OPTIMIZATION OF EXTRACTED BUILDING FOOTPRINTS FROM UAV IMAGES

Miloš Tutnjević[a], Miro Govedarica[b], Gordana Jakovljević*[c]
[a]NMT Geoplan, Bosnia and Herzegovina; [b], Faculty of Technical Science, University Novi Sad, Serbia [c] Faculty of Architecture, Civil Engineering and Geodesy, University of Banja Luka, Bosnia and Herzegovina

ABSTRACT

Building footprints is important information in many types of applications, including optimization of rescuer response in case of catastrophic events, urban planning, urban dynamic monitoring, 3D building modeling etc. Traditionally, in remote sensing, building footprints are detected from very high-resolution images or point clouds. Convolution Neural Network (CNN) based semantic image segmentation model has become a common way to extract buildings footprints from remote sensing data with high accuracy regardless of differences in landscapes, shapes, texture, and used materials. However, the results of extraction usually represent rooftop outlines with overhangs rather than true building footprints.

This paper presents the methodology for the optimization of building footprints by using contour information, which is derived from the UAV point cloud. First, the CNN model was used to extract rooftops from high-resolution UAV-based orthophoto. After that, the cross-section of the mesh model was performed in order to detect the outline of the building. The optimum height of the mesh cross section was derived based on an analysis of the Digital Elevation Model and Digital Surface Model. The generated results were compared with Open Street Map (OSM) and reference cadastral datasets. Quantitative and qualitative evaluations show that the proposed methodology can significantly improve the accuracy of CNN-extracted building footprints (and OSM data) compared to cadastral data. In addition, the high of buildings is simultaneously derived. Therefore, our approach opens up the possibility to use the full potential of UAV products for generating accurate building footprints and 3D building models of LoD1 with compatible accuracy as cadastral.

Keywords: Building footprint, UAV, semantic segmentation, Point cloud

1. INTRODUCTION

Buildings are one of the most important elements of the urban environment. The information about the spatial distribution of buildings' footprints is essential in many applications including cadastral, urban planning, disaster monitoring, smart city, 3D city modeling, and many other aspects [1], [2], [3]. The economical development and urbanization, as a complex and continuous process, have led to rapid urban expansion and a continuous increase in the number of building footprints. In addition, in undeveloped countries, there is a large number of illegal constructions making the official cadastral database outdated which represents a major challenge in urban planning and sustainable development. Therefore, accurate footprint detection and monitoring of changes in an urban environment are critical.

Traditional building survey methods provide high accuracy but they are also time and resource intensive. Remote sensing data is one of the most important data sources for monitoring urban areas and updating data in the cadastral system. In recent times, the implementation of remote sensing for land management and cadastral updating has been significantly improved due to advantages in both data acquisition and data processing technologies. Various approaches have been used to extract building footprints. Most of them are based on high-resolution imagery or point cloud data [4].

The automatic building footprint extraction method represents an effective way to acquire building information over large areas using satellite or airborne images. Accurate automatic extraction of building footprints remains challenging mostly due to the complexity of the shapes, sizes, color, materials of buildings, and textures, (2) buildings are often surrounded by obstacles such as tall trees, and their shadows.

Ninth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2023),
K. Themistocleous, D. G. Hadjimitsis, S. Michaelides, G. Papadavid, Eds., Proc. of SPIE Vol. 12786, 127860T
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2681315

Over the past decades, different approaches including pixel-based, object-based, edge-based [5], and shadow-based methods [6], [7], were proposed. Belgiu and Dragut [8] compared the supervised and unsupervised multi-resolution segmentation method combined with Random Forest (RF) classifies for building footprint extraction. Singh and Mehrotra [7] employed a Differential Morphological Shadow Operator and Differential Morphological Building Operator for extracting shadows and buildings respectively from high-resolution satellite images. Chen et al. [5] proposed edge regularity indices and shadow line indices as new features and tested three classifiers including AdaBoost, RF, and Supported Vector Machine (SVM) for house identification. Although machine learning algorithms can provide high accuracy, they rely on the manual design of input making them sensitive to expert knowledge which usually leads to poor generalization with different environmental backgrounds [9].

