TUM School of Engineering and Design
Technische Universität München

Enrichment of 3D building models by facade elements based on point clouds and confidence values

Olaf K. Wysocki

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz:
Prof. Dr.-Ing. Christoph Holst

Prüfer der Dissertation:
1. Prof. Dr. rer. nat. Thomas Kolbe
2. Prof. Dr.-Ing. habil. Ludwig Hoegner
3. Prof. Dr. tech. Sisi Zlatanova

Die Dissertation wurde am 03.01.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 07.10.2024 angenommen.

Abstract

Nowadays, the number of connected devices providing unstructured data in the form of images or point clouds is rapidly rising. These devices acquire unprecedented temporal and spatial resolution data, creating an influx of geoinformation that, however, lacks semantic information. Simultaneously, structured datasets such as semantic 3D building models are widely available and assure rich semantics and high global accuracy but are represented by relatively coarse geometries. The overarching theme of this research is: "How can we refine the existing 3D building models using unstructured data?"

In this thesis, the long-standing challenge of reconstructing semantic 3D building models at the level of detail (LoD) 3 is tackled. Unlike mesh-based models, they require watertight geometry and object-wise semantics at the façade level. The principal challenge of such demanding semantic 3D reconstruction is reliable façade-level semantic segmentation of 3D input data. In this thesis, three main contributions are presented: (i) enrichment strategy that accurately reconstructs semantic LoD3 building models by (ii) improving façade-level semantic 3D segmentation and (iii) fusing point clouds and 3D models.

To this end, laser physic traits and prior knowledge of 3D building models are leveraged to probabilistically identify model conflicts. These probabilistic physical conflicts propose locations of absent façade-objects in a model: Their final semantics and shapes are inferred in a Bayesian network fusing multimodal probabilistic maps of conflicts, 3D point clouds, and 2D images. To meet the demanding LoD3 requirements, the estimated shapes are used to cut openings in 3D building priors and fit semantic 3D objects from a library of façade objects. Additionally, an investigation is conducted to determine the accuracy improvement achievable by fusing point clouds with images.

Extensive experiments on the Ingolstadt and Munich datasets demonstrate the superior performance of the proposed strategy and its methods over the state-of-the-art methods in façade-level detection, semantic segmentation, and LoD3 building model reconstruction. The methods can foster the development of probability-driven semantic 3D reconstruction at LoD3 since not only the high-definition reconstruction but also reconstruction confidence becomes pivotal for various applications such as autonomous driving and urban simulations.

Kurzfassung

Heutzutage steigt die Zahl vernetzter Geräte, die unstrukturierte Daten in Form von Bildern oder Punktwolken bereitstellen, rasant an. Diese Geräte erfassen Daten mit beispielloser zeitlicher und räumlicher Auflösung und erzeugen so einen Zustrom von Geoinformationen, denen jedoch semantische Informationen fehlen. Gleichzeitig sind strukturierte Datensätze wie semantische 3D-Gebäudemodelle weit verbreitet und gewährleisten eine umfassende Semantik und hohe globale Genauigkeit, werden jedoch durch relativ grobe Geometrien dargestellt. Das übergeordnete Thema dieser Forschung lautet: „Wie können wir die vorhandenen 3D-Gebäudemodelle mithilfe unstrukturierter Daten verfeinern?“

In dieser Arbeit wird die seit langem bestehende Herausforderung der Rekonstruktion semantischer 3D-Gebäudemodelle auf der Detailebene (Eng. level of detail, LoD) 3 angegangen. Im Gegensatz zu netzbasierten Modellen erfordern sie eine wasserdichte Geometrie und objektbezogene Semantik auf Fassadenebene. Die größte Herausforderung einer derart anspruchsvollen semantischen 3D-Rekonstruktion ist die zuverlässige semantische Segmentierung der 3D-Eingabedaten auf Fassadenebene. In dieser Arbeit werden drei Hauptbeiträge vorgestellt: (i) eine Anreicherungsstrategie, die semantische LoD3-Gebäudemodelle genau rekonstruiert, indem (ii) die semantische 3D-Segmentierung auf Fassadenebene verbessert wird und (iii) Punktwolken und 3D-Modelle zusammengeführt werden.

Zu diesem Zweck werden physikalische Eigenschaften des Lasers und Vorkenntnisse über 3D- Gebäudemodelle genutzt, um Modellkonflikte probabilistisch zu identifizieren. Diese probabilistischen physikalischen Konflikte schlagen Standorte fehlender Fassadenobjekte in einem Modell vor: Ihre endgültige Semantik und Form wird in einem Bayes'schen Netzwerk abgeleitet, das multimodale Wahrscheinlichkeitskarten von Konflikten, 3D-Punktwolken und 2D-Bilder zusammenführt. Um die anspruchsvollen LoD3-Anforderungen zu erfüllen, werden die geschätzten Formen verwendet, um Öffnungen in 3D- Gebäudevorbauten zu schneiden und semantische 3D-Objekte aus einer Bibliothek von Fassadenobjekten anzupassen. Darüber hinaus wird eine Untersuchung durchgeführt, um die durch die Fusion von Punktwolken mit Bildern erreichbare Genauigkeitsverbesserung zu ermitteln.

Umfangreiche Experimente mit den Datensätzen von Ingolstadt und München zeigen die überlegene Leistung der vorgeschlagenen Strategie und ihrer Methoden gegenüber den aktuellen Methoden bei der Erkennung auf Fassadenebene, der semantischen Segmentierung und der Rekonstruktion von LoD3-Gebäudemodellen. Die Methoden können die Entwicklung einer wahrscheinlichkeitsgesteuerten semantischen 3D-Rekonstruktion am LoD3 fördern, da nicht nur die hochauflösende Rekonstruktion, sondern auch die Rekonstruktionssicherheit für verschiedene Anwendungen wie autonomes Fahren und Stadtsimulationen von entscheidender Bedeutung ist.

Abstrakt

W dzisiejszych czasach liczba podłączonych urządzeń pozyskujących nieustrukturyzowane dane w postaci zdjęć lub chmur punktów gwałtownie rośnie. Urządzenia te pozyskują dane o niespotykanej rozdzielczości czasowej i przestrzennej, powodując napływ geoinformacji, w której brakuje jednak informacji semantycznych. Jednocześnie ustrukturyzowane zbiory danych, takie jak semantyczne modele budynków 3D, są szeroko dostępne i zapewniają bogatą semantykę i wysoką globalną dokładność, ale są reprezentowane przez stosunkowo zgeneralizowaną geometrię. Nadrzędnym tematem tych badań jest: „Jak możemy udoskonalić istniejące modele budynków 3D przy użyciu nieustrukturyzowanych danych?”

W tej pracy podjęto wieloletnie wyzwanie, jakim jest rekonstrukcja semantycznych modeli budynków 3D na poziomie szczegółowości (LoD) 3. W przeciwieństwie do modeli opartych na siatce (ang. mesh) wymagają wodoszczelnej geometrii i semantyki obiektowej na poziomie fasady. Głównym wyzwaniem tak wymagającej semantycznej rekonstrukcji 3D jest niezawodna segmentacja semantyczna danych wejściowych 3D na poziomie fasady. Głównym wkładem tej pracy są trzy elementy: (i) strategia wzbogacania, która dokładnie rekonstruuje semantyczne modele budynków LoD3 poprzez (ii) poprawę semantycznej segmentacji 3D na poziomie fasady oraz (iii) fuzja chmur punktów i modeli 3D.

W tym celu wykorzystuję cechy fizyki lasera i istniejącą wiedzę na temat modeli budynków 3D, aby probabilistycznie zidentyfikować konflikty modeli. Te probabilistyczne konflikty fizyczne proponują lokalizacje nieobecnych obiektów fasadowych w modelu: Ich ostateczną semantykę i kształty wywnioskowano w sieci Bayesowskiej łączącej multimodalne probabilistyczne mapy konfliktów, chmury punktów 3D i zdjęcia 2D. Aby sprostać wymagającym wymaganiom LoD3, szacunkowe kształty służą do wycinania otworów w fasadach budynków 3D i dopasowywania semantycznych obiektów 3D z biblioteki obiektów fasadowych. Dodatkowo prowadzone jest badanie w celu określenia poprawy dokładności możliwej do osiągnięcia poprzez połączenie chmur punktów ze zdjęciami.

Szeroko zakrojone eksperymenty na zbiorach danych z miast Ingolstadt i Monachium wykazały wyższą skuteczność proponowanej strategii i jej metod w porównaniu z innymi metodami wykrywania na poziomie fasady, segmentacji semantycznej i rekonstrukcji modelu budynku LoD3. Metody te mogą wspierać rozwój semantycznej rekonstrukcji 3D opartej na prawdopodobieństwie w LoD3, ponieważ nie tylko rekonstrukcja w wysokiej rozdzielczości, ale także pewność rekonstrukcji staje się kluczowa dla różnych zastosowań, takich jak samochody autonomiczne czy symulacje miejskie.

List of Figures

List of Tables

Glossary

ALS airborne laser scanning. 33, 94, 100

BayNet Bayesian Network. xv, xvii, 16, 17, 37, 40, 41, 49, 51, 73, 74, 93

BIM Building Information Modelling. xiii, 8, 9, 13, 21–23

BoW Bag-of-Words. 60–62, 84

BRIEF Binary Robust Independent Features. 61

CAD computer-aided design. xvi, xvii, 8, 60, 61, 70, 86, 87

CC Conflict Classification method. xvii, 78, 81, 83, 84, 86, 94, 95

CI confidence interval. xv, xvii, 12, 13, 37, 39, 40, 42, 45, 46, 73, 75, 93

CityGML CityGML. xiii–xix, xxiv, 7–9, 13, 21, 29, 43, 44, 52, 58, 66, 67, 70, 73, 75, 85, 92, 96, 100, 102, 103

CL confidence level. 45, 46

CNN convolutional neural network. xiv, 25–27

CPT Conditional Probability Table. xiii, xv, 16, 17, 40, 41, 49, 51, 54, 73

CRS coordinate reference system. 2, 8, 44

CSG Constructive Solid Geometry. xvi, 59, 62–65, 96

DL deep learning. 28, 47, 77

DST Dempster-Shafer theory. 13, 19, 22

GIS Geographic Information System. 9, 17, 44, 93

GNSS global navigation satellite system. 2, 10, 11

HD high definition. 11, 70, 93

HOG Histogram of Oriented Gradients. 61, 62, 86–88

ICP Iterative Closest Point. 11, 18, 39, 73

IFC Industry Foundation Classes. 43

IoU intersection over union. 98

LiDAR light detection and ranging. 1, 34

LoA Level of Accuracy. xiii, 13, 22

LoD Level of Detail. xiii, xv–xix, 1–3, 7, 8, 10, 13, 35–37, 39, 43, 45, 57, 58, 62, 67, 70, 71, 74–77, 90, 93–97, 101–103

LoFG level of façade generalization. 43, 48, 54, 69

ML machine learning. 28

MLS mobile laser scanning. xv–xvii, 3, 10, 11, 18, 19, 29, 33, 34, 37–40, 45, 46, 59–61, 67, 73–76, 86, 88, 93, 96, 97

NLOS non-line-of-sight. 2

OGC Open Geospatial Consortium. 103

ORB Oriented FAST and Rotated BRIEF. 61, 88

OSM OpenStreetMap. 37, 44

PCA principal component analysis. 47, 77

PPD posterior probability distribution. xvii, 74

PT Point Transformer. xvii, 71, 78, 81, 83, 86

RANSAC RANdom SAmple Consensus. xv, 35, 38, 39, 75

rjMCMC reversible-jump Markov Chain Monte Carlo. xiv, 34, 35

RMS root mean square. 83, 84, 98

RMSE root mean square error. 40

SfM Structure from Motion. 18

SIFT Scale Invariant Feature Transform. 61

SVD singular value decomposition. 47

TLS terrestrial laser scanning. 18, 19, 22

Contents

1 Introduction

In our thriving modern society, where challenges are seen as opportunities for positive change, addressing resource shortages and combating climate change is not just a necessity but a chance to shine brighter. Unavoidably, we must reuse (recycle) the already-created goods to sustain our development. On the other hand, we are insufficiently weaponized to tackle various pressing challenges, such as climate change, with our current data. This fact, in turn, implies that we still need to create more sophisticated solutions and acquire more data.

