

Unsupervised Roofline Extraction from True Orthophotos for LoD2 Building Model Reconstruction

Weixiao Gao¹ and Ravi Peters² Jantien Stoter¹

¹ Dept. Urbanism, Delft University of Technology, The Netherlands,
(w.gao-1, j.e.stoter)@tudelft.nl,

² 3DGI, Zoetermeer, The Netherlands,
ravi.peters@3dgi.nl

Abstract. This paper discusses the reconstruction of LoD2 building models from 2D and 3D data for large-scale urban environments. Traditional methods involve the use of LiDAR point clouds, but due to high costs and long intervals associated with acquiring such data for rapidly developing areas, researchers have started exploring the use of point clouds generated from (oblique) aerial images. However, using such point clouds for traditional plane detection-based methods can result in significant errors and introduce noise into the reconstructed building models. To address this, this paper presents a method for extracting rooflines from true orthophotos using line detection for the reconstruction of building models at the LoD2 level. The approach is able to extract relatively complete rooflines without the need for pre-labeled training data or pre-trained models. These lines can directly be used in the LoD2 building model reconstruction process. The method is superior to existing plane detection-based methods and state-of-the-art deep learning methods in terms of the accuracy and completeness of the reconstructed building. Our source code will be released when the paper is accepted.

Keywords: building rooflines extraction, 3D building models, true orthophotos

1 Introduction

With the recent advancements in computer vision and photogrammetry technology, the acquisition of both 2D and 3D data for large-scale urban environments has become feasible. This has led to the derivation of urban 2D data from street view, aerial, and satellite images, while 3D data is derived from LiDAR point clouds, dense image matching generated point clouds, and textured meshes. Building models in 3D urban scenes are essential for various applications such as solar irradiation [1, 2], photovoltaic analysis [3, 4], building illumination analysis [5], and urban planning [6, 7]. However, for large-scale urban scene analysis, it is often necessary to find the optimal balance between simplicity and preserving essential, detailed geometric features. To this end, LoD2 building models, conforming to the CityGML standard [8], have gained popularity.

The automatic reconstruction of LoD2 building models from 2D and 3D data has been a topic of active research. Traditionally, point clouds obtained from airborne LiDAR have been used to generate such models [9]. However, due to the high costs and long intervals associated with acquiring and collecting LiDAR point clouds for large-scale urban scenes, it can be challenging to capture rapidly developing and changing areas in cities. Consequently, there may be missing data for these regions, which makes it difficult to update the models accurately. To address this issue, researchers have started exploring the use of point clouds generated from aerial images with the dense image matching pipelines for automatic LoD2 building model reconstruction [10]. Dense image matching point clouds offer several advantages over LiDAR point clouds, including being easily obtainable, cost-effective, and having color information. Furthermore, they can be frequently updated to reflect changes in the urban environment, making them more suitable for vector building reconstruction tasks in large-scale, high-frequency urban scenes where the data must be updated frequently to reflect urban development. As a result, dense image matching point clouds have become an area of interest for researchers studying automated LoD2 building model reconstruction.

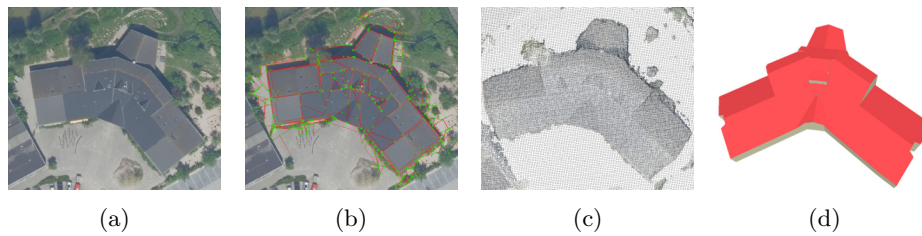


Fig. 1: The workflow of our method. First, we use a true orthophoto (a) as input. Next, we perform line extraction (b) to partition the building footprint, which generates separate roof parts. We then utilize a dense point cloud (c) to extrude the partition results and reconstruct a LoD2 building model (d).