In recent years, Convolution Neural Networks (CNN) are employed for developing algorithms for the automatic extraction of building footprints achieving state-of-the-art accuracy. Xu et al. [10] develop a building footprint extraction method based on ResUNet architecture combined with an object-oriented guided filter using ISPRS 2D labeled VHR satellite image dataset. Li et al. [11] combined UNet, Cascade R-CNN, and Cascade CNN models to provide semantic segmentation maps, boundary boxes, and corners of buildings respectively, in the WHU database while Zhang et al. [12] combined atrouse attention pyramid model and the dual graph convolution model for the detection of building footprints from WHU dataset. Li et al. [13] integrated the CNN and feature pairwise condition random field graph model to preserve sharp building boundaries and fine-grained segmentation.

Unmanned Aerial Vehicle (UAV) offers an affordable, effective, and flexible approach to the low-cost collection of high-resolution images. In addition, UAV represents a rich source of not only 2D but also 3D information about the environment due to the Structure-from-motion algorithm (SfM) producing point clouds, the Digital Elevation Model (DEM), and the Digital Surface Model (DSM). However, despite the ability to collect fine detail, due to limited payloads most UAV are limited to the collection of ultra-high resolution RGB images, making a low interclass and high intraclass variance on images [14]. Due to that, it is hard to accurately distinguish building boundaries from the complex background. It has been shown that the inclusion of different visual features in deep learning models can improve classification accuracy. Kokeza et all. [15] combined RGB images and normalized DSM (nDSM) to extract the buildings using ResUNet architecture. Results show that usage of nDSM has decreased the number of false negative pixels resulting in higher accuracy. Similarly, Boonpook et al. [16] combined RGB data with Visible Band Vegetation Index (VDVI) and DSM to extract building footprints using SegNet deep learning model providing better separation between building and vegetation (VDVI), building and ground object (DSM), and identification of small buildings obstructed by vegetation.

Although the ability to accurately extract buildings from remote sensing images has been increased with advantages in deep learning, the existing approaches assume that the building footprints and rooftops are well overlapped. However, the rooftops usually contain overhangs and therefore don't contain information about building footprints (contact line between building façade and ground). Taking that into account, the oblique images and terrestrial point cloud can be used for accurate building footprint generation. Wang et all. [17] proposed an offset vector learning scheme based on off-nadir areal images and Mask R-CNN model, which estimates footprints by translating the predicted roof masks according to the predicted offset vector. Similarly, Zhuo et al. [18] proposed the optimization of OSM building footprints by exploring contour information from oblique UAV images based on FCN architecture. The results reveal that the accuracy of OSM compared to official cadastral data is significantly improved from meter level to decimeter level.

In this paper, we proposed a methodology for the optimization of building footprints extracted from high-resolution UAV images in order to detect real building footprints excluding roof overhangs. Our approach is on the detection of building footprints in 3D space by creating a cross-section between mesh structures and detected rooftops from orthophoto.

2. MATERIALS AND METHODS

In this paper, optimization of extracted building footprints from high-resolution UAV images based on point cloud is proposed. The workflow which summarizes the approach followed in this paper is shown in Figure 1. The workflow consists of four main steps: data acquisition, preprocessing, processing, and accuracy assessment.