At the forefront of these challenges stand urban digital twins. The term digital twin has been first introduced in the domain of a product life cycle, where not only its structure is deemed pertinent but also its current and changing status; essentially formulating requirements for models representing geometry, semantics, and time-stamp. Urbanists have adopted this concept since a city can be described as a system with its own life cycle, analogous to any industrial machine [Batty, 2018; Kolbe & Donaubauer, 2021]. The critical element in the urban digital twins is acknowledging and addressing the influx of unstructured and structured data from various sensors and stakeholders depicting the same phenomenon. This trend inevitably necessitates creating reliable reference data, which will work as an anchor for the other data. Such an anchor for 3D building models frequently is governmental semantic 3D building models [Kolbe & Donaubauer, 2021]. At the same time, even the reference data is subject to constant changes that shall be identified and refined in the model. Also, as we become equipped with more accurate data, our reference model may be enriched in new features, enhancing its capabilities as the reference data. Yet, the complete removal of an "anchor" is infeasible as it will impose the destruction of existing connections to the reference data. My thesis delves into the reusability of existing 3D building models in alignment with this philosophy. This research aims to unlock the full potential of these models, leveraging modern point cloud acquisition techniques to craft more accurate and detailed representations; An overarching theme reads: "When should we update the existing information based on the new evidence?".

1.1 Motivation

Reconstructing detailed semantic 3D building models is a long-standing challenge in geoinformatics, photogrammetry, and computer vision fields [Szeliski, 2010; Haala & Kada, 2010]. Recent advancements have demonstrated that using 2D building footprints and aerial observations ensures the automatic reconstruction of building models up to the LoD 2, characterized by complex roof shapes and planar facades [Roschlaub & Batscheider, 2016; Haala & Kada, 2010]. The models' pivotal traits are their watertightness and object-oriented modeling. These have proved essential for numerous applications, for example, in estimating the solar potential of roofs or simulating urban wind flows [Biljecki et al., 2015]. Moreover, such LoD2 models are nowadays ubiquitous, and many of them are freely available, as underscored by approximately 140 million open-access building models in the United States, Switzerland, and Poland [1].

As shown in Figure 1.1, the current challenge lies in the automatic reconstruction of facade-detailed semantic LoD3 building models, which still remains in its infancy owing to its reconstruction complexity and data availability. Currently, LoD3-specific elements, such as windows and doors, are often manually modeled and typically only for landmark buildings[Uggla et al., 2023; Chaidas et al., 2021]. The typical workflow leading to 3D semantic building reconstruction comprises several steps. The first is to acquire the data, which might be performed using various sensors, such as light detection and ranging (LiDAR) scanners, cameras, or depth cameras. This data typically includes point clouds, images, and potentially other sensor data. The sensors can be placed on moving vehicles, such as planes or cars, but can also

[1] https://github.com/OloOcki/awesome-citygml

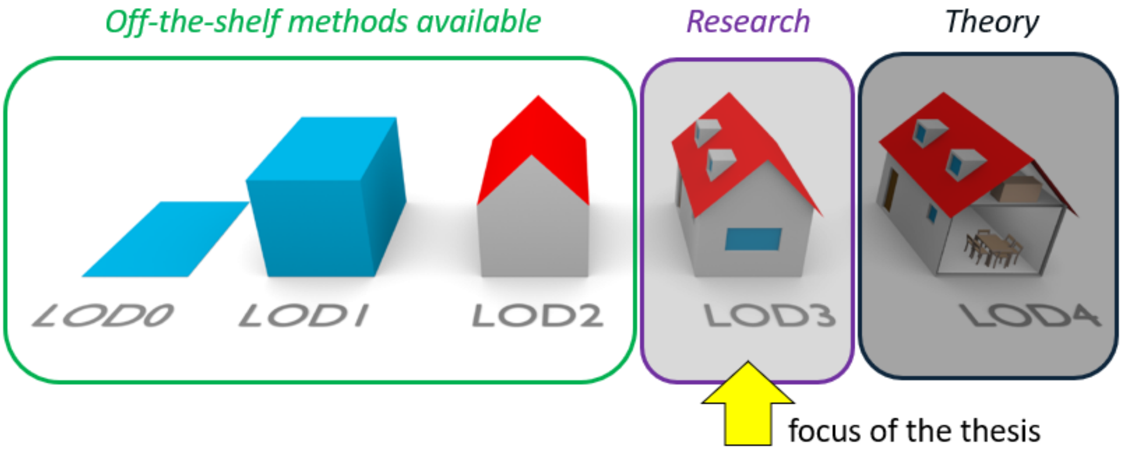


Figure 1.1 While established commercial and open-source 3D reconstruction methods for low LoD building models exist, the research of high-definition building model reconstruction remains in its infancy. This thesis contributes to the research community by introducing novel methods to reconstruct facade-detailed LoD3 building models. Adapted from Biljecki et al. [2016].

be taken in situ. Then, the data is pre-processed, cleaned of noise, corrected for sensor distortions, and compatibility between different data sources is ensured. This step may involve calibration and co-registration, especially in the case of multi-modal data; This step is crucial to ensure coherency. In the global context, the co-registration is also called geo-referencing, where a global coordinate reference system (CRS) is used as a harmonization framework of input datasets. The co-registration is followed by data understanding: The process is distilled into feature extraction and semantic labeling; The former deals with identifying and extracting features from the data. These features can include structural elements like walls, floors, doors, windows, and other architectural components; The latter assigns semantic labels to the extracted features. This step involves identifying and categorizing each component within the 3D model, e.g., labeling a set of connected planar surfaces as a wall. Finally, 3D semantic reconstruction is conducted, where a 3D geometrical model of the building is created using the annotated data. This step can involve techniques such as surface reconstruction and mesh generation. In this dissertation, contributions to each workflow step are presented.

Mobile mapping data appears to be the most suitable data source for semantic LoD3 facade modeling [Xu & Stilla, 2021], providing high local 3D point accuracy, dense street-level point cloud measurements with radial values, and offering high measuring flexibility at short acquisition time. The absolute 3D point accuracy, however, relies on a reliable global navigation satellite system (GNSS) signal, whose accuracy decreases in an urban environment due to the multipath and non-line-of-sight (NLOS) signal phenomena [Zhang et al., 2018]. This flaw directly impacts the co-registration accuracy of existing 3D models and acquired 3D point clouds, introducing a positioning uncertainty between them. Such uncertainty requires quantification to match corresponding objects and decide the possible enrichment rate.

While the datasets are co-aligned with specific confidence, the question of data semantics arises. Therefore, 3D semantic reconstruction is frequently preceded by 3D semantic segmentation: Its degree of robustness, accuracy, and completeness have an immediate impact on the final 3D reconstruction performance. Several learning-based solutions for facade-level 3D point cloud segmentation have shown promising performance in the past decade [Grilli & Remondino, 2020; Matrone et al., 2020]. However, their accuracy decreases while exposed to translucent objects (e.g., windows) and imbalanced training samples (e.g., doors) [Matrone et al., 2020]. Such challenging facade element detection has been tackled by ray-casting-driven methods, where the intersection of laser ray with model surface serves to find laser-penetrable objects, e.g., windows [Tuttas & Stilla, 2013; Hoegner & Gleixner, 2022]. This method is also not flawless, as it lacks semantic information (e.g., door or window decision) and is prone to field-of-view obstacles (e.g., window blinds) [Tuttas & Stilla, 2013; Hoegner & Gleixner, 2022]. The alternative approach to mitigating this issue assumes using images instead of point clouds. Image-driven methods

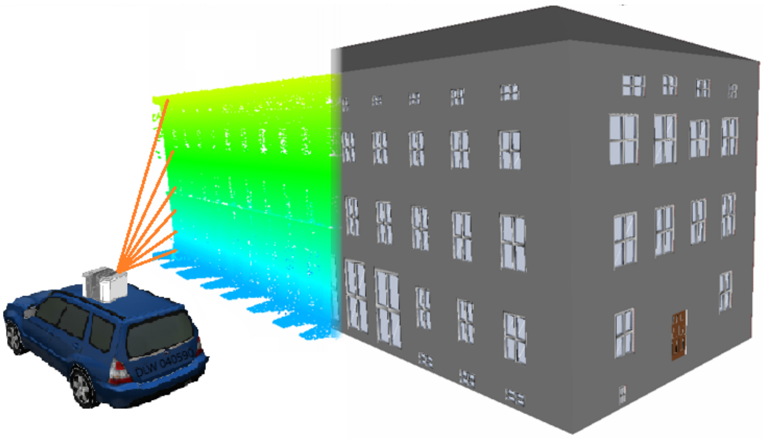


Figure 1.2 The refinement strategy: Laser scanning yields dense 3D point cloud, which, after the semantic segmentation, serves as a source for the enrichment of building models in facade details while the input modalities are subject to uncertainty [Wysocki et al., 2023b].

achieve high performance in facade segmentation, but their 2D nature hinders their direct application to 3D facade segmentation [Riemenschneider et al., 2012; Liu et al., 2020].

Considerable research efforts have been dedicated to the 3D semantic reconstruction of facade elements. However, methods that directly integrate these elements into the structure of semantic 3D building models are limited. One of the main challenges arises from the incomplete nature of acquired data, where only the street-accessible frontal facades are often captured. This limitation hampers the reconstruction process, as most approaches require full building coverage[Nan & Wonka, 2017]. Another significant challenge is the hierarchical complexity of the semantic data model, which extends beyond the typical segmentation derivation of single-object semantics. For instance, embedding 3D window objects into wall surfaces belonging to a building entity within a city model poses more complexities than detecting window and its extent [Pantoja-Rosero et al., 2022; Huang et al., 2020]. These challenges necessitate novel methodologies and techniques to overcome the limitations of embedding facade elements into the structure of semantic 3D building models. As depicted in Figure 1.2, the presented refinement strategy leverages MLS point cloud traits and ubiquity of semantic 3D building models while addressing their uncertainties to reconstruct LoD3 building models.

1.2 Research questions

In this thesis, to tackle the challenges mentioned above, the refinement strategy is introduced, which addresses the following research questions:

- To what extent can model priors of building models be used for the fusion of mobile mapping point clouds with building models?
- What level of geometric and semantic refinement of building models can be achieved using the conflict analysis approach?
- What level of improvement of accuracy and semantic completeness of building models can be achieved by:
  - Conflict analysis with model and library knowledge?
  - Conflict analysis with the model, library knowledge, and images?

1.3 Contributions

Within the scope of this dissertation, the presented methods aim to answer the research questions, ultimately contributing to the development of the refinement strategy studies. The methods advance the standard reconstruction workflow at each step, commencing with fusion, semantic segmentation, and closing with semantic reconstruction, as shown in Figure 1.3.

