm drain.

3.2. Mask generation

We used a pre-trained Mask R-CNN [18] through the MMDetection toolbox [20] with the COCO dataset. Mask R-CNN generates, for each detected object, the mask with the edges of the object and the corresponding bounding box. Nevertheless, to generate the mask files used in the photogrammetric process, the following categories were considered: person, bicycle, car, motorcycle, bus, and truck. Those categories were selected since they are, in general, moving objects in the frames. In the results section, we present a qualitative analysis and discuss the potential and limitations of using this technique as a mask generator.

4. RESULTS

Figure 1 presents the results using the Mask R-CNN method, in which bounding box and segmentation in cars, motorcycles, and pedestrians are generated.

In general, the achieved results are accurate, generating bounding boxes and segmentation masks for the objects of interest. However, in some situations, objects that are not from the classes of interest are detected, as depicted with a red circle in Figure 1. Finally, mask files (see Figure 2) are generated and used in photogrammetric software as input. This can be used in commercial or open-source software, such as Agisoft Metashape, Pix4D, and MicMac. With the adopted procedure, it is possible to remove moving objects that can interfere in the alignment procedure.

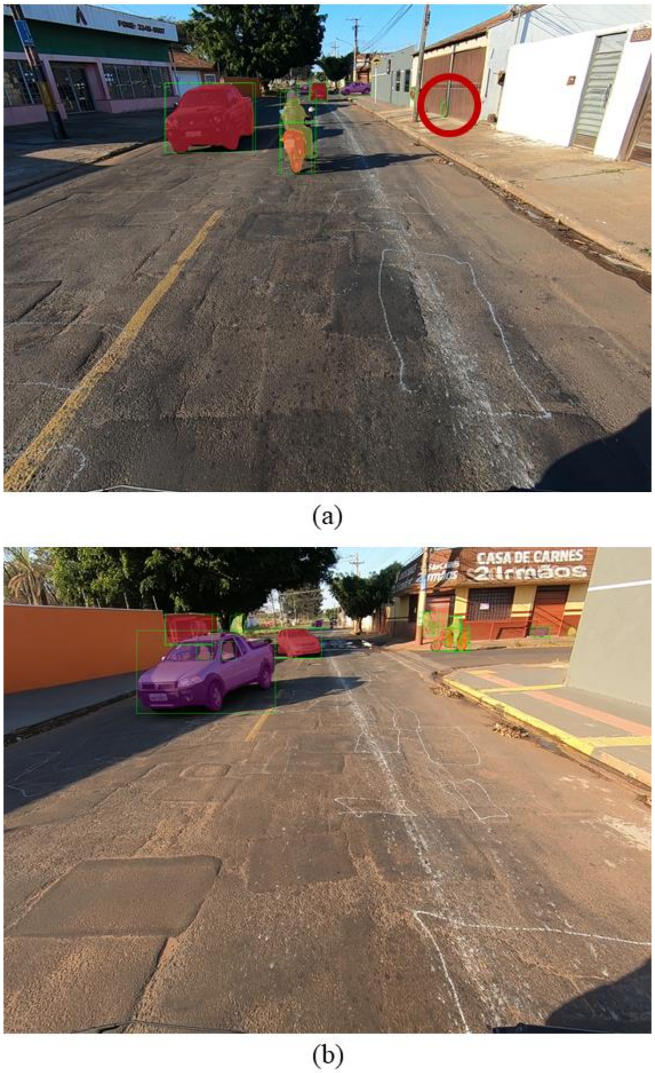

Figure 1: Examples of bounding box and masks generated using Mask R-CNN model pre-trained on Coco dataset.

Based on the experiments, we verified that even considering pre-trained models, satisfactory results are achieved. Even not explored, the same strategy of using masks (but now for the objects of interest) can be used when generating the point cloud, which is a subsequent step in the photogrammetric process. Consequently, only point clouds from the objects of interest would be generated, reducing the storage requirements and the processing cost. Also, there would be no need to apply deep learning methods again, as only points of the objects of interest would belong to the point cloud.

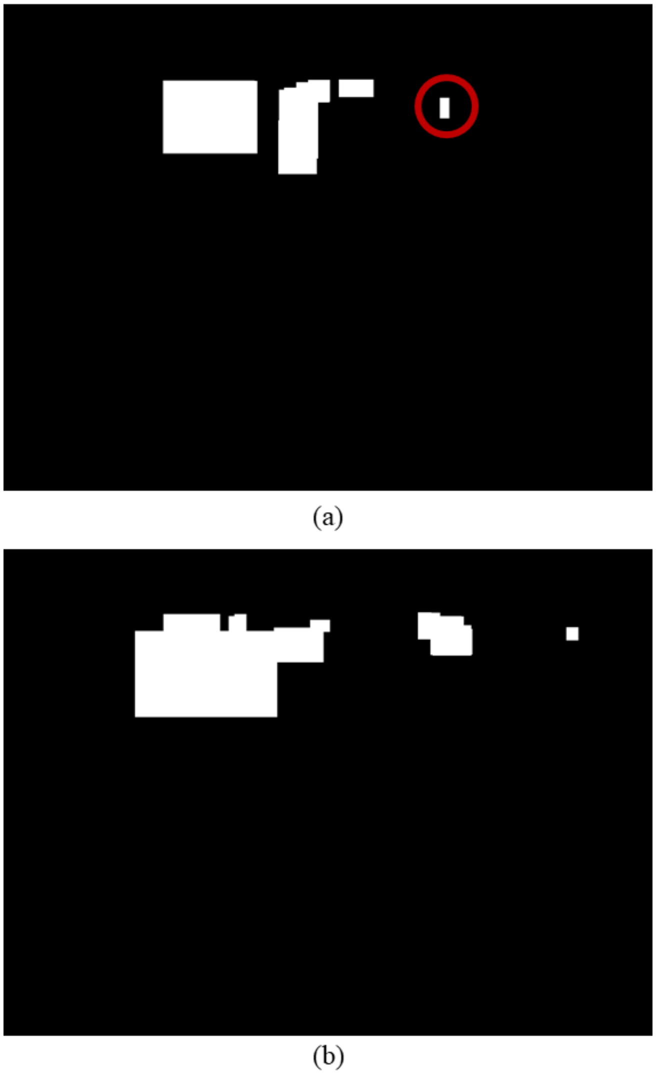

Figure 2: Bounding box masks used in the photogrammetric processing.

