Automated Building Extraction from Dense LIDAR Data

Damian WIERZBICKI, Elzbieta BIELECKA, and Olga MATUK, Poland

Key words: building extraction, LIDAR, cadaster, LAS data, GIS, Remote Sensing

SUMMARY

Airborne Laser Scanning (ALS) is one of the promising technologies for automated generating building models as it provides three dimensional coordinates of the scanned area. Furthermore, the availability of advanced software tools enables the automatic extraction of particular geographic objects located on the ground surface. The motivation of the conducted study was twofold, firstly – open and free access to dense LIDAR data (12 points per 1 sq. meter) stored in LAS binary format from national geoportal, maintained by the Surveyor General, and secondly – ongoing modernization of Polish cadaster. Hence this study aims to investigate possibility of automatically extraction building's data from LAS data for the modernization of buildings in the cadastre. The experiment was carried out in a housing estate in northern Warsaw, a typical residential area, dominated by detached buildings and green infrastructure. ENVI LIDAR tools were used for building extraction and planar rooftop surfaces delineation. Buildings, trees and other objects (e.g., cars) were separated based on geometric criteria, such as size, height and shape characteristics. The experimental result showed that the proposed methodology achieved good results and was robust after adjusting the model parameters to the specifics of the analyzed area. Dense LAS data, offer fast and cost-effective ways of extracting buildings for cadastral purposes.

Automated Building Extraction from Dense LIDAR Data

Damian WIERZBICKI, Elzbieta BIELECKA, and Olga MATUK, Poland

1. INTRODUCTION

Recent years have brought a rapid development of airborne laser scanning (ALS) technology. The automation of data acquisition and processing ensures quick and accurate creation of DTM (Digital Terrain Model) or DSM (Digital Surface Model), as well as the extraction of buildings, roads or other objects, proving the undoubted advantages of ALS. Of particular importance is automatic buildings' extraction, as it is the basis for 2 and 3-D models development used for spatial planning, land management or cadastre update (Kulesza 2007). However, the development of efficient methods for automatic building detection from ALS data remains a challenge due to many factors related to remote sensing data (point cloud density or spatial and spectral variability of the images) as well as the complexity of urban development (Wierzbicki et al. 2021, Gilani et al. 2016) found that small, shaded or partially occluded buildings were generally misclassified. This problem was successfully solved by the authors thanks to the fusion of the point cloud and image data. The achieved correctness above 95% proved the reliability of the approach. The accuracy of building assessment extracted from LiDAR data is complicated due to the zigzag shape of the building outline, devoid of many details and hardly comparable with reference building's data. Moreover, building footprints and outlines of rooftop shapes differ both in shape and location. These differences, albeit slight, make it necessary to correct the geometric building outlines in order to locate the building accurately. Dey et al. (2020) overcame this problem by introducing a new, robust corner correspondence (RCC) metric that al-lowed to assess the extra- and under-lap areas of extracted and reference buildings. The RCC metric constitutes a combined measure of the positional accuracy and shape similarity and allows for a more realistic assessment of the extracted building boundaries from LiDAR data. Nguyen et al. (2020) whereas introduced the Super-Resolution-based Snake Model (SRSM) that operates on high-resolution LiDAR data that involved a balloon force model to extract buildings. Lai et al. (2019), who noticed that the accuracy of building segmentation based on ENVI algorithms ranged from 80 to 90%. While, regularization technique using LiDAR data and orthoimages proposed by Gilani et al. (2016) obtained completeness from 83% to 93% for the area with a correctness of above 95%.

The literature showed that LIDAR data allows extraction of building outlines and construction of 2 or 3D models, thus facilitating automatic cadastre updating. Based on cadastral laws and previous practical and scientific achievements, it was hypothesized that buildings automatically extracted from ALS data can only be used for modernization of building records. Hence, the aim of the study was automatic extraction of buildings from dense ALS data for modernization of cadastre of Poland. The remainder of this paper is structured as follows. Section 2 describes the building register as part of the cadastre, Section 3 defines the study area, the data used and the methods employed. Sections 4 and 5 present the results and conclusions.