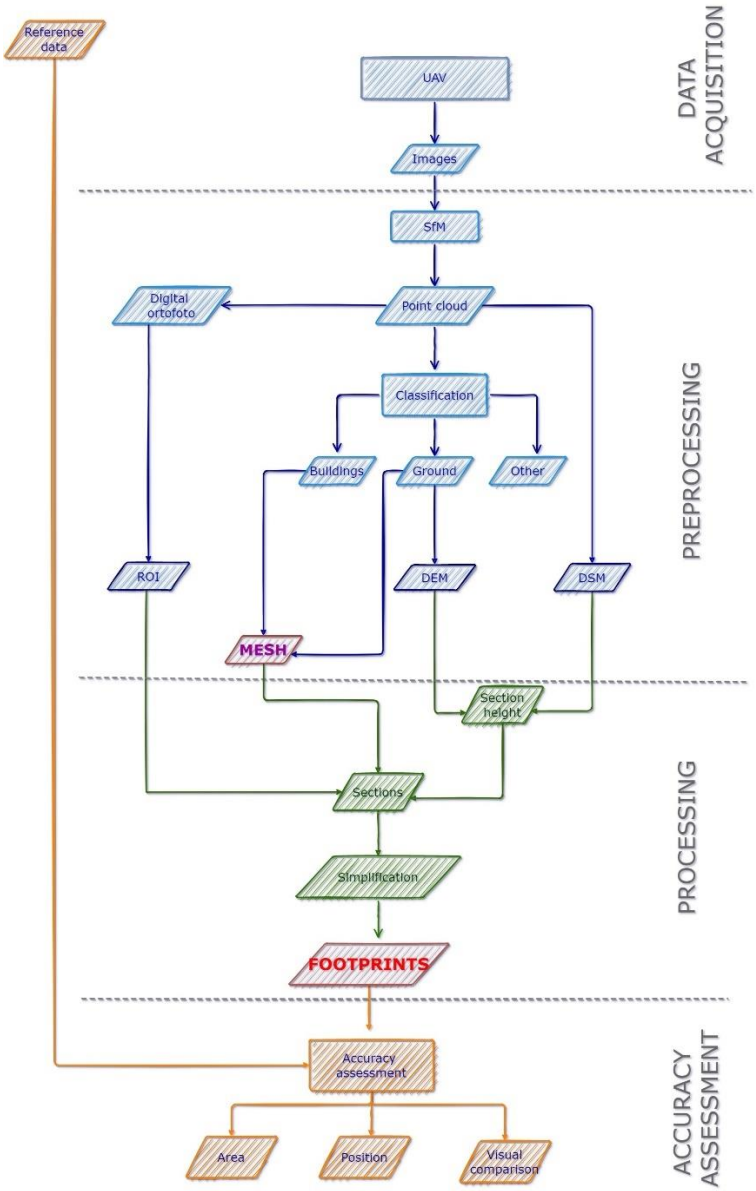


Figure 1. Workflow

1. Study area

The study area used in this research is part of Novo Naselje, city of Gradiška. City of Gradiška is located in the northwest part of the Republic of Srpska, Bosnia and Herzegovina, on the right bank of Sava river which represents the natural border between Croatia and Bosnia. According to the 2013 census, the total population of the city is approximately 52000, and it has an area of 721,62 km^2. The study area covers 9.31 ha representing a part of the Lijevče plain and it is characterized by dominantly flat terrain. The research area is part of a rural area with complex surface composites and high building density. In addition to single-family homes, there is a large number of accessory buildings such as garages, storage sheds, stables, etc. Moreover, some of the buildings are surrounded and shaded by tall trees or shrubs. The overview of the study area is presented in Figure 2.