5. CONCLUSION

This paper presented potential applications of deep learning to improve and optimize photogrammetric processing. Here, we present a solution for image orientation in close-ranging photogrammetry considering a mobile mapping application in urban areas, which is rich in moving objects. In the proposed approach, a deep learning model - trained in a generic dataset, was used to generate masks, which can minimize the detection of homologous points by image matching methods in moving objects.

Based on the experiments, we verified that even considering pre-trained models, satisfactory results are achieved. The adopted strategy can benificiate several Earth observation applications in urban environments, including tree species mapping and asphalt monitoring.

6. ACKNOWLEDGMENTS

This research was funded by CNPq (p: 433783/2018-4, 304052/2019-1 and 303559/2019-5), FUNDECT (p: 59/300.066/2015), and CAPES PrInt (p: 88881.311850/2018-01). The authors acknowledge the support of the UFMS (Federal University of Mato Grosso do Sul) and CAPES (Finance Code 001).

7. REFERENCES

[1] M.B. Campos, A.M.G. Tommaselli, J. Marcato Junior, and E. Honkavaara, "Geometric model and assessment of a dual-fisheye imaging system," *The Photogrammetric Record*, vol. 33, no. 162, pp. 243–263, 2018.

[2] S. Ullman. The interpretation of structure from motion. Proc. R. Soc. Lond. B Biol. Sci. 1979, 203, 405–426.

[3] S.M., Seitz, B. Curless, J. Diebel, D. Scharstein, R. A. Szeliski. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; pp. 519–528.

[4] L. P. Osco, M. S. Arruda, J. Marcato Junior, N. B. Silva, A. P. M. Ramos, E. A. S. Moryia, N. N. Imai, Da. R. Pereira, J. E. Creste, E. T. Matsubara, J. Li, and W. N. Goncalves, "A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020.

[5] L.P. Osco, M.S. de Arruda, D.N. Goncalves, A. Dias, J. Batistoti, M. Souza, F.D.G. Gomes, A.P.M. Ramos, L.A.C. Jorge, V. Liesenberg, J. Li, L. Ma, J. Marcato Junior, and W.N. Goncalves, "A cnn approach to simultaneously count plants and detect plantation-rows from uav imagery," 2020.

[6] A.A. Santos, J. Marcato Junior, M.S. Araujo, D.R. Di Martini, E.C. Tetila, H.L. Siqueira, C. Aoki, Anette Eltner, E.T. Matsubara, H. Pistori, R.Q Feitosa, V. Liesenberg, and W.N. Goncalves, "Assessment of cnn-based methods for individual tree detection on images captured by rgb cameras attached to uavs," *Sensors.*, vol. 19, no. 16, pp. 3595, 2019.

[7] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS J. Photogramm. Remote Sens.*, vol. 159, pp. 296–307, 2020.

[8] S. Ren, K. He, R. Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," 2015.

[9] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," 2017.

[10] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S.Z. Li, "Bridging the gap between anchorbased and anchor-free detection via adaptive training sample selection," 2019.

[11] M. Gomes, J. Silva, D. Gonc¸alves, P. Zamboni, J. Perez, E. Batista, A. Ramos, L. Osco, E. Matsubara, J. Li, J. Marcato Junior, and W. Gonc¸alves, "Mapping utility poles in aerial orthoimages using atss deep learning method," *Sensors*, vol. 20, no. 21, pp. 6070, 2020.

[12] L.J Biffi, E. Mitishita, V. Liesenberg, A. A. Santos, D. N. Gonc¸alves, N. V. Estrabis, J. A. Silva, L. P. Osco, A. P. M. R. Ramos, J. A. S. Centeno, M. B. Schimalski, L. Rufato, S. L. R. Neto, J. Marcato Junior, and W. Goncalves, "Atss deep learning-based approach to detect apple fruits," *Remote Sensing*, vol. 13, no. 1, pp. 54, 2021.

[13] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481–2495, 2017.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015.

[15] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A.L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

[16] D.L. Torres, R.Q. Feitosa, P.N. Happ, L.E.C. La Rosa, J. Marcato Junior, J. Martins, P.O. Bressan, W.N. Goncalves, and V. Liesenberg, "Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery," *Sensors.*, 2020.

[17] L.P. Osco, K. Nogueira, A.P.M. Ramos, et al. "Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery," *Precision Agric*, 2021.

[18] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-cnn," in 2017 *IEEE International Conference*.

[19] A.A. Santos, J. Marcato Junior, J.A. Silva, R. Pereira, D. Matos, G. Menezes, L. Higa, A. Eltner, A.P. Ramos, L. Osco, and et al., "Stormdrain and manhole detection using the retinanet method," *Sensors*, vol. 20, no. 16, pp. 4450, 2020.

[20] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C.C. Loy, and D. Lin, "Mmdetection: Open mmlab detection toolbox and benchmark," 2019.