2. POLISH CADASTRE - BUILDING REGISTER

The basic legal act for all geodetic and cartographic issues is the *Geodetic and Cartographic Law* (Law, 2021) and the implementing provisions listed in the Minister's Regulation on *Land and building registry* (EGiB Regulation 2021). The amendment to the *Geodetic and Cartographic Law* and the Ministry Regulation of 2021 introduced comprehensive administrative procedures for updating, verifying and modernizing the cadastre to ensure data accuracy and reliability. The database structure followed the INSPIRE Directive and Land Administration Domain Model ISO standard. A building, as a geographic object type, adopts the <<Feature type>> stereotype and is described by 12 attributes, including the geometry as a polygon (GM_Multisurface) as it is seen in Figure 1.

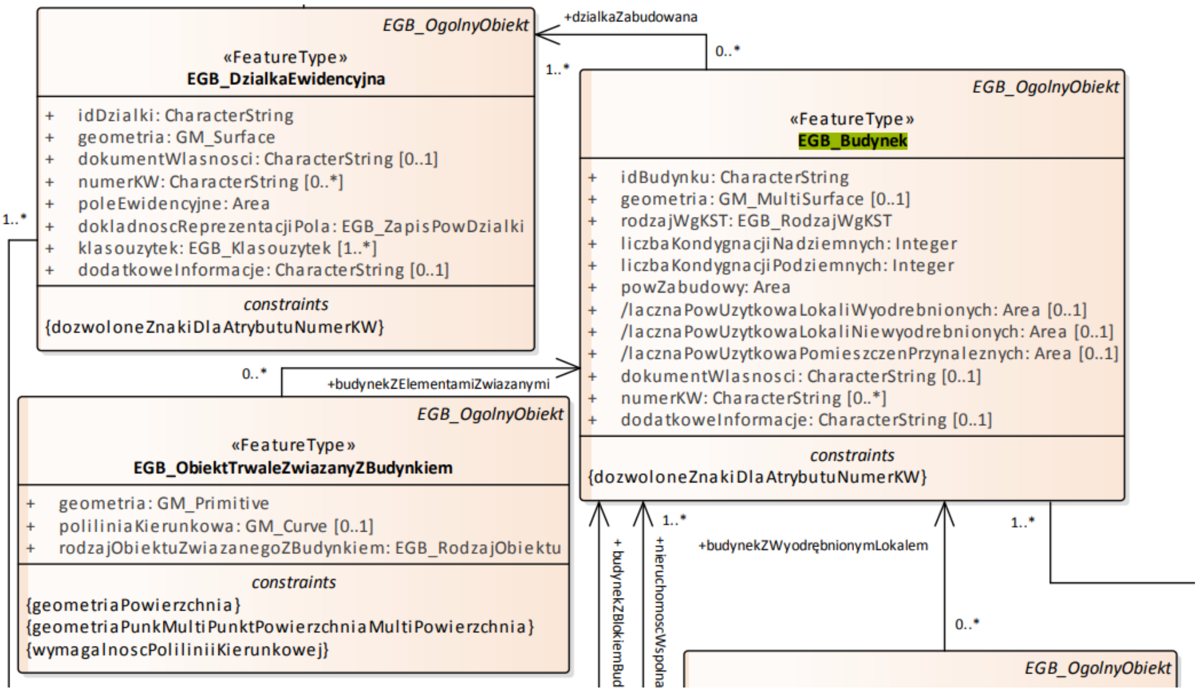


Fig. 1. Land and Building Registry conceptual schema, ULM diagram (only selected part).

The attribute "powZabudowy: Area" writes the area defined by the outline of the above-ground part of the building. Due to the complexity of the provisions of the EGiB Regulation and the multitude of building’s shapes determination its value, caused many problems. Buśko (2016, 2017) and Hanus et al. (2017) gave many examples of such difficulties, including buildings with an overhang without columns (Fig. 2a) and an overhang supported by columns (Fig. 2b)