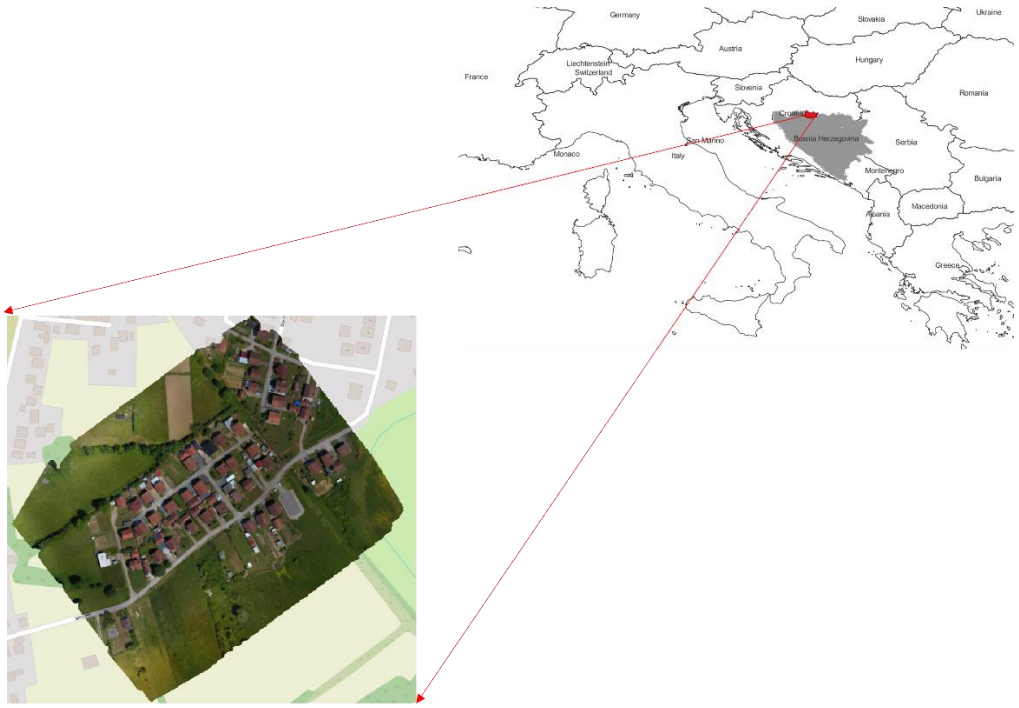


Figure 2. Studyarea

2. Data acquisition

Reference data: Reference datasat containing the 2D representation of buildings and surrendering objects was used as the true data. The data were collected by classical surveying methods. According to the law, the horizontal error must not exceed 10 cm which is accomplished with polar methods from different points determined by classical or GNSS positioning methods. The survey of the study area was conducted in 2022.

UAV data: The experimental data utilized in this research were collected by using a four-rotor UAV DJI Phantom 4 Pro equipped with onboard capabilities, a positioning system (Glonass/GPS), and an IMU sensor and RGB camera. Characteristics of UAVs are presented in Table 1. The 3D Survey Pilot application was used for flight planning. The double grid mission, which consists of the consecutive nadir rectangle missions, was used to ensure that images are taken from multiple sides avoiding dead angles. The front and side overlap was 80% and the flight height was set to 60 m. A total of 576 images were acquired with a spatial resolution of 2.26 cm/pix.

In order to provide a comparison with reference data, georeferencing by using 11 Ground Control Points (GCP) was performed. The position of GCP was determined by the GNSS RTK method resulting in a horizontal position higher than 2 cm and vertical accuracy higher than 3 cm. The GCP is marked using the color spray in order to provide easy and accurate identification of images.

Table 1. Characteristics of DJI Phantom 4 Pro

Item	Characteristics
Sensor	1” CMOS with 20 MP
FOV	84 8.8 mm/24 mm
Hover accuracy range	Vertical: up to +-0.1 m Horizontal: up to +- 0.3 m

3. Preprocessing

The image acquired by UAV and the SfM algorithm were used to generate high-resolution orthophoto. SfM represents the process of reconstruction of 3D structure from a set of 2D images taken from different viewpoints. SfM consist of three main steps [19]: (1) the SIFT algorithm is used to detect and describe set of local features that are invariant under radiometric and geometric changes so that the RANdom Sample Consensus (RANSCAN) method can match unique features across multiple images enabling initial estimation of camera positions and object coordinates. The bundle block adjustment of matched features was used to refine the extrinsic and intrinsic camera parameters that minimize the reprojection error between the image locations of observed and predicted image points, and three-dimensional (3D) coordinates for a sparse unscale point cloud. The camera positions derived from SfM lack scale and orientation. Due to that, the point clouds are generated in an image-space coordinate system. The transformation from image space to an absolute coordinate system is achieved by using GCP; (2) point cloud densification; and (3) digital terrain model and orthophoto generation.

In this research, SfM implemented in AgiSoft PhotoScan was used. Root mean square errors of georeferencing were 1.6 cm in the horizontal and 6.5 cm in the vertical plane. The UAV images and dense point clouds were used to create a digital orthophoto. Additionally, dense point clouds were classified as ground, buildings, high vegetation, roads, and other artificial objects. The ground class was used for the production of DEM while all classes were used for the creation of the Digital Surface Model (DSM). The 3D mesh was created by using points classified as ground and buildings.

In this phase, the building footprints are created by using orthophoto and the ResUNet-50 network following the approach presented in [15].

4. Processing

The building footprints addressed in previous research are the rooftops with overhangs. In order to be able to detect the real building footprints i.e. edges where building walls meet ground excluding roof overhangs methode direct building footprint detection in 3D space is proposed. The model is based on a cross-section of a 3D mesh model with detected building rooftops at a suitable height. The concert of the proposed methodology is presented at Figure 3.