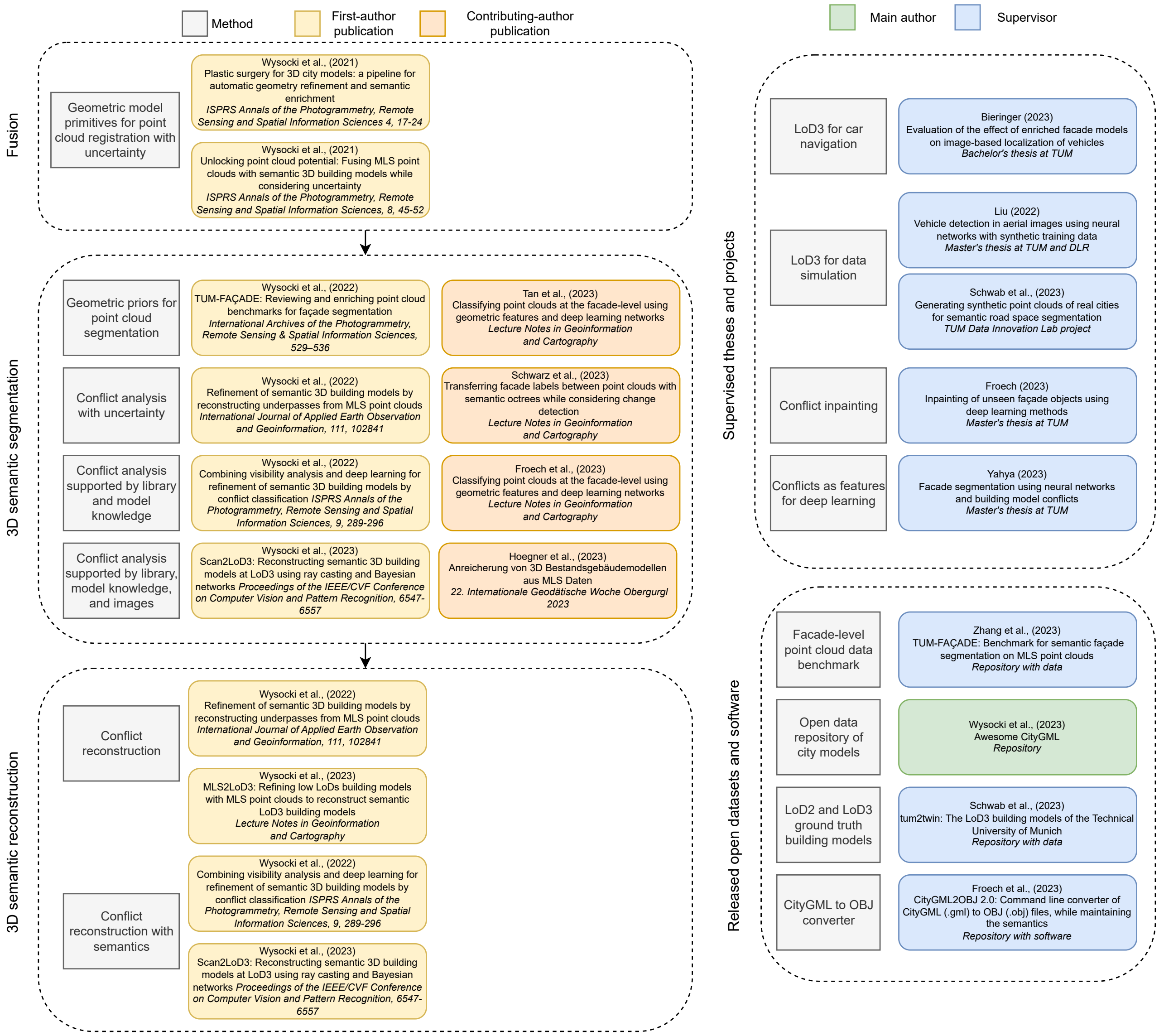

Method
First-author publication
Contributing-author publication
Main author
Supervisor
Fusion
Geometric model primitives for point cloud registration with uncertainty
Wysocki et al., (2021) Plastic surgery for 3D city models: a pipeline for automatic geometry refinement and semantic enrichment *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4, 17-24*
Wysocki et al., (2021) Unlocking point cloud potential: Fusing MLS point clouds with semantic 3D building models while considering uncertainty *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 8, 45-52*
3D semantic segmentation
Geometric priors for point cloud segmentation
Wysocki et al., (2022) TUM-FAÇADE: Reviewing and enriching point cloud benchmarks for façade segmentation *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 529–536*
Tan et al., (2023) Classifying point clouds at the facade-level using geometric features and deep learning networks *Lecture Notes in Geoinformation and Cartography*
Conflict analysis with uncertainty
Wysocki et al., (2022) Refinement of semantic 3D building models by reconstructing underpasses from MLS point clouds *International Journal of Applied Earth Observation and Geoinformation, 111, 102841*
Schwarz et al., (2023) Transferring facade labels between point clouds with semantic octrees while considering change detection *Lecture Notes in Geoinformation and Cartography*
Conflict analysis supported by library and model knowledge
Wysocki et al., (2022) Combining visibility analysis and deep learning for refinement of semantic 3D building models by conflict classification *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 9, 289-296*
Froech et al., (2023) Classifying point clouds at the facade-level using geometric features and deep learning networks *Lecture Notes in Geoinformation and Cartography*
Conflict analysis supported by library, model knowledge, and images
Wysocki et al., (2023) Scan2LoD3: Reconstructing semantic 3D building models at LoD3 using ray casting and Bayesian networks *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6547-6557*
Hoegner et al., (2023) Anreicherung von 3D Bestandsgebäudemodellen aus MLS Daten *22. Internationale Geodätische Woche Obergurgl 2023*
3D semantic reconstruction
Conflict reconstruction
Wysocki et al., (2022) Refinement of semantic 3D building models by reconstructing underpasses from MLS point clouds *International Journal of Applied Earth Observation and Geoinformation, 111, 102841*
Wysocki et al., (2023) MLS2LoD3: Refining low LoDs building models with MLS point clouds to reconstruct semantic LoD3 building models *Lecture Notes in Geoinformation and Cartography*
Conflict reconstruction with semantics
Wysocki et al., (2022) Combining visibility analysis and deep learning for refinement of semantic 3D building models by conflict classification *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 9, 289-296*
Wysocki et al., (2023) Scan2LoD3: Reconstructing semantic 3D building models at LoD3 using ray casting and Bayesian networks *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6547-6557*
Supervised theses and projects
LoD3 for car navigation
Bieringer (2023) Evaluation of the effect of enriched facade models on image-based localization of vehicles *Bachelor's thesis at TUM*
LoD3 for data simulation
Liu (2022) Vehicle detection in aerial images using neural networks with synthetic training data *Master's thesis at TUM and DLR*
Schwab et al., (2023) Generating synthetic point clouds of real cities for semantic road space segmentation *TUM Data Innovation Lab project*
Conflict inpainting
Froech (2023) Inpainting of unseen façade objects using deep learning methods *Master's thesis at TUM*
Conflicts as features for deep learning
Yahya (2023) Facade segmentation using neural networks and building model conflicts *Master's thesis at TUM*
Released open datasets and software
Facade-level point cloud data benchmark
Zhang et al., (2023) TUM-FAÇADE: Benchmark for semantic façade segmentation on MLS point clouds *Repository with data*
Open data repository of city models
Wysocki et al., (2023) Awesome CityGML *Repository*
LoD2 and LoD3 ground truth building models
Schwab et al., (2023) tum2twin: The LoD3 building models of the Technical University of Munich *Repository with data*
CityGML to OBJ converter
Froech et al., (2023) CityGML2OBJ 2.0: Command line converter of CityGML (.gml) to OBJ (.obj) files, while maintaining the semantics *Repository with software*

Figure 1.3 Peer-reviewed publications, supervised projects, and released open source tools: Contributions to refining strategy research within the dissertation scope.

1.4 Thesis structure

The remainder of the thesis is organized as follows. Chapter 2 describes the state of the art in the refinement of semantic 3D building models with its pertinent topics, for example, common semantic 3D city model standards, basics of point cloud acquisition, principles of uncertainty estimation, change detection, 3D semantic segmentation, and reconstruction. Chapters 3, 4, and 5 present the methods addressing the thesis's research questions, while chapters 6 and 7 show the methods' experiments and results, respectively. The discussion of the presented methods and experiments is described in chapter 8. The thesis is concluded in chapter 9, followed by the outlook presentation and implications of the work in chapter 10.

2 State of the art in the refinement of semantic 3D building models

Several elements can describe buildings' exteriors. Generally, a building shell consists of a ground surface, a roof, and walls. The façade (or facade) according to Cambridge Dictionary [2023a] is a *front of a building*. To be more precise, the Cambridge Dictionary [2023b] also explains the word *front*, which refers to *the part of a building that faces forward or is most often seen or used*. Thus, each element of a façade constitutes the building exterior, such as walls, ornaments, cornices, or adjacent objects, e.g., rain gutters. From this group of elements, the roof structure and the ground surface should be excluded as they are neither designed to face forward nor to be most frequently used or seen by dwellers. Elements of building architecture, such as underpasses, should belong to the façade group as they are intended to be often used by pedestrians or cars.

According to the Cambridge Dictionary, the word refinement means "a small change that improves something" or "the process of making a substance pure" [Cambridge Dictionary, 2023c]. Within the scope of this thesis, both definitions are valid. On the one hand, the refinement strategy performs relatively minor changes to the overall spatial 3D building model presence, yet it significantly improves its accuracy, completeness, and semantics. On the other hand, the strategy retains validity and removes undesired model parts while upgrading the low LoD to high LoD model, purifying the building's representation.

This idea contrasts the commonly used 3D reconstruction strategy, which creates 3D models based on solely unstructured sensor data. This chapter provides a comprehensive overview of methods pertinent to the proposed façade-driven refinement strategy.

2.1 Semantic 3D building models

3D building models are generally divided into the visual 3D models and semantic 3D models [Kolbe & Donaubauer, 2021]. The former is commonly represented by mesh-based 3D surfaces with local topological relations and appearance; the latter typically describes object-based 3D entities with global semantic and topological relations. While the visual models of 3D buildings are frequently used for human-centered visualization purposes, the semantic 3D building models convey machine-readable information, invaluable in numerous computer-based applications, such as simulating floods or estimating heating demand [Biljecki et al., 2015]. This section presents the most commonly used standard in semantic 3D building models and their current and potential applications.

2.1.1 CityGML

CityGML, short for City Geography Markup Language, is an open data model and an XML-based format for describing and representing 3D city models. It is developed to enable the exchange and storage of digital 3D representations of city models, the whole city infrastructure, and even rural areas. CityGML has been developed by the Open Geospatial Consortium (OGC) and is widely used in urban planning, geospatial analysis, and virtual city applications [Gröger et al., 2012; Kolbe et al., 2005]. Recently, an encoding of the standard, called CityJSON, has been introduced. CityJSON is based on the JSON (Java Script Object Notation) format, making it easy to parse and generate using a wide range of programming languages on the expense of simplifying the CityGML data structure [Ledoux et al., 2019].

The representation of objects in CityGML can range from LoD0, allowing 2.5D geometries (elevation and footprint), up to LoD4, which also enables modeling of object interiors and exteriors (e.g., buildings), as illustrated in Figure 2.1. Besides the Building module, CityGML introduces modules such as Bridge, CityFurniture, CityObjectGroup, Generics, LandUse, Relief, Transportation, Tunnel, Vegetation, and WaterBody [Gröger et al., 2012]. To depict the 3D geometric characteristics and contours of distinct entities, boundary representations (B-Reps) have become a prevalent choice. The accumulated bounding surfaces are then used to create volumetric geometries. The rationale supporting this approach owes to the source data typically stemming from photogrammetry and remote sensing data, which are outdoor-acquired and unavoidably lead to modeling only outer-observable surfaces. Another consequence of the outdoor acquisition, unlike in many other fields, is global CRS for vertices of the geometries [Kolbe & Donaubauer, 2021]. This characteristic contrasts the CAD and BIM modeling, where volume representation and local coordinate systems are typically chosen. These types of models, however, are meant to represent single objects rather than whole cities.

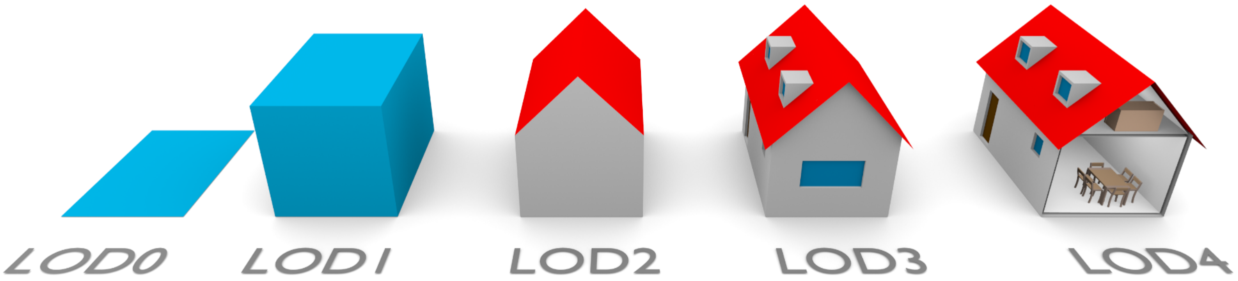


Figure 2.1 Level of details (LoD) according to the CityGML 2.0 standard. Note that LoD4 is removed in the revised 3.0 version. Source: [Biljecki et al., 2016].