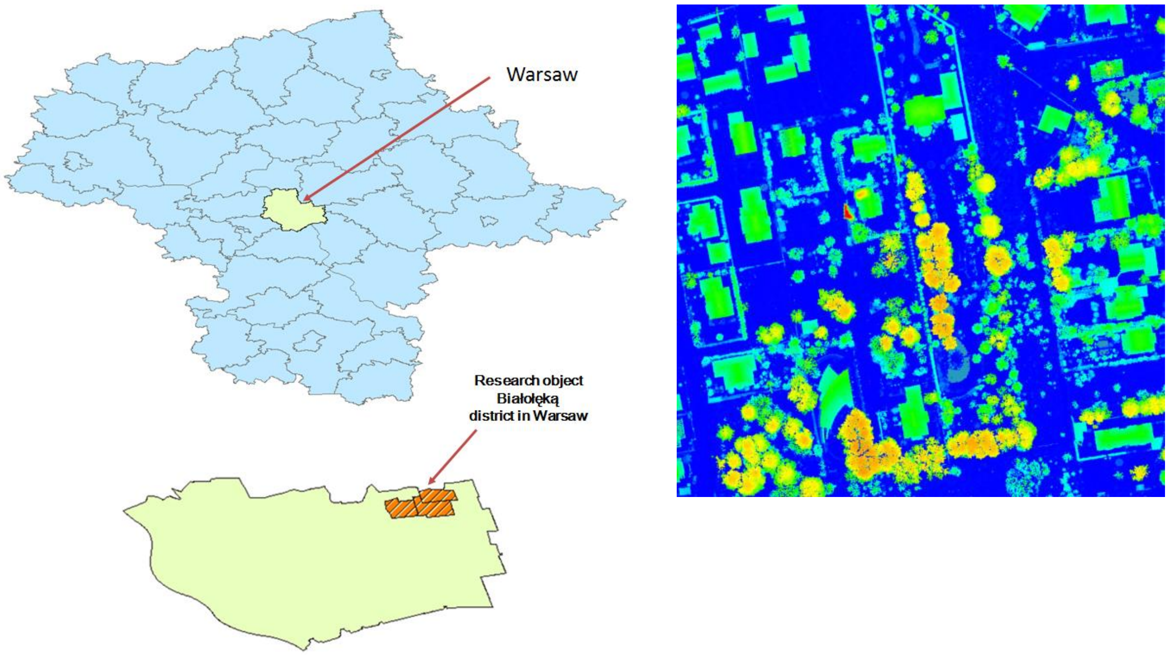
Fig. 2. Buildings with an overhang: a - without columns; b- supported by columns.
(Source: Buśko M. 2016. Analysis of the influence of amendments to the legal provisions relating to building structures on keeping the real estate cadastral database updated (in Polish). *Infrastruktura I Ekologia Terenów Wiejskich,* 395–410)

Current legal regulations allow the use of modern photogrammetric techniques and remote sensing data to update cadastral data on buildings, provided that the positioning accuracy of the building ground contour (outline) is not less than 0.10 m (in relation to the points of the geodetic network points). However, practitioners and scientists still note many inconsistencies between cadastral building data and field survey data, resulting in limited reliability of building data, with positional accuracy ranging from 0.7 to 1.5 m (Hanus et al. 2017, Mika 2018).

The implementation of remote sensing in cadastre has developed significantly since the 1990s due to technological advances, in particular image acquisition capabilities, processing algorithms, internet and mobile technologies. Currently, remote sensing data available from multiple sensors (ground, aerial and satellite) are widely used by public administration for many tasks, in particular planning, spatial management and multi-purpose cadastre. However, it is worth mentioning that photogrammetric data and methods used for cadastral purposes have long history in Poland. The first book describing the possibilities of using photogrammetry in cadastral measurements by Prof. Edmund Wilczkiewicz was published in 1925 (Bryś 2021). This book aimed to promote aerial photogrammetry by describing aerial photographs and techniques of their processing for a wide range of measurement tasks, in particular cadastre creation and updating.

3. STUDY AREA AND DATA

The research area covered 800 ha of Kobiałka, a peripheral part of Warsaw (Baiłołęka district), stretching in the east from 21°02′52 ″.2 to 21°03′44 ″.7, and in the north from 52°03′44 ″.7 to 52°21′55 ″.6 (Figure 3a). It is a typical residential area, dominated by detached buildings, roads, and green infrastructure. This area is characterized by large and in many cases uncontrolled growth of buildings due to favorable housing conditions, good communication with the center of Warsaw, and proximity to forest.



(a) (b)

Fig. 3. Study area location (a), LIDAR data (b).

LIDAR dense classified point clouds (12 points per 1 m^2) stored in LAS binary format were employed for building vectors extraction (Figure 3b). This is an open data available via Polish Spatial Infrastructure Geoportal maintained by the Head Office of Geodesy and Cartography the National Mapping Agency (Izdebski 2020; Izdebski et al. 2021; Bielecka et al. 2018). Buildings data from the Warsaw cadastre was used for training and final evaluation of building extraction. The data comprised the building geographical location, number of storeys, source of geometry and building type (e.g., residential, outbuilding, office).

Data processing deploying deep learning methods was performed in the ENVI software, developed by Harris Geospatial Solutions, Inc., while geo-processing and final evaluation of remotely extracted building were done with ArcGIS software, provided by ESRI.