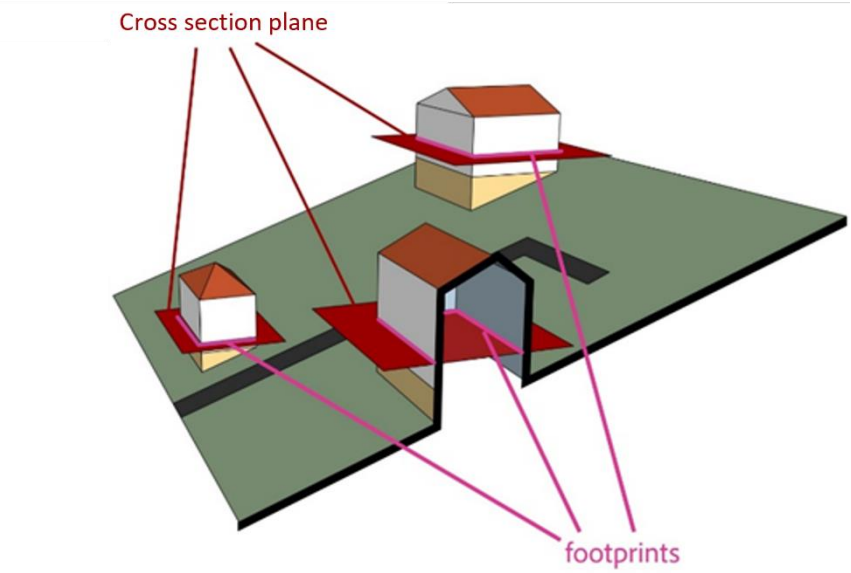


Figure 3. Proposed methodology for extraction of building footprints in 3D space

The optimal height of the cross-section was determined based on DEM, DSM, and Region Of Interest (ROI) which represents building rooftops detected in the preprocessing phase. Firstly, the minimum and maximum value of DEM in ROI was extracted. The maximum height within ROI should represent the optimal height of the cross-section. However,

the dense point cloud wasn't uniforme dense. Due to dead angles, shadows etc. the holes are presented in 3D point clouds mostly in areas under overhangs and near to the ground. Taking that into account during mesh creation the interpolation was used to create triangles in those areas. Therefore, the created mesh of the building will not accurately represent the objective shape between vertical plane and overhangs and vertical plane and ground (Figure 3.). Consequently, creating cross-section at maximum height of DEM within ROI would result in a larger building footprint.



Figure 4. 3D model of the building represented by mesh structure.

Due to that the optimal height of the cross section was determined by introducing the bias Δh which is determined based on building height. Building height is calculated as the difference between maximum value of DSM and minimum value of DEM within the ROI (Figure 5). Based on testing the optimal height of cross section (H_{ROI}) in this study is defined using the following expression:

$$H_{ROI} = H_{DEM}^{max} + \Delta h$$

where Δh = 0.5 if the building height is less than 2.35 m and Δh = 1.5 m if the building height is greater than 2.35 m

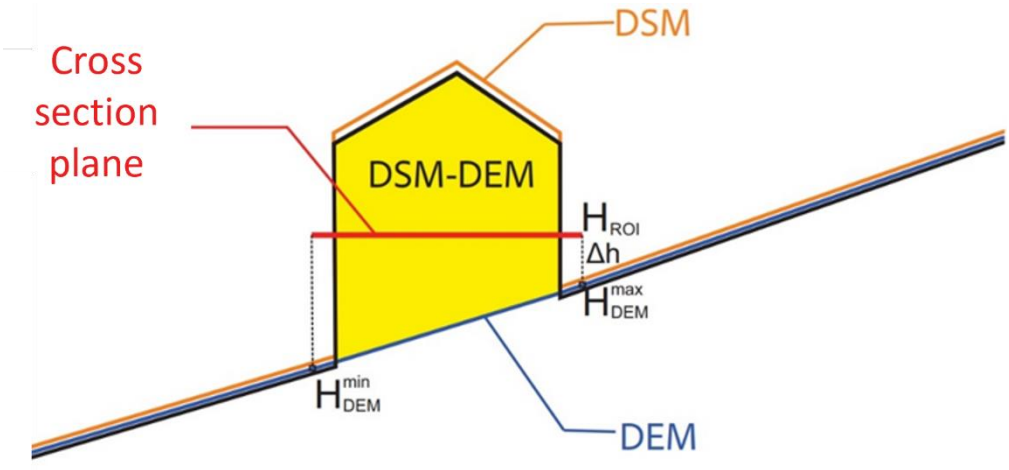


Figure 5. Determining the optimum height of the cross-section

Buildings mostly have regular boundaries that consist of straight lines, while the created cross-sections, due to the triangle mesh structure, result in buildings footprints with irregular outlines and redundant vertices (Figure 6). Therefore, in order to provide optimization of building footprints it is necessary to perform generalization of extracted outlines. Although, methods for generalization by reducing the number of vertices, such as Douglas-Peucker, are well

established, in the case of building shapes it is necessary to force several geometrical rules such as diagonal lenght equality, forcing right angles, etc.