Generating dense point clouds from aerial oblique photogrammetry images can result in significant errors and introduce noise into the generated point clouds, compared to high-precision LiDAR point clouds. This can make originally planar regions, such as roof planes of buildings, become uneven, which creates challenges for reconstructing LoD2 building models. Traditional point cloud plane detection-based methods [11, 12] can also exacerbate the issue caused by dense image matching point clouds by increasing complexity and reducing accuracy in the reconstructed models. However, the original images used to generate these point clouds and the synthesized orthoimagery typically contain less noise and fewer errors, providing an opportunity to extract roof structures directly from the images and combine them with height information from the point

clouds extracted from the same source images for reconstructing high-precision LoD2 building models.

In recent years, more and more research has focused on how to use deep learning methods to extract roof structures from images [13–18]. Although these methods have performed well on public datasets [14], they still require a large amount of manually labeled training data and cannot guarantee good generalization ability. More importantly, they cannot ensure the geometric completeness of the extracted roof structures, especially for complex buildings.

This paper presents a method for extracting rooflines using line detection, which can be used for the reconstruction of building models at the Lod2 level Figure 1. As our results will show, our approach has the advantage of being able to extract relatively complete roof lines without the need for any pre-labeled training data or pre-trained models. Moreover, our method surpasses both traditional plane detection-based methods and state-of-the-art deep learning methods based on transformers in terms of both the accuracy and completeness of the reconstructed building. Our method is applicable to large-scale urban scenes and can be used for extracting roof structures of complex buildings.

2 Recent Advances in Roofline Extraction

In recent years, deep learning methods have received increasing attention for extracting roof structures from images using neural networks. For instance, Fatemeh Alidoost et al. (2019) [13] proposed a CNN-based approach for 3D building reconstruction from a single 2D image, which achieved accurate height prediction, roofline segmentation, and building boundary extraction. However, the accuracy of the method degraded for test data with different spatial-spectral characteristics and complicated buildings. Another pioneering work by Nelson Nauata et al. (2020) [14] used CNNs to detect geometric primitives and infer their relationships, fusing all information into a planar graph through holistic geometric reasoning for reconstructing a building architecture. This work made significant improvements over the existing state-of-the-art but has limitations in handling missed corners, curved buildings, and weak image signals.

Furthermore, Fuyang Zhang et al. (2020) [15] proposed Conv-MPN, a message passing neural architecture, for structured outdoor architecture reconstruction, which achieved significant performance improvements over existing prior-free solutions but has the drawback of extensive memory consumption and limitations on the number of corner candidates. They further improved the method by presenting a novel explore-and-classify framework [16] for structured outdoor architecture reconstruction. The method learns to classify the correctness of primitives while exploring the space of reconstructions via heuristic actions, and demonstrated significant improvements over all state-of-the-art methods, with a few limitations related to slow test-time inference and corner detection failures in extreme cases. Another similar work is presented by Wufan Zhao et al. (2022) [17], who proposed RSGNN, an end-to-end learning framework for planar roof structure extraction in polygon format from VHR remote sensing

images, which combines primitive detectors and GNN-based relationship inference and shows superior results in both qualitative and quantitative evaluations. However, the method has limitations in handling complex roof structures and labeling accuracy of reference data.

A recent work HEAT is proposed by Jiacheng Chen et al. (2022) [18], which is an attention-based neural network for reconstructing a planar graph from 2D raster images. It uses transformer decoders and iterative inference at test time but still faces challenges in missing corners and rare structures. Despite their advantages, deep learning methods require a large amount of training data with manual labels and may not ensure good generalization and geometric completeness for extracting roof structures, especially for complex buildings. Even a single missing roof line can result in significant geometric errors in the reconstructed building models.

3 Methodology

Our method involves three main steps: 1) building image cropping; 2) extracting line segments from the building images; and 3) reconstructing LoD2 building models.

3.1 Building image cropping

Due to the large size of the original true orthophoto and the presence of other objects in addition to buildings, direct line detection would extract many redundant line segments. Therefore, our first objective is to extract a single building image for each building from the true orthophoto.