4. METHOD APPLIED

The automated building extraction consisted of two main steps: building roof extraction and geoprocessing and quality control. The ENVI LIDAR algorithm recognized the position, aspect and slope of roof planes in the area in question as well as extracted consistent and geometrically correct building models (Africani et al. 2013). Buildings, trees, and other recorded objects were separated based on size, height, and shape characteristics. Building rooftop patches were extracted, after extensive tests, by applying a threshold to the following parameter: minimum building area – 10 m^2, Near Ground Filter Width – 5 m, Buildings Points Range 1.5 m, and Plane Surface Tolerance (PST) – 0.5 m. PST recognized curved roofs based on a series of successive planes and delineated a new roof plane when the distance between the analysed and previous points in the cloud reached the declared value (Harris Geospatial Solutions 2020). PST value was set-up after series of tests using 0.15 m, 0.30 and 0.5 m values. The best results were obtained for the PST of 0.5 m, achieving the completeness of 95% compared to the buildings from cadastral data.

Geoprocessing of building roof outlines consisted of geometrically fitting adjacent roof facets and creation a polygon of the building outline. The polygons were further simplified by the Douglas-Peucker algorithm. Quality control (QA) consisted in comparing cadastral and extracted building's outline and calculating mean displacement, standard deviation, relative standard deviation (RSD), coefficient of variance to the mean (VMR).

5. RESULTS AND DISCUSSION

Kobiałka, as a residential district located on the outskirts of Warsaw, is characterized by diverse building architecture, which had a large impact on the accuracy of building extraction. In particular, the multi-sloped roofs, dormers and the accompanying vegetation partially covering the roofs (trees and high shrubs) hindered the performance of the extraction algorithm and required further geoprocessing (see Figure 4).

Fig.4. Multi-slope roof patches delineated by ENVI algorithm

Figure 5 shows correctly extracted single-family buildings, while Figure 6 demonstrates errors in determining building contours from ALS data. Red points indicate building corners derived from the cadastre, while green points show corners determined by the ENVI algorithm from the ALS data.

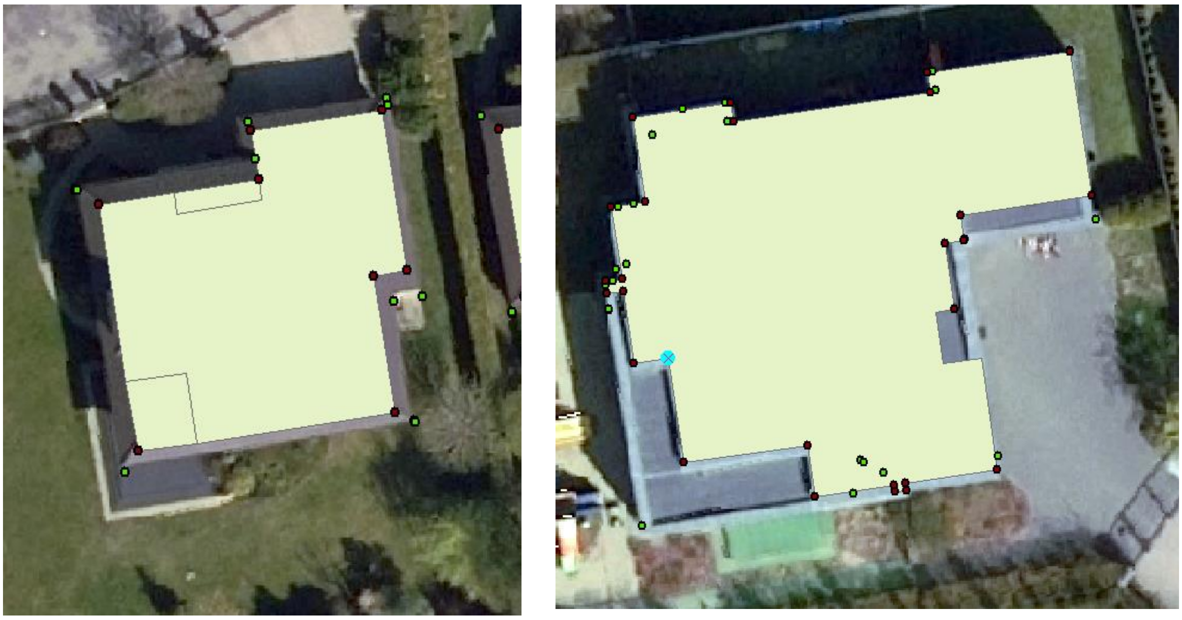


Fig. 5. Correctly delineated outline of the building's roof with a clear offset from the ground survey points marking the ground level.

a b

Fig. 6. Errors in building extraction: a - omitted part of the building with roof partially obscured by vegetation; b – a red paved terrace erroneously included as the building roof outline.