Kutzner et al. [2020] announce a revised version of CityGML 2.0 standard: CityGML 3.0. The main changes concern the Building and Transportation modules. Also, the concept of LoD is rewritten and removes LoD4 (Figure 2.1); In turn, building interiors can be now modeled in the remaining LoD 0-3. Besides those adaptations, new modules such as Versioning, Dynamizer, PointCloud, and Construction are also introduced. For example, the Versioning module can track changes in a building's semantics or geometry, enabling the storage of old features via a common identification number. PointCloud is designed to represent a phenomenon with MultiPoint geometry explicitly or by linking to external files. The creators allow conversion from version 2.0 to 3.0 by utilizing solely syntactical transformations. The

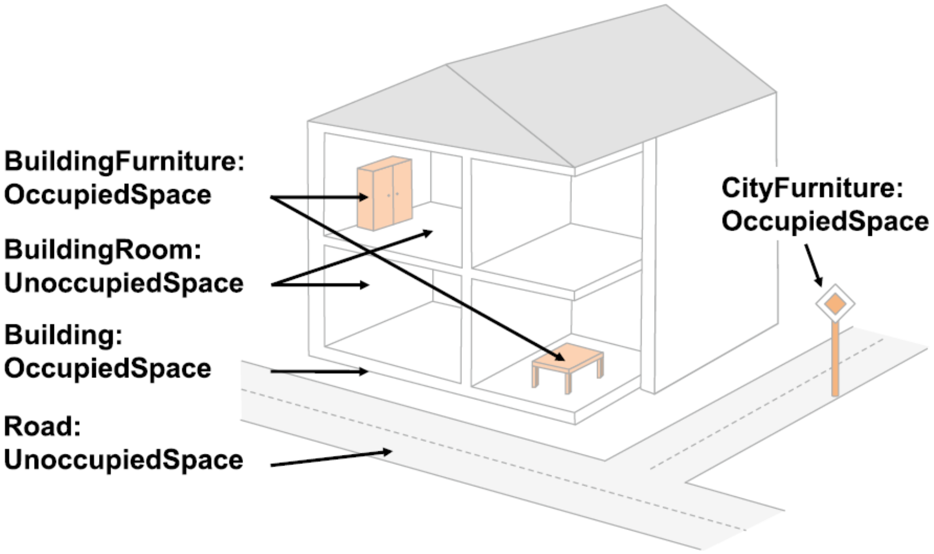


Figure 2.2 Concept of occupied and unoccupied spaces used in the revised 3.0 CityGML version. Source: [Kutzner et al., 2020].

changes in the revised CityGML version are introduced to adapt to new applications of the standard, such as the conversion of BIM to Geographic Information System (GIS) models or automatic driving simulations [Schwab & Kolbe, 2019].

The Space concept is a novelty in the CityGML standard, inspired by the field of robotics, where the term occupancy grids map navigable spaces. Thus, a road can indicate not only possible driveable 2D extent but also the maximum possible height for a vehicle on a particular road segment (ClearanceSpace). The Space represents the 3D border's entity in the real world, subdivided into *physical spaces* and *logical spaces*. While logical space is an abstract volumetric object (e.g., industrial zone), physical space depicts a physical object with a further subdivision into *occupied spaces* and *unoccupied spaces*. Occupied space stands for an existing volumetric object in a specific place. In turn, unoccupied space does not occupy volumetric space and thus can still be taken by other urban objects or actors (see Figure 2.2). For example, this concept finds its application in modeling road environments. As illustrated in Figure 2.3, the available traffic space (*TrafficSpace*) for cars, pedestrians, and cyclists is modeled based on footprint-extruded 3D prismatic volumes. However, conflicts between other objects may occur, for instance, with unmodeled building underpasses or vegetation [Beil et al., 2020].

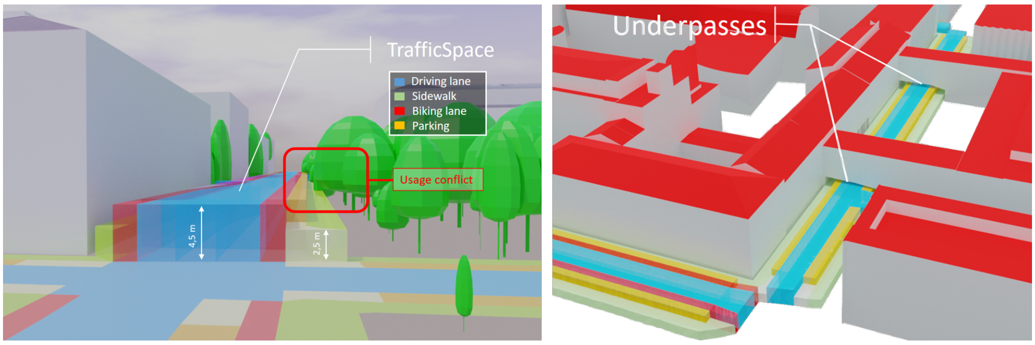


Figure 2.3 Concept of TrafficSpace in CityGML 3.0 and potential clashes with vegetation and underpasses. Source: [Beil et al., 2020].

2.1.2 Applications of refined semantic 3D building models at LoD3

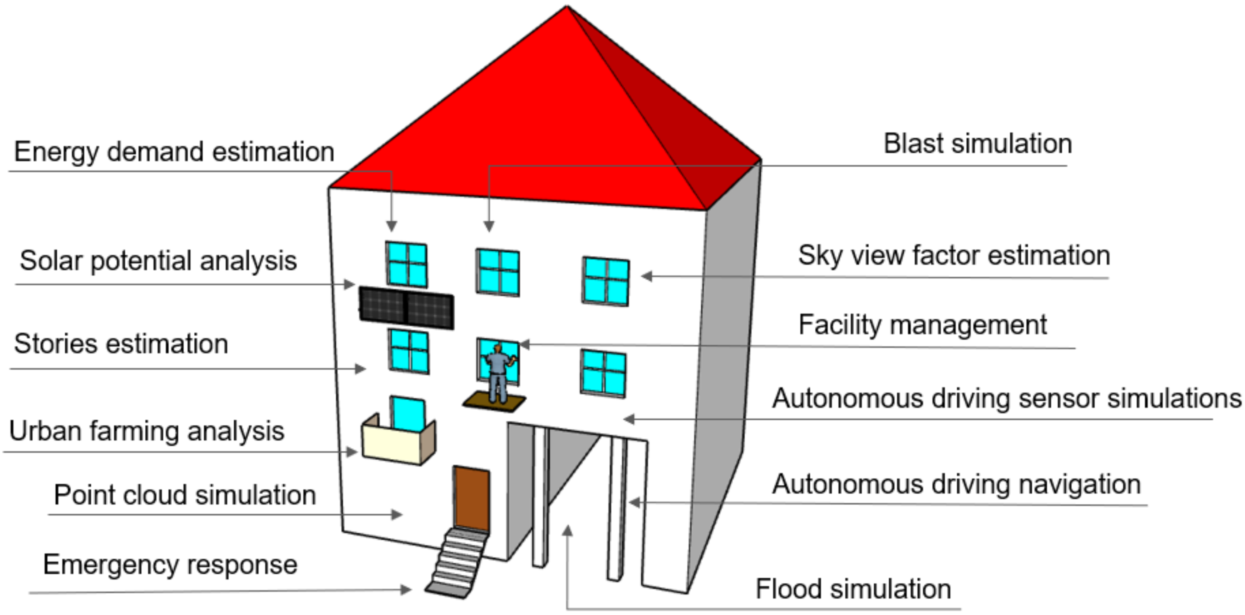


Figure 2.4 Applications of semantic 3D building models at LoD3. Source: [Wysocki et al., 2023a].

Much research has been devoted to analyzing applications of semantic 3D building models, mainly concentrating on LoD 1 and 2 models [Biljecki et al., 2015; Willenborg et al., 2018b]. This phenomenon owes to the widespread availability of LoD 1 and 2 models, which are available as open- and proprietary-datasets. At the time of writing, 59 governmental bodies are releasing their building models at no cost [1]. The availability of LoD 1 and 2 building models has rapidly evolved due to the robust methods for semantic 3D reconstruction using aerial observations in conjunction with cadastre-derived building footprints [Haala & Kada, 2010; Roschlaub & Batscheider, 2016] and non-footprint approaches [Nan & Wonka, 2017]. Even though the LoD3 are barely available, the researchers have already investigated generalizing high-level building models into low-level building models while maintaining minimum information loss [Muñumer Herrero et al., 2018].

However, the need for methods automatically reconstructing façade-rich LoD3 building models is growing. As illustrated in Figure 2.4, potential applications of LoD3 models concern calculating photovoltaic potential of façades [Willenborg et al., 2018a], improving management of building facilities [Hijazi et al., 2018], calculating number of stories [Biljecki & Ito, 2021], simulating flood impact [Amirebrahimi et al., 2016], creating test-beds for simulating automated driving functions [Schwab & Kolbe, 2019], enhancing positioning and navigation of autonomous cars [Wysocki et al., 2022b; Bieringer, 2023; Bieringer et al., 2024], estimating accurately heating demand [Nouvel et al., 2013], calculating sky-view factor [Lee et al., 2013], analysing façade potential for vertical farming [Palliwal et al., 2021], simulating blasts impact [Willenborg et al., 2016], enhancing navigation for emergency units [Kolbe et al., 2008], and generating semantic-rich simulation data [Liu, 2022; Schwab et al., 2023b].

2.2 Mobile laser scanning point clouds

Mobile laser scanning (MLS) point clouds typically stem from a mobile mapping unit (MMU), a system that maps an environment while moving through an area. The primary purpose of an MMU is to capture various types of data, such as geographic locations, 3D point clouds, images, and other relevant information, to create accurate and detailed maps of the surroundings. An example of such MMU is shown in Figure 2.5; The frequent setup of MMU comprises:

- GNSS receiver providing direct positioning in the global coordinate system, the accuracy may be further enhanced with the real-time kinematic positioning (RTK).
- Inertial Measurement Unit (IMU), which measures the vehicle's orientation and motion to enhance the accuracy of GPS data.
- Cameras capturing images to provide visual context to the collected data; can be used to color point clouds.
- MLS comprising one or more light detection and ranging (LiDAR) sensors, the primary MMU sensor emits high-frequency laser pulses to measure distances to objects, creating 3D point clouds.
- Other sensors such as radar, thermal, or multispectral cameras may be integrated for specific applications.

Mobile mapping systems are often mounted on vehicles, such as cars or vans, making them capable of efficiently surveying large street spaces. The technology has significantly improved the speed, accuracy, and cost-effectiveness of data collection for mapping purposes compared to traditional terrestrial surveying methods.