The presented approach is implemented in the Python programming language. The optimal height of the cross section was determined by using the *zonal_stats* function which is available in the ***rasterstat*** library. The cross-section of the mesh model with ROI at optimal height was implemented by using ***trimesh*** library. The generalization of extracted building footprints is performed in ArcGIS software by using *Regularize Building Footprint* tool. This tool provides normalization and regulation of polygons by defining several parameters such as tolerance, density, precision, forcing right angles shapes, etc. More information about this tool is available at [20].

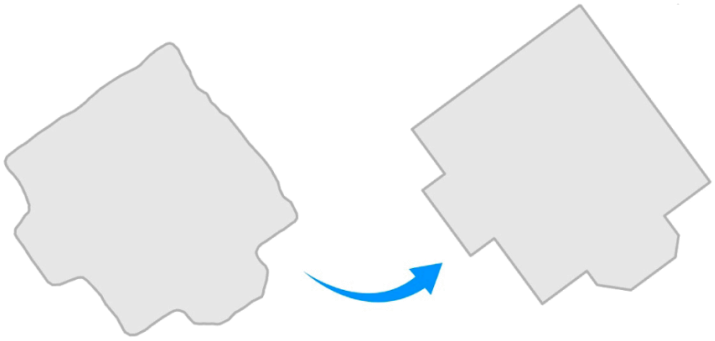


Figure 6. Left: Building footprint derived from mesh model, Right: Building footprint after generalization

3. RESULTS AND DISCUSSION

In this paper, a novel framework for optimizing building footprints extracted from remote sensing data based on detection in 3D space is proposed. The extracted data were compared with reference data and the OSM building dataset. In addition to the visual comparison of results, the quantitative analysis of the results is performed including the investigation of the positional accuracy of building footprints by calculating the average distance between the corresponding vertices pairs from optimized footprint and reference data and the difference in detected building area between correspondent polygons. The visual inspection of the results is shown at Figure 7. Based on the presented results it can be concluded that the performance of the proposed workflow is depended on the building surrounding. The extracted building footprints for buildings in close neighborhoods don't have any objects greatly overlapped with reference data (Figure 7 (a) and (b)). The algorithm produces lower accuracy in the environment where several buildings were located near neighborhoods. Due to the proximity of the objects, dead angles, and shadows appears making it difficult for the SfM algorithm to detect and match key features between different images. Consequently, the holes in the point cloud will appear decreasing the accuracy of the mesh model and therefore of generated results. The accuracy is also decreased due to the presence of accessory objects that are connected to the main building but not represented in reference datasets such as outdoor staircases, covered terraces, or temporal warehouses resulting in an increase in building footprint areas.

The comparison of detected building area between referenced and generated data is 15.5 % while the difference between generated and OSM data is 20.7 % indicating that the proposed methodology provides area improvement of 5.2% compared with OSM data. The results show that area differences have extreme values for some of the buildings such as in Figure 7 (e) where the detected area is higher by 54% compared with reference data. In this specific case, the two covered terraces lean against the building producing higher error.

Positional accuracy assessment was calculated by paring the nearest nebrhoods between generated and reference data. Totaly 384 were used for assessment. The average positional error was 0.69 m compared with reference data. Regarding the positional comparison between OSM and reference data, a total of 208 points were used. The average positional error was 1.54 m. The lower positional accuracy of OSM dataset is expected due to errors of transformation between different coordinate systems i.e. EPSG:4326 to EPSG: 3908. In addition, the intersection between reference

data and generated polygons was performed. Additionally, Intersection over Union (IoU) was calculated using the following expression:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

The IoU between reference and generated data and reference and OSM data was 0.76 and 0.66 respectively. Higher IoU indicate the more overlap between polygones i.e. greter positional accuracy. Usually, IoU greater than 0.5 is considered as good overplap.