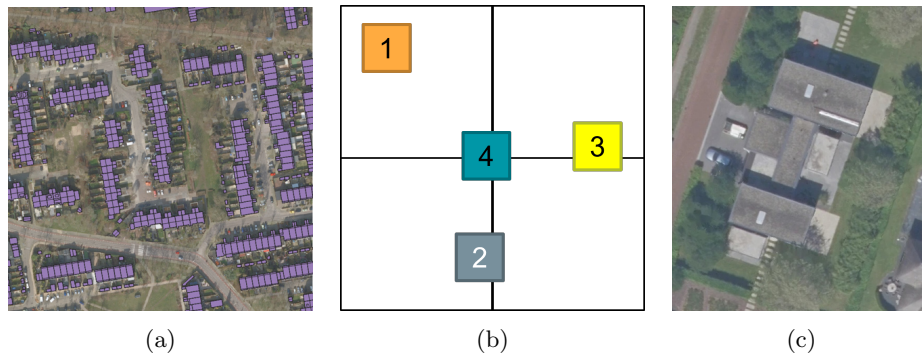


Fig. 2: The workflow of building image cropping. (a) Building footprints on the map. (b) Four types of correspondences that can exist between a bounding box of a building and true orthophotos. (c) Cropped building image.

We use GIS data, specifically the building footprints on the map (see Figure 2a), to assist with the extraction process. First, we merge adjacent buildings according to their footprints. Next, we establish a bounding box based on the footprint of a single building and determine the corresponding region in the true orthophotos by using the coordinates of the four corners of the bounding box (as illustrated in Figure 2b). The geographic coordinates of the top-left corner of the true orthophoto are already known. The correspondences between a specific bounding box and orthophotos can be divided into four cases: 1) a single image, 2) two images joined horizontally, 3) two images joined vertically, and 4) four images joined both horizontally and vertically. In the final step, we use the building boundary boxes to crop and stitch the corresponding images, generating a single composite image for each building (see Figure 2b). These images will serve as input data for the next stage of line segment detection.

3.2 Line Extraction

Our objective is to detect line segments in the single building image obtained in the previous step and convert the detection results from image coordinates to geographic coordinates. We use the KIPPI [19] algorithm as our primary tool for line detection. Firstly, it employs a line segment detector to identify line segments in the image. These detected line segments are then globally regularized based on geometric properties like parallelism, perpendicularity, and collinearity. Finally, the line segments are bidirectionally extended using kinetic data structures to segment the image. As shown in Figure 3b, rooflines represent a subset of these segmented lines.

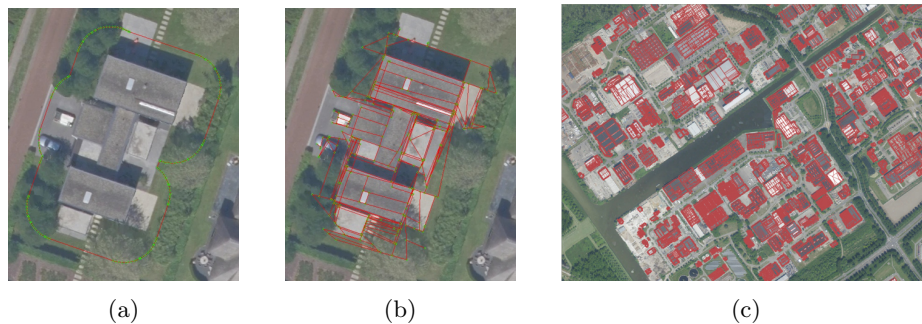


Fig. 3: Line extraction workflow on the single building image. (a) Buffered building footprints. (b) Cropped line segments. (c) Georeferenced rooflines.

However, the single building image may include redundant information such as non-building features surrounding the building. To filter out such information, we use a buffer (of 60 pixels in all experiments) generated from the building footprint to refine the segmented lines (see Figure 3a). Next, we convert

the coordinates of each node of every line segment in the image to geographic coordinates using Equation 1 as follows:

$$\begin{aligned} g_x &= t_x + (p_x + b_x) \times s \\ g_y &= t_y - (p_y + b_y) \times s \end{aligned} \quad (1)$$

where g_x and g_y represent the geographic coordinates of the line node, while p_x and p_y represent its pixel coordinates. The t_x and t_y denote the geographic coordinates of the top left corner of the corresponding true orthophoto. The s represents the pixel size. The results provide geometric information for the next step of building reconstruction (see Figure 3c).