[1] https://github.com/OloOcki/awesome-citygml

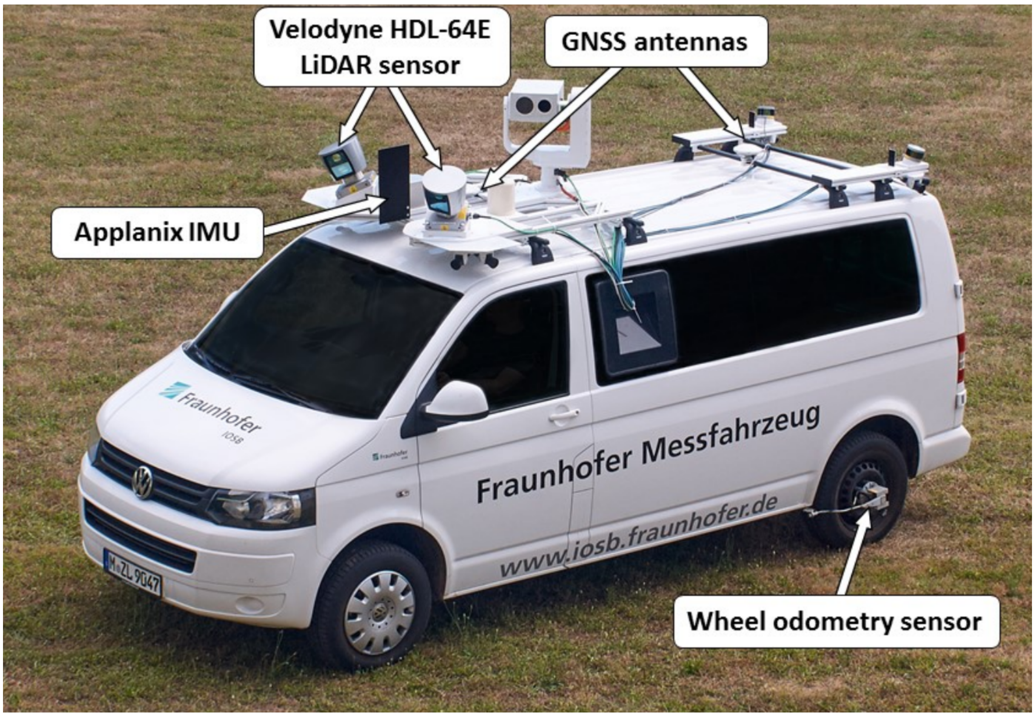


Figure 2.5 Exemplary mobile mapping unit and its typical sensors. Source: [Gehrung, 2022].

2.2.1 Georegistration of mobile mapping point clouds

Currently, MLS point clouds relative accuracy reaches 1 to 3 cm [3D Mapping Solutions, 2023]. Such accuracy unlocks multiple applications, from creating high definition (HD) maps to 3D visual tours. On the other hand, the global positioning accuracy of MLS point clouds relies on accurate system positioning of the vehicle's trajectory. This, in turn, is mainly based on the GNSS receiver, whose signal is prone to obstructions, ubiquitous in urban scenarios of high-rise buildings. To mitigate the trajectory drift, the intra-epoch registration is performed. One of the critical components is the so-called loop closure, which describes an event of measuring the previously captured scene; it allows for determining correction vectors for trajectory.

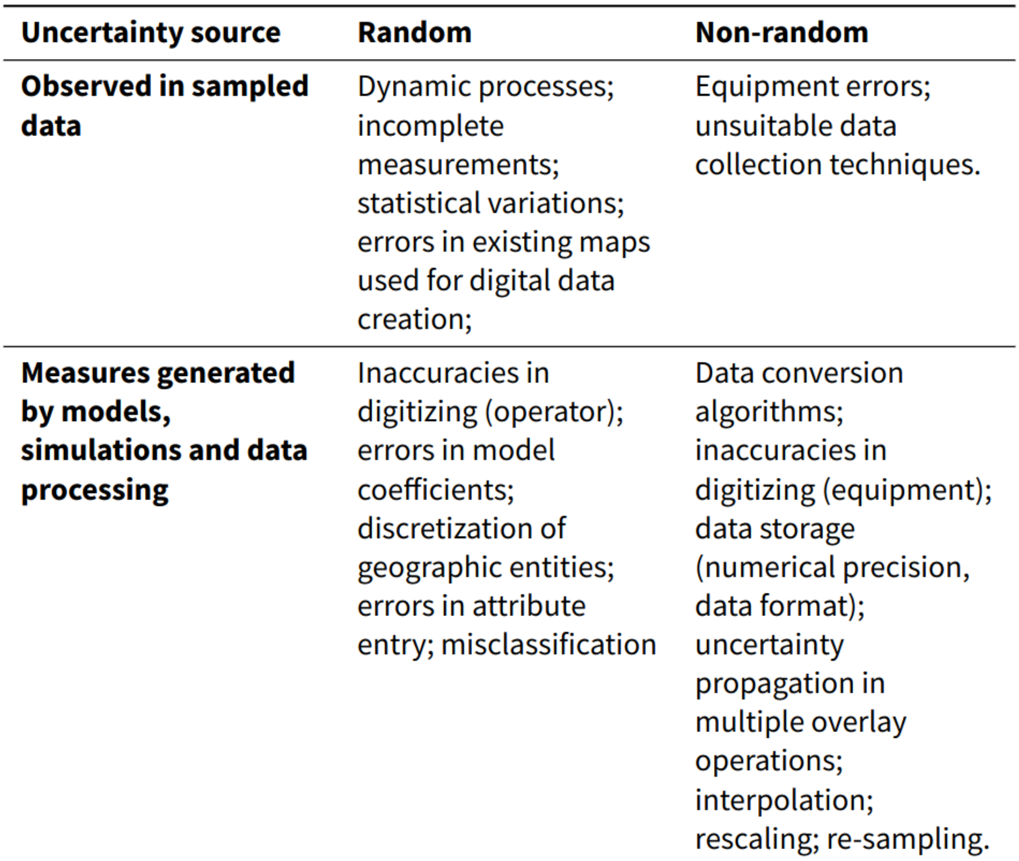
As shown by Gehrung [2022], correcting loop closures in a measurement run involves utilizing a method known as graph-based SLAM (Simultaneous Localization and Mapping). This approach creates a pose graph, where nodes represent the measured vehicle poses at different points in time, consisting of both translation and rotation information. The nodes are interconnected in chronological order through odometry constraints obtained from the post-processed trajectory. When a loop is detected, an additional loop closure constraint is inserted between the current node and the matched node [Carlone et al., 2015].

The pose graph is treated as an optimization problem to optimize the trajectory and solved using standard optimization techniques, such as Levenberg-Marquardt. The first vehicle pose remains fixed during optimization to enable a solvable problem. Loop closure constraints are assigned higher weights than odometry constraints due to their presumed higher accuracy. The correction vectors for rotation and translation are derived from the optimized graph and applied to the individual measurements, correcting loop closure errors.

Loop closures are determined by storing references to all poses in a two-dimensional grid. The vehicle poses are processed sequentially, and before inserting a pose into a cell, a neighborhood search on that cell and its neighbors is performed. If a reference is found, the corresponding measurements are compared, and the correction vector between the two scans is determined using an automatic registration algorithm, such as Iterative Closest Point (ICP) [Rusinkiewicz & Levoy, 2001]. The success of the ICP procedure is determined by applying a threshold to the ICP error measure. If a reasonable match is found, a new loop closure constraint is added to the pose graph, correcting registration errors. Although interinaccuracies are minimized with this approach, global geo-registration may still occur. Such inaccuracies are subject to uncertainties resulting from the approach above.

2.3 Uncertainty quantification using confidence values

The uncertainty quantification plays a key role in deciding upon the final state of the estimated phenomenon. Therefore, the uncertainty is explored in multiple domains, especially those dealing with high-risk decision-making, for example, avalanche prediction [Grêt-Regamey & Straub, 2006]. As shown in Figure 2.6, the uncertainties stem from data- and processing- sources. This section is based on the research published in my first-author publications: Wysocki et al. [2022b, 2021b].



Uncertainty source	Random	Non-random
Observed in sampled data	Dynamic processes; incomplete measurements; statistical variations; errors in existing maps used for digital data creation;	Equipment errors; unsuitable data collection techniques.
Measures generated by models, simulations and data processing	Inaccuracies in digitizing (operator); errors in model coefficients; discretization of geographic entities; errors in attribute entry; misclassification	Data conversion algorithms; inaccuracies in digitizing (equipment); data storage (numerical precision, data format); uncertainty propagation in multiple overlay operations; interpolation; rescaling; re-sampling.

Figure 2.6 Uncertainty sources in the acquired and processed data. Adopted from: [Chuprikova, 2019].

2.3.1 Uncertainty assignment

Each measurement and subsequent fusion is inevitably related to uncertainty issues. This might be expressed by, for example, acquisition technique, registration accuracy, or expert belief in metadata [Chuprikova, 2019]. One of the measures of uncertainty in the observed dataset is the confidence interval (CI) [Chuprikova, 2019]. Within the scope of 3D building reconstruction, the CI measure is used, for example, to accommodate for a map, roof extensions, and feature extraction inaccuracies [Suveg & Vosselman, 2000]. However, this measure is often obscured in map accuracy assessment, which results in false full certainty [Anderson et al., 2017].

The CI for 2D vector maps has been addressed by numerous studies [Neis et al., 2012; Haklay, 2010; Minghini et al., 2018]. Based on derived standard deviations and identified beliefs, a linear position in 2D is quantified. Additionally, there exist manuals and standards for specific urban assets that may be used as an initial 3D uncertainty value, for example, for road width [Fiutak et al., 2018; FGSV, 1996; Chacon, 2020] and clearance space [Chacon, 2020; U.S. Department of Transportation, 2014; Holst & Holst, 2004].

In this context, a frequently used method is using confidence levels. The belief in the error prior information is expressed by confidence level (CL_n). The found CI results in the estimation of the standard deviation (σ_n). The total standard deviation (σ) is found using the formula [Suveg & Vosselman, 2000]:

$$\sigma = \sqrt{\sigma_n^2 + \sigma_{n+1}^2} \tag{2.1}$$

The final upper and lower bounds are found on the basis of $[\mu - 2\sigma, \mu + 2\sigma]$ while assuming a certain confidence interval, typically 95 %. This assumption is valid for the inaccuracies representing Gaussian distribution and overestimating the bounds by operating in the L1 norm.

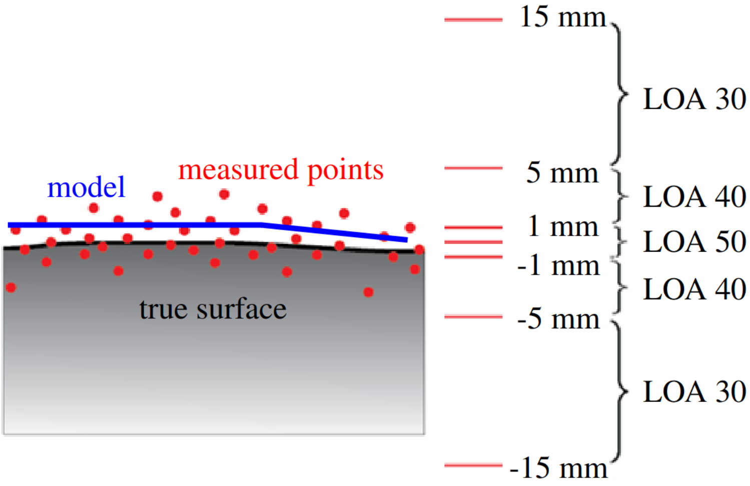


Figure 2.7 Schematic depiction of typical issue in comparison of as-is to as-planned state, where LoA indicates BIM-derived plausible acceptance thresholds. Source: [Meyer et al., 2021].

In the context of 3D model-oriented uncertainties, the uncertainty range might also be expressed by assuming a plausible range of a model based on standards. As shown in Figure 2.7, BIM models possess the LoA attribute, which describes allowed model deviations and can be directly used for uncertainty quantification. Although there are studies evaluating CityGML accuracy and providing suggested accuracy to the specific LoD, in practice, there are no fixed rules regarding the model accuracy [Biljecki et al., 2014; Gröger et al., 2012].