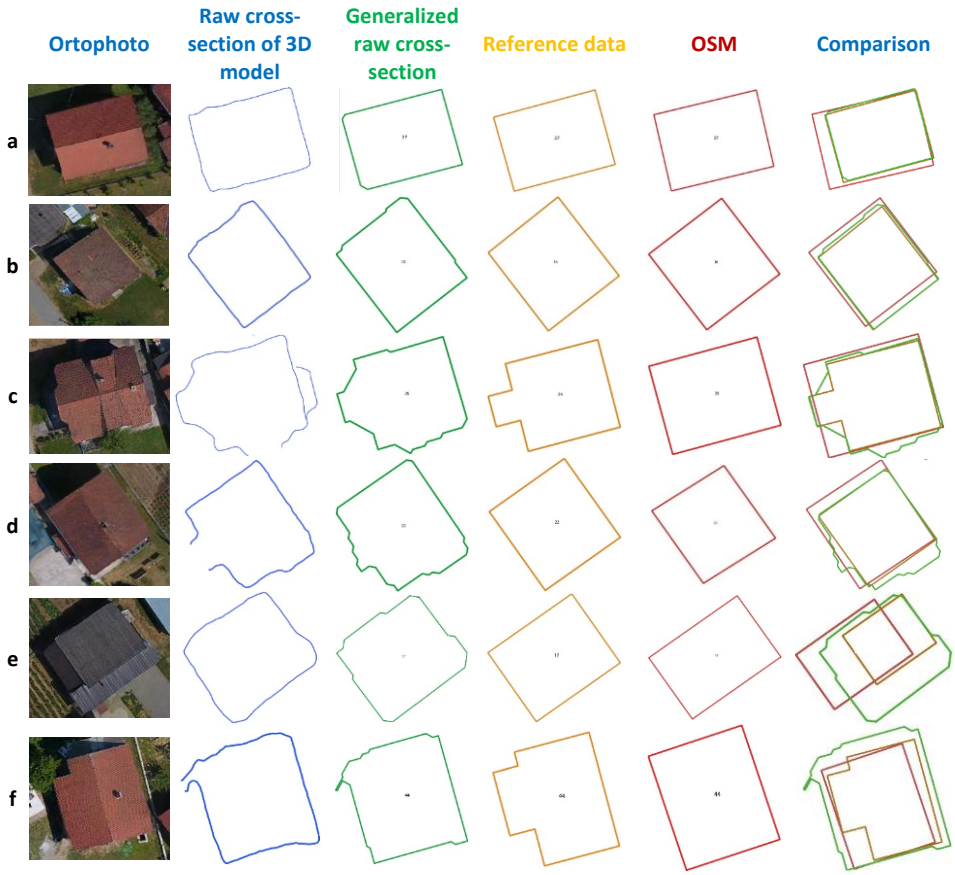


Figure 6. Visual comparasion of results

4. CONCLUSION

In this paper, a novel framework for optimizing building footprints extracted from remote sensing data based on detection in 3D space is proposed. The cross-section of the mesh model was performed in order to detect the outline of the building. The optimum height of the mesh cross section was derived based on an analysis of the DEM and DSM. The generated results were compared with OSM and reference cadastral datasets. Quantitative and qualitative evaluations show that the proposed methodology can significantly improve the accuracy of CNN-extracted building footprints (and OSM data) compared to cadastral data The extracted data were compared with reference data, collected by using the classical surveying method, and the OSM building dataset.

The results of visual inspection and accuracy assessment indicate that the proposed methodology provides optimization of OSM footprints. The accuracy of the building footprint area is increased by 5.2 %. Also, a positional error was reduced from 1.54 m to 0.69 m. The lower positional accuracy of OSM dataset is expected since it includes the error of transformation between different coordinate systems i.e EPSG: 4326 to EPSG:3908. In addition, IoU

metric was used. The results show that both datasets have a good overlap with reference data however extracted dataset has greater IoU compared with OSM i.e. 0.76 and 0.66 respectively. Visual inspection of results shows that model performance is highly limited by building neighborhood. The model provides high accuracy for dislocated objects that don't have accessory buildings in the immediate vicinity while the presence of high vegetation and accessory buildings such as garages, terraces, staircases, etc decrease accuracy.

In future research, the extraction of building footprint from the point cloud should be analyzed.