3.3 LoD2 building Reconstruction

In this step, our goal is to reconstruct LoD2 building models using the set of line segments extracted in the previous step, the building footprint, and the point cloud generated by the dense image matching pipeline from the same source images. Our approach is based on the method proposed by Ravi Peters et al. (2022) [9], with the main difference being the replacement of the line segments generated through point cloud detection of planes with the set of line segments extracted from the image. We then use the set of line segments extracted from the image to subdivide the building footprint and obtain the roof plane structure. Finally, based on the height information provided by the point cloud, we extrude the roof parts to generate a watertight and 2D-manifold solid building model in LoD2.

4 Experimental Results

4.1 Dataset

We utilized an orthophoto dataset with a resolution of 8cm [20], covering three cities including Almere, Den Haag, and Leiden. The building footprint data was obtained from the Building and Address Register of the Netherlands (BAG). The dense point cloud was generated using oblique aerial images with Nframes SURE software by a workflow designed and executed by the Dutch Kadaster. Our experimental setup consisted of a computer equipped with an AMD Ryzen Threadripper 1920X 12-Core Processor and 32GB of memory.

4.2 Evaluation and Comparison

Our method for line extraction was tested on three cities in the Netherlands, with extraction times of 1.62 hours for 27,395 buildings in Almere, 3.22 hours for 43,520 buildings in Den Haag, and 2.85 hours for 53,502 buildings in Leiden. The time required for line extraction in these cities is mainly influenced by the number of buildings and the complexity of their external geometric structures.

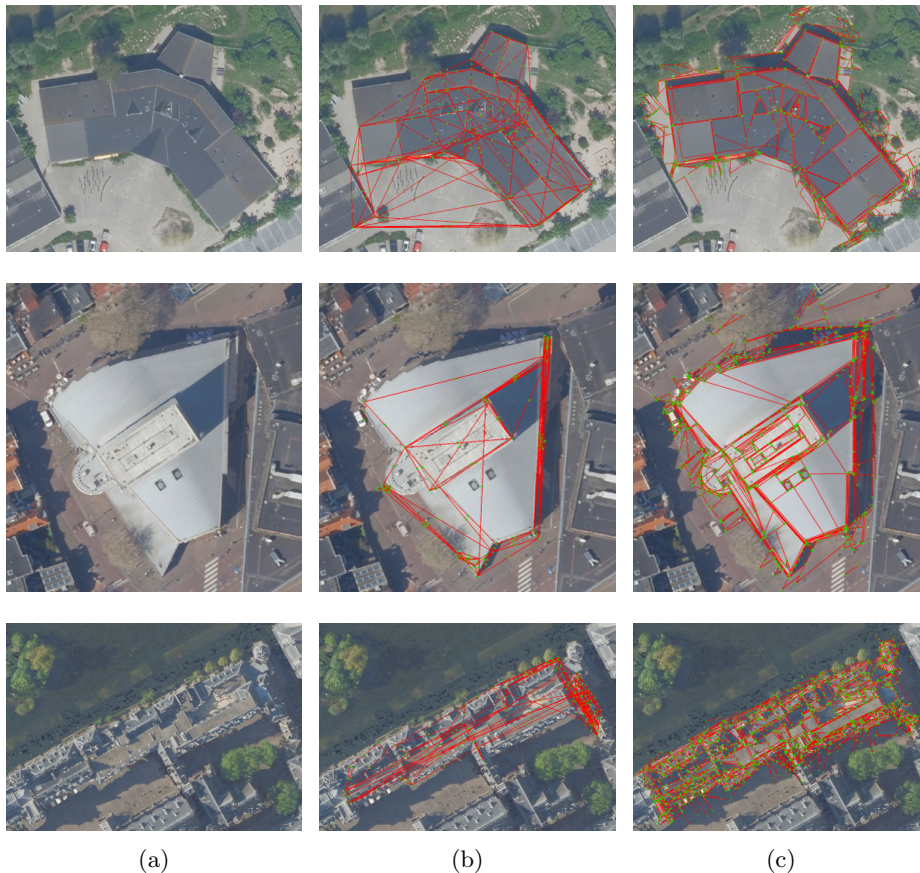


Fig. 4: Roofline comparison. (a) Input building image. (b) HEAT results. (c) KIPPI results.