2.3.2 Uncertainty estimation

Dempster-Shafer theory (DST) [Shafer, 1992] is widely used for data fusion due to its ability to handle conflicting information implicitly and facilitate commutative and associative combinations of evidence. This is particularly valuable for dealing with objects scanned from different angles and epochs. The occupancy state can be represented by a universal set $U = empty, occupied$, and its power set, 2^U, consists of $\emptyset, empty, occupied, U$, encompassing all possible states. DST assigns a belief mass, ranging from 0 to 1, to each element in the power set. Notably, the empty set $\emptyset$ has zero mass, while the masses of the remaining sets are summed to one [Huang, 2021; Hebel, 2012]:

$$m : 2^U \rightarrow [0, 1], m(\emptyset) = 0,$$

$$\sum_{A \in 2^U} m(A) = 1$$

A distribution of masses assigned to the elements of the power set of U, meeting these criteria, is known as a basic mass function (BBA). Within Dempster-Shafer's theory, the assignments of the basis mass function are employed to define an interval encompassing classical probability, with boundaries termed degree of belief and plausibility. With the exception of the relationships described earlier, the mass of the universe $m(U) = m(\{\text{empty}, \text{occupied}\})$ does not directly influence the individual elements $\{\text{empty}\}$ and $\{\text{occupied}\}$ themselves, as they possess their own respective mass values. Instead, $m(U)$ represents the degree of uncertainty. A value of $m(U) = 1$ indicates that the space occupation at the given position is entirely unknown. As shown in Figure 2.8, Figure 2.9, and the following formulas, this concept is applicable to laser scanning point clouds:

$$m_{q,p}(\emptyset) = 0, \tag{2.2}$$

$$m_{q,p}(\text{emp}) = (1 - \frac{1}{1 + e^{-\lambda d_x - c}}) \cdot e^{-\kappa d_y^2} \tag{2.3}$$

$$m_{q,p}(\text{occ}) = (1 - \frac{1}{1 + e^{-\lambda d_x - c}} - \frac{1}{1 + e^{-\lambda d_x + c}}) \cdot e^{-\kappa d_y^2} \tag{2.4}$$

$$m_{q,p}(U) = 1 - m_{q,p}(\text{emp}) - m_{q,p}(\text{occ}) \tag{2.5}$$

Where:

$$d_x = (q - p) \cdot r_0 \tag{2.6}$$

$$d_y = ||(q - p) \times r_0|| \tag{2.7}$$

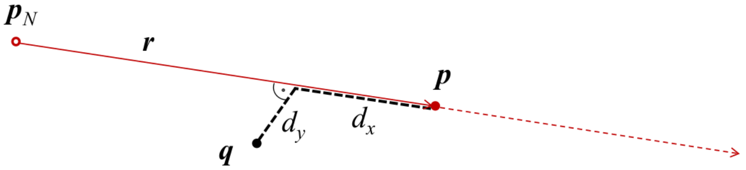


Figure 2.8 Longitudinal and transverse distance from the ideal ray (r). Source: [Hebel, 2012].

Dempster's combination rule is used to calculate a fused mass function, where the conflict C caused by m_1 and m_2 is specified: $C = m_1(e)m_2(o) + m_1(o)m_2(e)$. The applied combination rule ignores conflicts (contradictions) by using a normalization factor $(1 - C)$ [Hebel, 2012]:

$$m(e) = \frac{m_1(e)m_2(e) + m_1(e)m_2(U) + m_1(U)m_2(e)}{1 - C}, \tag{2.8}$$

$$m(o) = \frac{m_1(o)m_2(o) + m_1(o)m_2(U) + m_1(U)m_2(o)}{1 - C}, \tag{2.9}$$

$$m(U) = \frac{m_1(U) \cdot m_2(U)}{1 - C}, \tag{2.10}$$

$$m(\emptyset) = 0. \tag{2.11}$$

An alternative to uncertainty estimation with the Dempster-Shafer theory is the Bayesian reasoning [Chuprikova, 2019]. Bayes's theorem uses probabilities to express beliefs about hypotheses and allows for updating of these probabilities as new evidence is obtained. However, it does not explicitly handle conflicting evidence. Bayesian reasoning allows the inclusion of objective but also subjective beliefs in the form of prior probabilities, which provides a framework for including data-extracted and expert-extracted priors. They found its applications in numerous fields, including statistics, machine learning, risk assessment, and decision analysis.

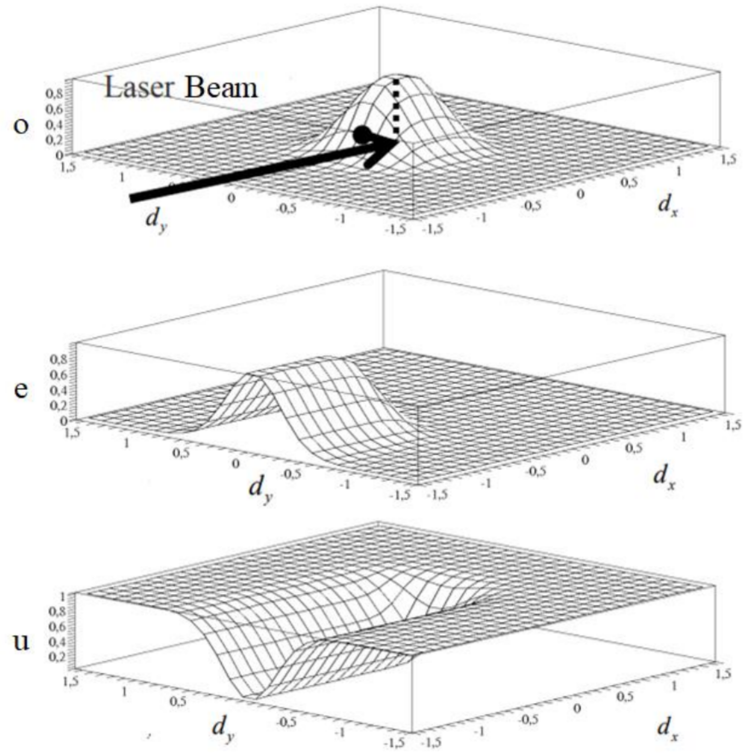


Figure 2.9 Visualized belief masses occupied (o), empty (e), and unknown (u). Source: [Huang, 2021; Hebel, 2012].

The characteristic of probability-based evidence updates is exploited to provide reliable 3D mapping frameworks [Hornung et al., 2013; Tuttas et al., 2015; Gehrung et al., 2017]. The probability $P(n|z_{1:t})$ of a leaf node n to be occupied given the sensor measurements $z_{1:t}$ is estimated according to:

$$P(n|z_{1:t}) = \left[1 + \frac{1-P(n|z_t)}{P(n|z_t)}\frac{1-P(n|z_{1:t-1})}{P(n|z_{1:t-1})}\frac{P(n)}{1-P(n)}\right]^{-1} \tag{2.12}$$

The given update formula (Equation 2.12) relies on three main components: the current measurement z_t, a prior probability $P(n)$, and the previous estimate $P(n|z_{1:t-1})$. The term $P(n|z_t)$ represents the probability of event n being measured based on the measurement z_t obtained from a specific sensor. It is important to emphasize that this value is specific to the characteristics and performance of the sensor used to generate z_t. As a result, the update process combines prior knowledge about sensor model with the sensor's measurement z_t to refine the estimate $P(n|z_{1:t})$. In the practical setup, efficient updates for large evidence databases are achieved by using log-odds notation; these allow to replace multiplication with additions (Equation 2.13):

$$L(n|z_{1:t}) = L(n|z_{1:t-1}) + L(n|z_t), \tag{2.13}$$

Where:

$$L(n) = \log\left[\frac{P(n)}{1-P(n)}\right] \tag{2.14}$$

Clamping policy (Equation 2.15) limits the number of updates required to alter an evidence's state. There are also further benefits. Firstly, it ensures that the map's confidence remains within a defined range, facilitating swift adaptation to changes in the surrounding environment. Additionally, the policy allows for the compression of neighboring evidence estimation. Yet, the clamping process is not entirely lossless: Information close to zero or one might be compromised, but total probabilities are effectively preserved for values lying between the clamping thresholds (Figure 2.10, Equation 2.16).

$$L(n|z_{1:t}) = \max\left(\min\left(L(n|z_{1:t-1}) + L(n|z_t), l_{\text{max}}\right), l_{\text{min}}\right) \tag{2.15}$$

Where:

$$\text{Lower log-odds bound: } l_{min}; \text{Upper log-odds bound: } l_{max} \tag{2.16}$$

The BayNets are based on the same Bayes's theorem. These are acyclic graphs incorporating joint probability distribution [Stritih et al., 2020; Chen & Pollino, 2012]. As shown in Figure 2.11, such graphs comprise nodes that represent random variables, where each node corresponds to a particular variable or event that can take on different states; Edges represent probabilistic dependencies between nodes, where an edge connecting two nodes indicates that the value of one node influences the probability distribution of the other node. These dependencies are often described using conditional probabilities along with Conditional Probability Table (CPT).

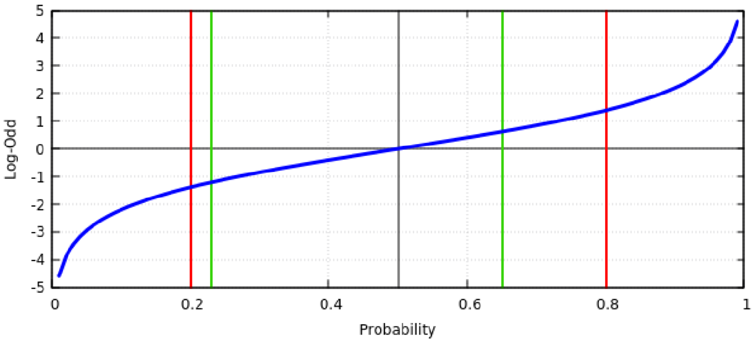


Figure 2.10 Symmetric (red) and asymmetric (green) setup of log-odd bounds. Source: [Gehrung et al., 2017].

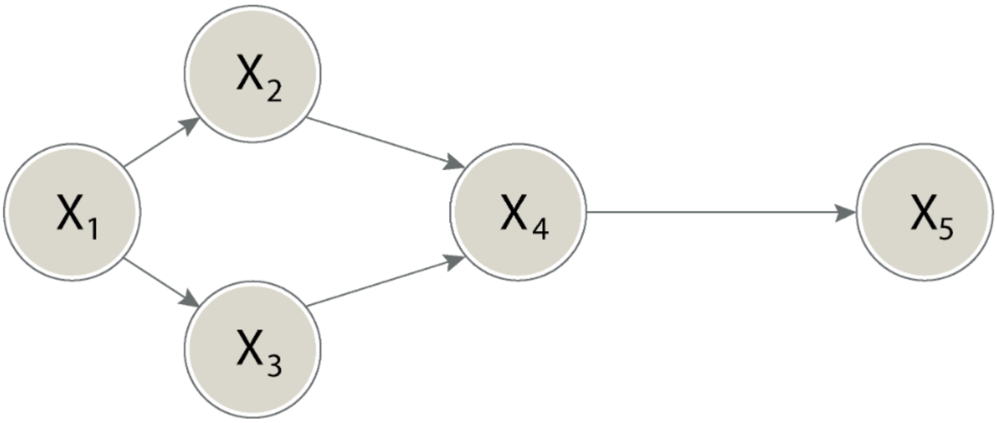


Figure 2.11 Relations among variables in the Bayesian network: Nodes represent variables, while edges describe direction and causal dependencies in the graph. Source: [Chuprikova, 2019]

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} \tag{2.17}$$

In the context of BayNet, $P(Y|X)$ signifies the probability of observing the evidence (Y) given that the hypothesis (X) is true (Equation 2.17). The prior $P(X)$ indicates the probability of the hypothesis before considering the evidence. The marginal $P(Y)$ represents the probability of observing the evidence under all possible hypotheses. Finally, the posterior $P(X|Y)$ denotes the probability of the hypothesis being true, given the observed evidence. BayNet provides a framework for reasoning under conditions of uncertainty and allows us to update the beliefs about a hypothesis based on new evidence [Chuprikova, 2019].

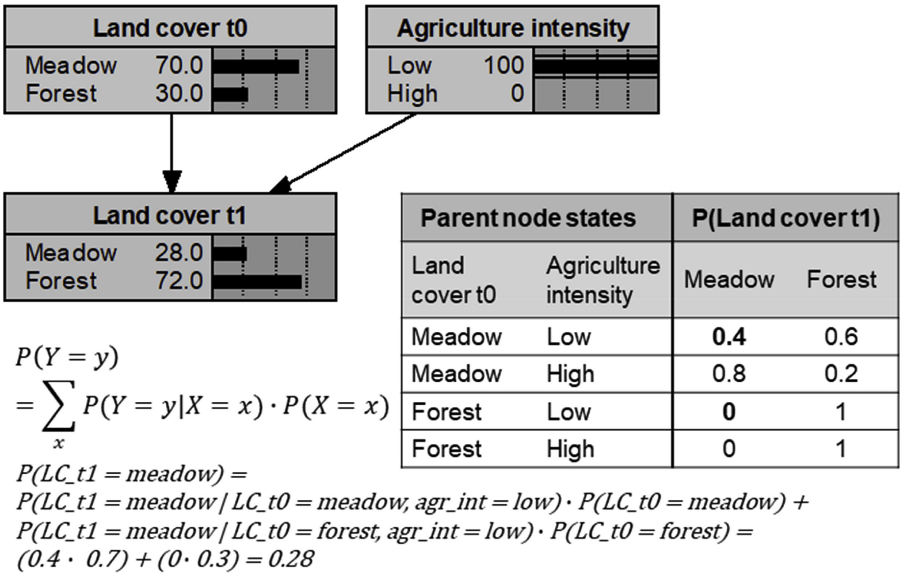
The target state (X_5 in Figure 4.10) is calculated based on the joint probability distribution (see Equation 2.18) and the CPT, which prescribes the probability weights of each node and each combination of parent node states.