References

[1] H. Mharzi Alaoui, H. Radoine, J. Chenal, H. Hajji and H. Yakubu, "Deep building footprint extraction for urban risk assessmnet-remote sensing and deep learning based approach," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*, pp. 83-86, 2022.

[2] S. Ural, E. Hussain and J. Shan, "Building population mapping with aerial imagery and GIS data," *International Journal of Applied Earth Observation and Geoinformatics*, vol. 13, pp. 841-852, 2011.

[3] P. Liu, X. Liu, M. Liu, Q. Shi, J. Yang, X. Xu and Y. Zhang, "Building Footprint Extraction from High-Resolution Images via Spatial Residual Inception Convolutional Neural Network," *Remote Sens*, vol. 11, no. 7, 2019.

[4] S. Gilani, M. Awrangjeb and G. Lu, "An automatic building extraction and regularisation technique using lidar point cloud data and orthoimage," *Remote Sens*, vol. 8, 2016.

[5] R. X. Chen, X. H. Li and J. Li, "Object-Based Features for House Detection from RGB High-Resolution Images," *Remote Sens*, vol. 10, no. 451, 2018.

[6] H. Xin, W. Yuan, J. Li and L. Zhang, " New Building Extraction Postprocessing Framework for High-Spatial-Resolution Remote-Sensing Imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*, vol. 10, pp. 654-668, 2017.

[7] K. K. Singh and A. Mehrotra, "Building extraction from VHR imagery using morphological shadow and building operator," in *In Computing for Sustainable Global Development (INDIACom)*, 2014.

[8] M. Belgiu and L. Dragut, "Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery," *ISPRS J. Photogramm. ,* vol. 96, pp. 67-75, 2014.

[9] M. Yu, X. Chen, W. Zhang and Y. Liu, "AGs-Unet: Building Extraction Model for High Resolution Remote Sensing Images Based on Attention Gates U Network," *Sensor*, vol. 22, 2022.

[10] Y. Xu, L. Wu, Z. Xie and Z. Chen, "Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters," *Remote Sens*, vol. 10, no. 144, 2018.

[11] Z. Li, Q. Xin, Y. Sun and M. Cao, "A Deep LearningBased Framework for Automated Extraction of Building Footprint Polygons from Very High-Resolution Aerial Imagery.," *Remote Sens.*, vol. 13, no. 3630, 2021.

[12] W. Zhang, M. Yu, X. Chen, F. Zhou, J. Ren, H. Xu and S. Xu, "Combining Deep Fully Convolutional Network and Graph Convolutional Neural Network for the Extraction of Buildings from Aerial Images," *Buildings*, vol. 12, 2022.

[13] Q. Li, Y. Shi, X. Huang and X. Zhu, "Building Footprint Generation by Integrating Convolution Neural Network With Feature Pairwise Conditional Random Field (FPCRF)," *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, vol. 58, pp. 7502-7519, 2020.

[14] Q. Wang, J. Lin and Y. Yuan, "Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 27, no. 6, pp. 1279-1289, 2016.

[15] Z. Kokeza, M. Vujasinovic, M. Govedarica, B. Milojevic and G. Jakovljevic, "Automatic building footrpint extraction from UAV images using Neural Networks," *Geodetski Vesnik*, vol. 64, no. 4, pp. 545-561, 2020.

[16] W. Boonpook, Y. Tan and B. Xu, "Deep learning-based multi-feature semantic segmentation in building extraction from images of UAV photogrammetry," *Int. J. Remote Sens*, vol. 42, pp. 1-19, 2021.

[17] J. Wang, L. Meng, W. Li, W. Yang, L. Yu and G. Xia, "Learning to Extract Building Footprints from Off-Nadir Aerial Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 45, no. 1, pp. 1294-1301, 2022.

[18] X. Zhuo, F. Fraundorfer, F. Kurz and P. Reinatz, "Optimization of OpenStreetMap Building Footprints Based on Semantic Information of Oblique UAV Images," *Remote Sens.*, vol. 10, no. 4, 2018.

[19] J. L. Schonberger and J.-M. Frahm, "Structure-from-Motion Revisited," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 2016.

[20] ArcGis, "Regularize Building Footprint (3D Analyst)," ArcGis, [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/regularize-building-footprint.htm. [Accessed 10 2 2023].