Evaluation of rooflines. We selected twelve representative buildings from the three cities and we projected the rooflines extracted from the 3D BAG building models [9] onto the orthophotos as a reference. We compared our roof line detection method (i.e., KIPPI method) with the HEAT method [18], which currently holds the top position in public benchmark datasets. To predict the roof structure of the generated single-building images (as shown in Figure 4), we used the trained model provided by the authors of HEAT [18]. We calculated the completeness of our method by measuring the percentage of extracted lines that overlapped with the reference lines (within 25 pixels offset in all experiments). Our method achieved around 90% overlap with the reference lines, while the HEAT method only achieved about 44%. This comparison clearly shows that our method produces more complete results, which is further evident in the subsequent 3D reconstruction outcomes. In addition, deep learning-based methods

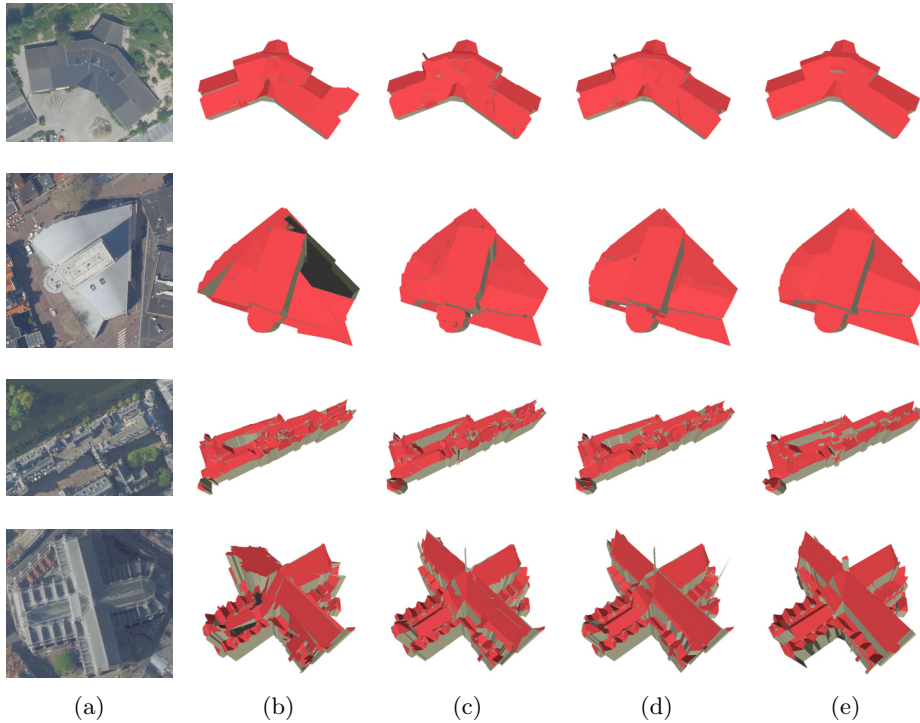


Fig. 5: Comparison of LoD2 Building Models: Input building images are shown in (a), while the reconstruction results using HEAT generated rooflines are shown in (b), plane detection method in (c), and our line segments in (d). Ground truth 3D BAG models reconstructed from LiDAR point clouds and plane detection are shown in (e).

require fixed-size image inputs due to limited GPU memory, and HEAT [18] limits the size of each building image to 256×256 . This limitation results in a significant loss of image resolution when scaling building images, especially for individual large and complex buildings (see Figure 4).