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \Pi_{X_i}) \tag{2.18}$$

The probability that the node Y is in the state y is estimated using the so-called marginalization process, which sums conditional probabilities of the states x belonging to the parent nodes X [Stritih et al., 2020; Chuprikova, 2019]; see an example of CPT and target node estimation in Figure 2.12.

When the network is compiled, datasets are added as evidence for updating the joint probability distribution. There are two types of input evidence: soft and hard evidence. The former assumes prior probability given to the evidence being true in a range of $P \in [0, 1]$, while the latter assumes Boolean $P = 1$ or $P = 0$. The update is performed by an inference process, which estimates the posterior probability distribution (PPD) yielding the most likely node states [Stritih et al., 2020].

Such connected graphs enable uncertainty propagation throughout the whole analyzing process. While they incorporate soft evidence coming from the dataset's uncertainties, they enable the incorporation of qualitative measures, too. The BayNets proved to be an efficient tool where multimodal datasets and the operator's expertise are the important factors. A perfect example serves the modeling of complex environment phenomena investigated by GIS community, such as predicting avalanche release [Chen & Pollino, 2012; Stritih et al., 2020; Chuprikova, 2019].



Parent node states		P(Land cover t1)	
Land cover t0	Agriculture intensity	Meadow	Forest
Meadow	Low	**0.4**	0.6
Meadow	High	0.8	0.2
Forest	Low	**0**	1
Forest	High	0	1

$$P(Y = y) = \sum_{x} P(Y = y | X = x) \cdot P(X = x)$$

$$\begin{aligned} &P(LC\_t1 = meadow) = \\ &P(LC\_t1 = meadow \,/\, LC\_t0 = meadow, agr\_int = low) \cdot P(LC\_t0 = meadow) + \\ &P(LC\_t1 = meadow \,/\, LC\_t0 = forest, agr\_int = low) \cdot P(LC\_t0 = forest) = \\ &(0.4 \cdot 0.7) + (0 \cdot 0.3) = 0.28 \end{aligned}$$

Figure 2.12 Consider a basic Bayesian Network representing land cover change with nodes for current land cover (LC_t0), agriculture intensity (agr_int), and future land cover (LC_t1). LC_t1 is influenced by both LC_t0 and agr_int. CPT define the probabilities of future land cover states (e.g., meadow) based on the values of LC_t0 and agr_int. Hard evidence is given for 'agriculture intensity,' while soft evidence is assigned to 'land cover t0'. To estimate the posterior probability of LC_t1 being a meadow $P(LC_t 1 = meadow)$, marginalization considers all possible values of LC_t0 and agr_int. Source: [Stritih et al., 2020]

2.4 Change detection

Detecting changes between the new acquisition epoch and the previous acquisition epoch is referred to as change detection, which has long been a challenge in photogrammetry, remote sensing, and computer vision communities [Xiao et al., 2015]. As shown in Figure 2.13, to decide upon a change, a coregistration between measurement sets is frequently applied, whereby a coregistration of two heterogeneous data types is especially challenging [Stilla & Xu, 2023]. This work distinguishes between changes detected in

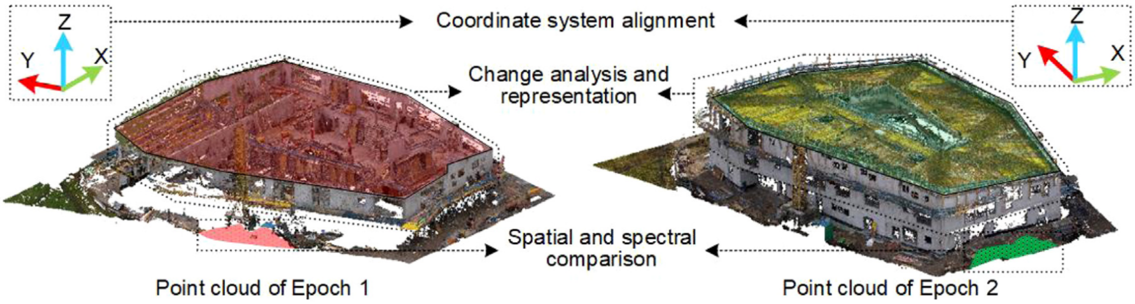

Coordinate system alignment
Z
Y
X
Change analysis and representation
Spatial and spectral comparison
Point cloud of Epoch 1
Point cloud of Epoch 2

Figure 2.13 Typical change detection workflow. Source: [Stilla & Xu, 2023].

unimodal (i.e., point cloud to point cloud) and multimodal (i.e., point cloud to vector) datasets; these are related to multimodal conflicts independent of time. This Section is primarily developed on the basis of the research published in my first-author publications: Wysocki et al. [2022b, 2021b].

2.4.1 Coregistration

In the context of urban scenarios, there are three main strategies of multiple-point sets' registration, namely: point-based, primitive-based, and global-based [Xu & Stilla, 2021]. The point-based strategy aims to find corresponding pairs of points from different point clouds, whereas primitive-based registration searches for geometric primitive shapes within point clouds to which correspondence can be found. The global-based approaches, however, neglect local information and focus on global features for entire point clouds.

One of the well-established examples of point-based strategy is the ICP algorithm and its variations point-to-point [Besl & McKay, 1992] and point-to-plane [Rusinkiewicz & Levoy, 2001]. Although it is a generic algorithm feasible in many applications, it is challenging to apply it for point clouds depicting an outdoor environment [Dong et al., 2020]. The main obstacles involve varying density, low overlap, and occlusions, among others [Dong et al., 2020; Xu & Stilla, 2021]. On the contrary, primitive-based solutions seem to be tailored to the harsh nature of registration in an urban environment. This is implied by many man-made structures localized within urban areas that are often represented by geometrical primitives. So far, researchers have investigated possibilities to formulate the registration process on the basis of found edges, crest curves, planes, and other geometric primitives [Xu & Stilla, 2021]. However, this strategy encounters another set of challenges, such as quality and consistency issues of extracted primitives or high computational time. While numerous works have tackled the challenge of terrestrial laser scanning (TLS) point cloud registration [Dong et al., 2020; Xu & Stilla, 2021], the registration of such point clouds based on semantic 3D city models seems to be obscured.

To the best of my knowledge, only a few research groups specifically tackled the coregistration and matching of point clouds and semantic 3D city models [Goebbels & Pohle-Fröhlich, 2018; Goebbels et al., 2019; Lucks et al., 2021]. While [Goebbels & Pohle-Fröhlich, 2018; Goebbels et al., 2018] incorporate a Mixed Integer Linear Program to find correspondences between modalities, the latest work uses a modified ICP point-to-plane [Goebbels et al., 2019]. Yet, the publications of Goebbels et al. focus on point clouds generated from images using the Structure from Motion (SfM) algorithm. This implies the usage of radiometric features for prefiltering [Goebbels et al., 2019] that may remove valid building features. Yet another approach is presented by Lucks et al. [2021], who incorporate the ICP point-to-plane algorithm for matching of MLS point clouds and semantic 3D city models. To increase the matching accuracy, they introduce random forests to select only point clouds' points depicting façades.

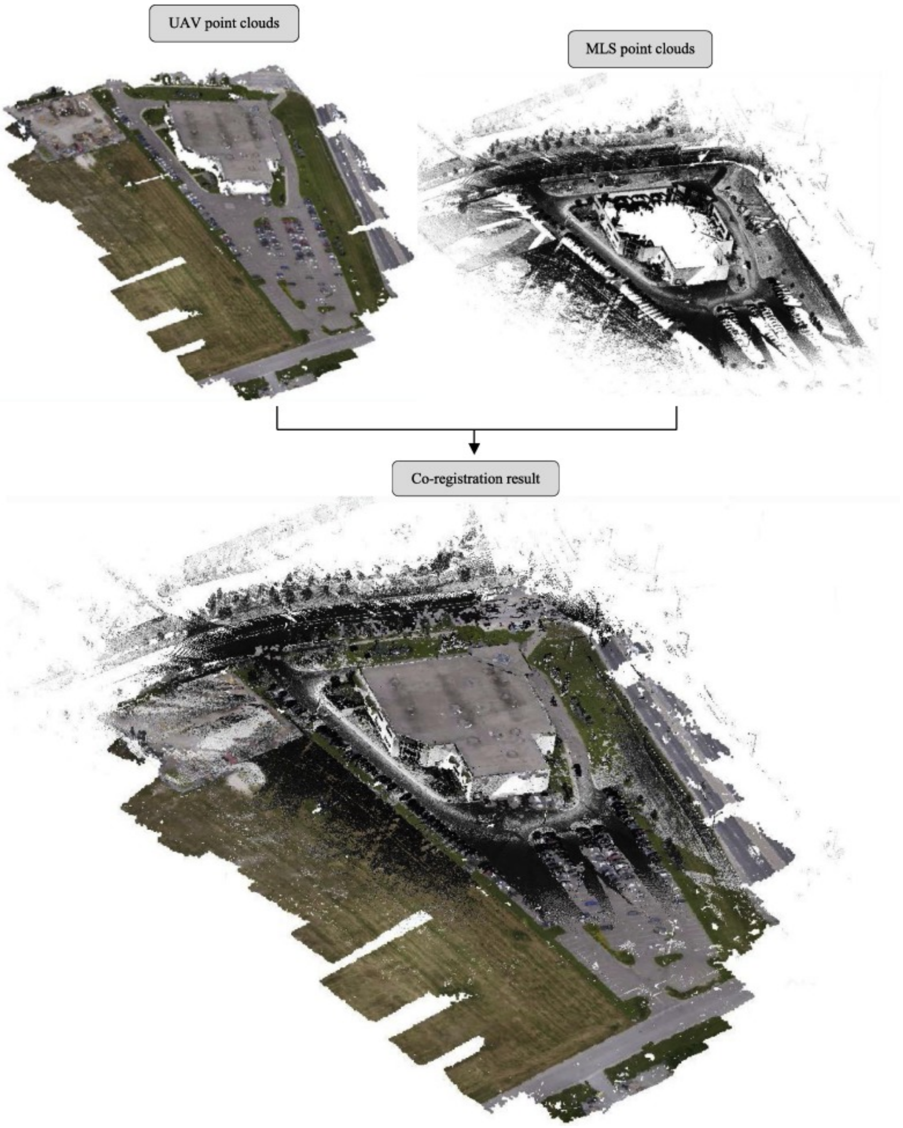


Figure 2.14 Example of a multi-modal coregistration of point clouds. Source: [Persad & Armenakis, 2017].

2.4.2 Cloud-to-cloud change detection

A method of detecting changes in different TLS point clouds is presented by Zeibak & Filin [2008]. To perform a point-wise comparison, they transform each scan into image panoramas. They define three possible states between point cloud epochs, such as *change*, *no change*, and *occlusion*.

Hebel et al. [2013] introduce a voxel grid for change detection. The uncertainty is modeled using DST. Based on ray tracing, they distinguish between three states: *occupied*, *empty*, and *unknown*, subsequently deriving the states: *consistent*, *disappeared*, and *appeared*. Hirt et al. [2021] pursue this idea using MLS point clouds to identify changes of urban trees. The concept of ray tracing on voxels is shown in Figure 2.15. The works by Hebel et al. [2013] and Hirt et al. [2021] differ in the role of the voxel grid and the choice of the voxel size. The former utilizes the occupancy grid as a search method for changes and acceleration of the search process and thus choose a coarse voxel size [Hebel et al., 2013]; Whereas the latter uses an occupancy grid for direct change identification and thus choose fine voxel size [Hirt et al., 2021]. The discretization impact of voxelization on object representation is shown in Figure 2.16.