Figure 6 indicates buildings that should be checked during cadastre modernization. Without field verification, it is difficult to determine whether the two buildings were recorded correctly in the cadastral building registry. Particular doubts are raised by the example shown on the right, where only one parking lot is visible and the distance between the buildings is very short.

Fig.6. Errors between data in the building registry and buildings extracted from LIDAR data.

After statistical analysis, an average offset (distance of corresponding building corners) of 1.18 m (with STD equal to 0.688) was found between building outlines and cadastral data. The dispersion index (VMR) amounted to 0.4, indicating a binomial distribution (under dispersion). Less than 5.6% of the building contour points (190 of 3406) were considered outliers because the near distances between corresponding edges in the compared datasets were greater than 2.44 m. On the other hand, the shape of the buildings recorded in the cadastre and extracted from the LIDAR data were very similar, with the shape coefficient took an average of 0.78 and 0.74, respectively. There was also little variation in the shape coefficient, with a VMR value of 0.01 for the cadastral data and 0.02 for the LAS.

The number of buildings in the cadastre and extracted from LAS data differed for several reasons. One of them was the provisions of the Geodetic and Cartographic Law, namely Article 2(7b) stated that cadastre building register does not store buildings that do not require a geodetic as-built inventory e.g. designed buildings and buildings under construction. The second one - errors in cadastral data and delays in updating, and the third one - inaccuracies in extracting buildings from the LAS point cloud.

CONCLUSIONS

Open access to dense LAS data and advancement in remotely sensed data processing, offer fast and cost-effective ways of extracting buildings for spatial database updating and cadastral purposes. The displacement of the building outline by 1.18 m on average indicates that the buildings extracted on the basis of the dense LIDAR point cloud cannot be directly upload to the cadastre, but they are a valuable source for its modernization. This means that the research hypothesis has been proved. Summarizing the discussion on the possibility of updating the cadastre with remote sensing data, in particular with dense LIDAR data, it should be noted that deep learning for building extraction is a promising technology. However, the final registration of automatically excavated buildings in the Polish cadastral system yet requires more accurate field measurements in accordance with national cadastral standards.

ACKNOWLEDGMENTS

Data used in this study, i.e.: buildings and LAS data, were derived from the national geoportal of Poland https://mapy.geoportal.gov.pl/imap/Imgp_2.html, maintained by the Head Office of Geodesy and Cartography. The study was conducted under financial support by Military University of Technology, Faculty of Civil Engineering and Geodesy, grant number 795/2022.

REFERENCES

Africani P., Bitelli G., Lambertini A., Minghetti A., Paselli E. 3013. Integration of LIDAR data into amunicipal GIS to study solar radiation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*. XL–1-W1-1, 1–6, doi:10.5194/isprsarchives-XL-1-W1-1-2013.

Bielecka E., Janczr E., Dukaczewski D. 2018. Spatial Data Infrastructure in Poland – lessons learnt from so far achievements. Geodesy and Cartography 67(1), 3–20, DOI: https://doi.org/10.24425/118702

Bryś H. 2021. The first Polish publication on the use of aerial photogrammetry for cadastral purposes by prof. dr hab. inż. Edmund Wilczkiewicz from 1925. *Przegląd geodezyjny* 11, 10-13, doi: 10.15199/50.2021.11.1

Buśko M. 2016. Analysis of the influence of amendments to the legal provisions relating to building structures on keeping the real estate cadastral database updated (in Polish). Infrastruktura i Ekologia Terenów Wiejskich, 395–410.

Buśko, M. 2017. Modernization of the Register of Land and Buildings with Reference to Entering Buildings into the Real Estate Cadastre in Poland. Environmental Engineering. In Proceedings of the International Conference on Environmental Engineering. Vilnius Gediminas Technical University, Vilnius, Lithuania, 27–28 April 2017, ICEE; Vilnius, t. 10, (2017), doi:10.3846/enviro.2017.176.

Dey E.K., Awrangjeb M. 2020. A Robust Performance Evaluation Metric for Extracted Building Boundaries from Remote Sensing Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*. 13, 4030–4043.