Evaluation of reconstructed buildings. We selected twelve buildings from the three cities that were previously identified and used them to reconstruct LoD2 models and evaluate our results. We used 3D BAG building models [9] reconstructed from LiDAR point clouds as the ground truth for evaluation. To compare the reconstructed building models, we utilized the same image-dense matching generated point clouds as input and applied three different approaches for roofline generation: 1) plane detection from the dense matching point cloud, 2) HEAT, and 3) our method (i.e., KIPPI). The evaluation criteria focused on model complexity, Hausdorff distance [21], and RMSE. We measured the complexity of the models based on the number of reconstructed faces and calculated

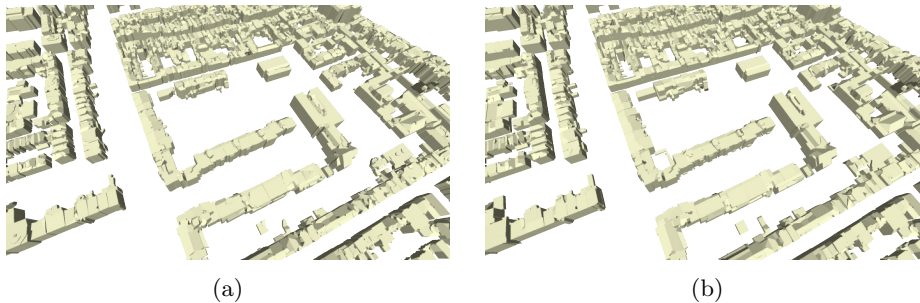


Fig. 6: Comparison of large-scale LoD2 building models reconstructed from extracted rooflines and image dense matching generated point clouds. (a) Building models reconstructed using rooflines extracted from plane detection. (b) Building models reconstructed using rooflines extracted from our method(i.e., KIPPI method).

the Hausdorff distance using sampled points from the building models. We computed the average Hausdorff distance and root mean square error (RMSE) of all the selected buildings in comparison to their respective ground truth models. It should be noted that if a model contains more faces with larger errors, then it is likely to have redundant faces. Conversely, models with smaller errors and more faces tend to capture more details.

	Faces	Mean	RMSE
3D BAG	1762.5	-	-
HEAT	3150.75	0.67	1.16
Planes	5660.08	0.49	0.67
KIPPI	4356.33	0.38	0.65

Table 1: A comprehensive quantitative evaluation of LoD2 building models using the averaged number of faces, mean Hausdorff distance, and RMSE metrics.

Table 1 and Figure 5 show the quantitative and qualitative results of the different methods, respectively. Our method produced results with higher accuracy and completeness than the HEAT method [18], which had the lowest complexity but the highest error due to missing line segment completeness. Furthermore, the reconstructed building models using our rooflines exhibited lower complexity and lower errors compared to the rooflines obtained from the plane detection method [9]. Figure 5d demonstrate that even when orthophoto images are somewhat blurry, our method can maintain high accuracy and low complexity, indicating its high robustness.

Figure 6 illustrates the outcomes of the LoD2 building model reconstruction for a large-scale urban scene. Notably, our method demonstrates its capacity to be seamlessly applied to such scenes, as depicted in Figure 6b. This suggests its potential as a viable alternative to image-based approaches for reconstructing the complete 3D BAG building model. Additionally, we observe that our method produces building models with enhanced geometric regularity compared to the plane detection-based method (see Figure 6b).

5 Conclusion

In conclusion, this paper presented a method for extracting rooflines using line detection, which is used for the improved reconstruction of building models at the LoD2 level from point clouds generated from images. The approach has the advantage of being able to extract relatively complete rooflines without the need for any pre-labeled training data or pre-trained models. Furthermore, the method surpassed both traditional plane detection-based methods and state-of-the-art deep learning methods based on transformers in terms of both the accuracy and completeness of the reconstructed building. The method is applicable to large-scale urban scenes and can be used for extracting roof structures of complex buildings. Although deep learning methods have received increasing attention for extracting roof structures from images, they still require a large amount of manually labeled training data and cannot ensure the geometric completeness of the extracted roof structures, especially for buildings with complex structures. The proposed method has the potential to overcome these limitations and improve the LoD2 model reconstruction from dense image matching point clouds which can then become a promising alternative for automated LoD2 building model reconstruction from LiDAR. Further research could be directed toward developing more accurate algorithms for line detection from low-resolution images, such as satellite imagery, to improve the reconstruction of building models also from those sources.