EGiB Regulation. 2021. Regulation of the Minister of Economic Development and Technology of 27 July 2021 for Land and Building Registry. Official Journal of Laws (Dz.U. 2021 poz. 1390), Warsaw, Poland.

Gilani S.A.N., Awrangjeb M., Lu G. 2016. An Automatic Building Extraction and Regularisation Technique Using LiDAR Point Cloud Data and Orthoimage. *Remote Sens.* 8, 258, doi:10.3390/rs8030258

Hanus P., Benduch P., Penska-Siwik A. 2017. Budynek na mapie ewidencyjnej, kontur budynku i bloki budynku. *Przeglad Geodezyjny* 89(7), 615-20, DOI: 10.15199/50.2017.7.1

Harris Geospatial Solutions. 2020. Train Deep Learning Models. Available online: https://www.l3harrisgeospatial.com/docs/TrainDeepLearningModels.html (accessed on 7 February 2022).

Izdebski W. 2020. Infrastruktura Danych Przestrzennych w Polsce. [Spatial data Infrastructure in Poland]. Geo-System Sp. z o.o., ISBN 978-83-943086-4-3.

Izdebski W., Zwirowicz-Rutkowska A., Nowak da Costa J. 2021. Open data in spatial data infrastructure: the practices and experiences of Poland. *International Journal of Digital Earth*, 14(11), 1547-1560, DOI: 10.1080/17538947.2021.1952323

Kulesza Ł. 2007. Automatic building detection and modelling with the Terrascan software (in Polish). Archiwum Fotogrametrii, Kartografii i Teledetekcji, Vol. 17a, 415-424.

Lai X., Yang J., Li Y., Wang, M. 2019. A Building Extraction Approach Based on the Fusion of LiDAR Point Cloud and Elevation Map Texture Features. *Remote Sens.* 11, 1636, doi:10.3390/rs11141636

Law. 2021. Geodetic and Cartographic Law. Official Journal 2021 item 1990, Warsaw, Poland.

Mika M. 2018. An analysis of possibilities for the establishment of a multipurpose and multidimensional cadastre in Poland. *Land Use Policy* 77, 446–453, doi:10.1016/j.landusepol.2018.05.060.

Nguyen T.H., Daniel S., Guériot D., Sintès C., Le Caillec J.-M. 2020. Super-Resolution-Based Snake Model—An Unsupervised Method for Large-Scale Building Extraction using Airborne LiDAR Data and Optical Image. *Remote Sens*. 12, 1702, doi:10.3390/rs12111702.

Wierzbicki D., Matuk O., Bielecka E. 2021. Polish Cadastre Modernization with Remotely Extracted Buildings from High-Resolution Aerial Orthoimagery and Airborne LiDAR. *Remote Sensing* 13(4):611. https://doi.org/10.3390/rs13040611

BIOGRAPHICAL NOTES

Prof. Damian Wierzbicki, received the Ph.D degree in photogrammetry and remote sensing from Military University of Technology, Warsaw, Poland, in 2015. He is currently an Associate Professor with the Department of Imagery Intelligence, Faculty of Civil Engineering and Geodesy, Military University of Technology His research interests include UAV navigation and image processing, deep learning in remote sensing. His research interests also include development of new algorithms for object detection and classification in image sequences from UAV.

Prof. Elzbieta Bielecka, full professor at the Faculty of Civil Engineering and Geodesy at the Military University of Technology. Scientific research focuses on the theoretical issues of GIS, especially data quality and uncertainty, spatial modeling as well as national SDI.

Dr eng. Olga Matuk, obtained her master's and doctoral degrees at the Faculty of Geodesy and Cartography of the Warsaw University of Technology. Research issues relate to cadastre and real estate management.

CONTACTS

Dmian Wierznicki
Military University of Technology
Faculty of Civil Engineering and Geodesy
2 gen. Sylwestra Kaliskiego St. Warsaw
POLAND
Email: damian.wierzbicki@wat.edu.pl
Elżbieta Bielecka
Military University of Technology
Faculty of Civil Engineering and Geodesy
2 gen. Sylwestra Kaliskiego St. Warsaw
POLAND
Email: elzbieta.bielecka@wat.edu.pl
Olga Matuk
Military University of Technology
Faculty of Civil Engineering and Geodesy
2 gen. Sylwestra Kaliskiego St., Warsaw
POLAND
Email: olga.matuk@wat.edu.pl