References

1. Biljecki, Filip, et al. "Propagation of positional error in 3D GIS: estimation of the solar irradiation of building roofs." *International Journal of Geographical Information Science* 29.12 (2015): 2269-2294.
2. Besuievsky, Gonzalo, Benoit Beckers, and Gustavo Patow. "Skyline-based geometric simplification for urban solar analysis." *Graphical Models* 95 (2018): 42-50.
3. Catita, Cristina, et al. "Extending solar potential analysis in buildings to vertical facades." *Computers & Geosciences* 66 (2014): 1-12.
4. Eerenstein, W., et al. "Tessera: maximizing PV yield performance with size flexibility for bipv." *30th European PV Solar Energy Conference and Exhibition*. 2015.
5. Saran, Sameer, et al. "CityGML at semantic level for urban energy conservation strategies." *Annals of GIS* 21.1 (2015): 27-41.
6. Chen, Ran. "The development of 3D city model and its applications in urban planning." *2011 19th International Conference on Geoinformatics*. IEEE, 2011.

7. Czyńska, Klara, and Paweł Rubinowicz. "Application of 3D virtual city models in urban analyses of tall buildings: today practice and future challenges." *Architecturae et Artibus* 6.1 (2014): 9-13.
8. Kolbe, T.H.; Kutzner, T.; Smyth, C.S.; Nagel, C.; Roensdorf, C.; Heazel, C. (Eds.) *OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard, Ver 3.0.0*. 2021.
9. Peters, Ravi, et al. "Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the Netherlands." *Photogrammetric Engineering & Remote Sensing* 88.3 (2022): 165-170.
10. Wang, Feng, et al. "Reconstruction of LoD-2 Building Models Guided by Façade Structures from Oblique Photogrammetric Point Cloud." *Remote Sensing* 15.2 (2023): 400.
11. Schnabel, Ruwen, Roland Wahl, and Reinhard Klein. "Efficient RANSAC for point-cloud shape detection." *Computer graphics forum*. Vol. 26. No. 2. Oxford, UK: Blackwell Publishing Ltd, 2007.
12. Lafarge, Florent, and Clément Mallet. "Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation." *International journal of computer vision* 99 (2012): 69-85.
13. Alidoost, Fatemeh, Hossein Arefi, and Federico Tombari. "2D image-to-3D model: Knowledge-based 3D building reconstruction (3DBR) using single aerial images and convolutional neural networks (CNNs)." *Remote Sensing* 11.19 (2019): 2219.
14. Nauata, Nelson, and Yasutaka Furukawa. "Vectorizing world buildings: Planar graph reconstruction by primitive detection and relationship inference." *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII* 16. Springer International Publishing, 2020.
15. Zhang, Fuyang, Nelson Nauata, and Yasutaka Furukawa. "Conv-mpn: Convolutional message passing neural network for structured outdoor architecture reconstruction." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
16. Zhang, Fuyang, et al. "Structured outdoor architecture reconstruction by exploration and classification." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
17. Zhao, Wufan, Claudio Persello, and Alfred Stein. "Extracting planar roof structures from very high resolution images using graph neural networks." *ISPRS Journal of Photogrammetry and Remote Sensing* 187 (2022): 34-45.
18. Chen, Jiacheng, Yiming Qian, and Yasutaka Furukawa. "HEAT: Holistic Edge Attention Transformer for Structured Reconstruction." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
19. Bauchet, Jean-Philippe, and Florent Lafarge. "Kippi: Kinetic polygonal partitioning of images." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
20. Dutch Kadaster. *Luchtfoto Actueel Ortho 8cm RGB*. <https://opendata.beeldmateriaal.nl/>, accessed: 2023-02-05.
21. Aspert, Nicolas, Diego Santa-Cruz, and Touradj Ebrahimi. "Mesh: Measuring errors between surfaces using the hausdorff distance." *Proceedings. IEEE international conference on multimedia and expo*. Vol. 1. IEEE, 2002.