In the work by Gehrung et al. [2017] a Bayesian approach is favored over DST for a fusion of MLS single sensor measurements. They present a method of removing dynamic objects in scenes by accumulating probabilities of voxel occupancy, which decrease when the voxel is traversed by a laser ray. The work

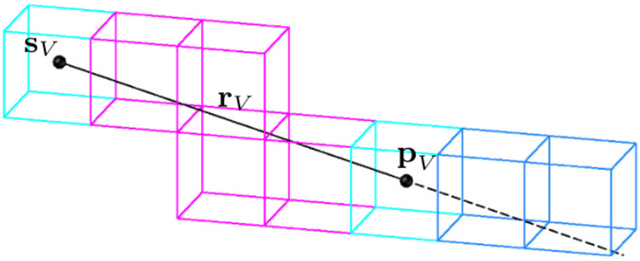


Figure 2.15 In voxel space V, there exists a scanner source s_V and a hit point p_V (i.e., laser point in point cloud). As the laser beam r_V travels from s_V to p_V, it intersects various voxels in its path. To represent this interaction, the traversed voxels are highlighted with magenta-colored voxels, while occupied voxels are displayed in cyan. Additionally, voxels that are obscured from view are indicated with a dark blue color, signifying their unknown status. Source: [Meyer et al., 2022].

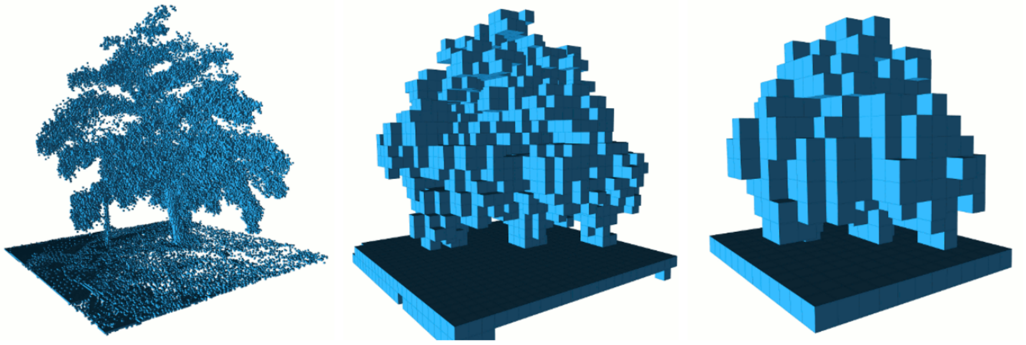


Figure 2.16 Impact of a voxel resolution on object's representation on an example of a tree: Voxel resolution of 8 cm, 64 cm, and 128 cm. Source: [Hornung et al., 2013].

is incorporated in the probabilistic *OctoMap* framework [Hornung et al., 2013] using an efficient octree structure for calculations. The octree structure is schematically shown in Figure 2.17.

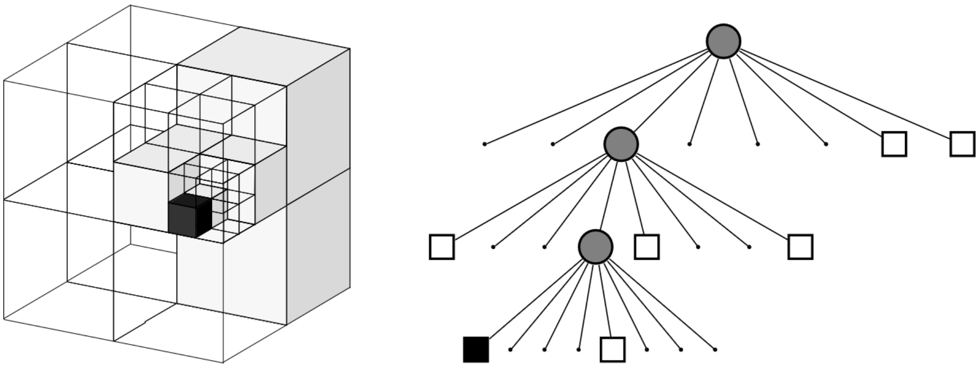


Figure 2.17 Octree volumetric representation (left) and its tree representation (right). Occupied (black) and free (white) voxels represent the leaves of the tree. Source: [Hornung et al., 2013].

2.4.3 Model-to-model change detection

The change detection is also well-studied in the context of model-to-model comparison. For example, Nguyen & Kolbe [2022] employ a Path-tracing Semantic Network (PSN) based on a graph-based path-tracing technique, which operates on an entire network to analyze the semantic meaning of its components.

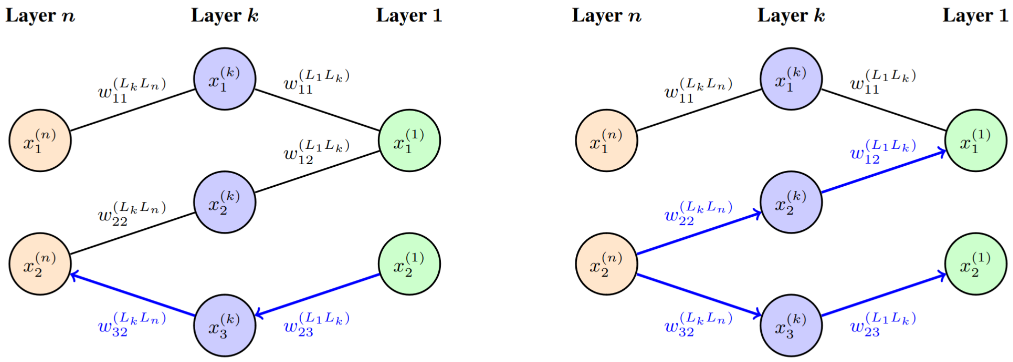


Figure 2.18 Forward path-tracing (left) and backward path-tracing (right). x_i^k indicates the weight of vertex v_i^k of layer L_k, and $w_{ij}^{(L_{k_1}\ L_{k_2})}$ indicates the weight of the edge connecting vertices $v_i^{k_1}$ and $v_j^{k_2}$ of layers L_{k_1} and L_{k_2}, respectively. Source: [Nguyen & Kolbe, 2021].

For effective forward and backward path tracing between the input and output layers, it requires that there is bidirectional traversal between two vertices (see Figure 2.18). This bidirectional traversal is achieved by establishing connections between the vertices using either undirected edges or two directed opposite edges. The network ensures navigation from one vertex to another and then back again, allowing for seamless exploration of paths in both directions. They distinguish five layers: stakeholder, actor role, reasoning layer, change type, and context layer. Their main network application is for detecting stakeholder changes. In other work of Nguyen & Kolbe [2021], they also propose graph-based solutions concentrating on finding changes in the CityGML appearance, semantics, geometry, and topology (see Figure 2.19).

Heeramaglore & Kolbe [2022] propose an alternative approach to identify changes. Their concept of RichVoxel transforms any input data into a homogeneous voxel grid. The comparison is then performed on a voxel-to-voxel basis using constructive solid geometry and its derived semantics. An example of such an approach is shown in Figure 2.20 Also, Aleksandrov et al. [2019] use voxels to compare 3D models. They utilize a ray tracing concept to analyze agent-based visibility in the 3D environment. To do that, they represent 3D objects as voxels and perform the ray tracing on a voxel grid as well; essentially performing the comparison on voxels.

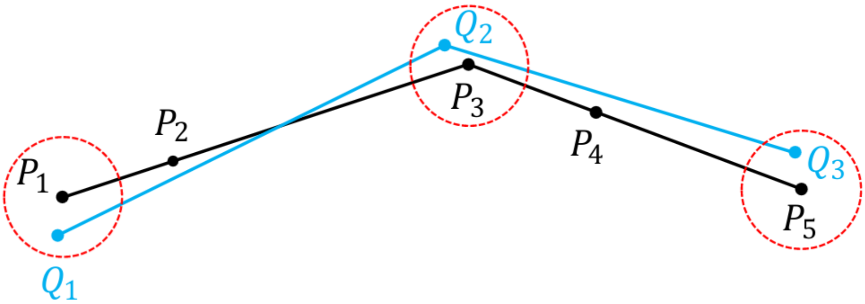


Figure 2.19 Two different representations of the same LineString in CityGML. Based on the chosen error tolerance (red), points may be considered *unchanged* or *changed*. Source: [Nguyen & Kolbe, 2021].

2.4.4 Cloud-to-model change detection

Recently, much research has been devoted to investigating methods for as-planned and as-built states in the context of the built environment. Such comparison necessitates the juxtaposition of different 3D representations, for example, point clouds and 3D models [Stilla & Xu, 2023]. Tuttas et al. [2015] propose a probability-based method of progress monitoring in construction sites to compare an as-planned state derived from a BIM model to an as-built state represented by a photogrammetric point cloud. The model

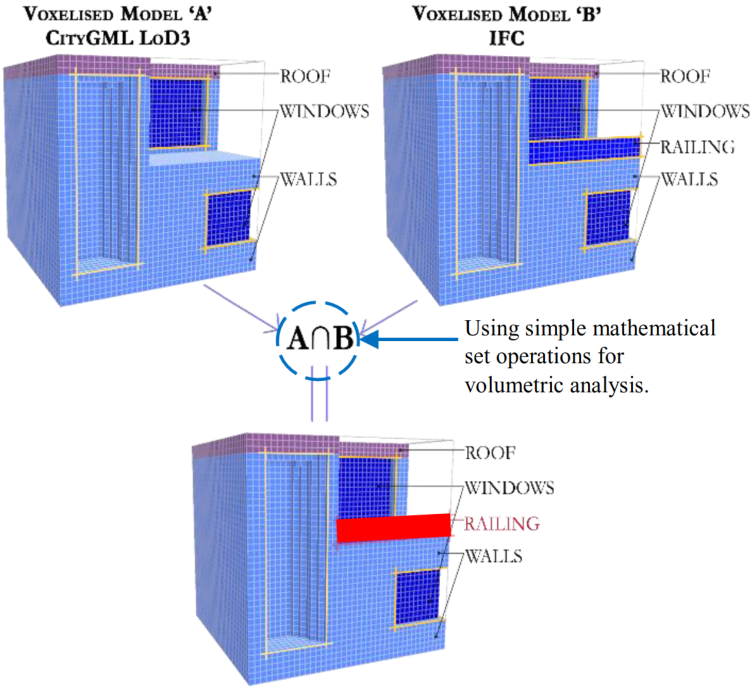


Figure 2.20 Constructive solid geometry operation to identify changes in voxel-represented building models; red indicates changes. Source: [Heeramaglore & Kolbe, 2022].

plausibility range is designed as a hyperparameter of bounding box size (Figure 2.21), and the final confirmation or conflict decision is based on a fuzzy-logic approach. They also elaborate on the challenges they encountered in such multimodal comparisons.

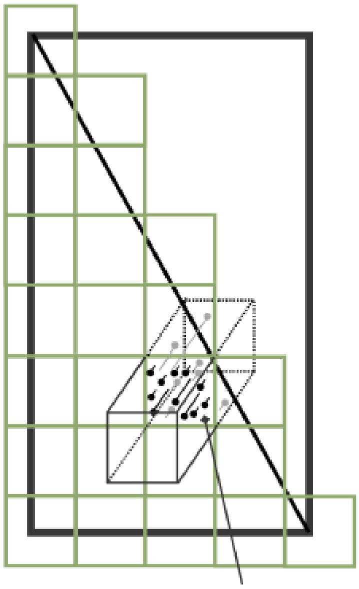


Figure 2.21 Point clouds in bounding boxes compared to the rasterized (green) 3D model. Source: [Tuttas et al., 2015].

Another approach is presented by Meyer et al. [2022], who decide upon a change based on DST theory (subsection 2.3.2). They discretize space into a voxel grid where multiple pieces of evidence are combined both for TLS measurements and BIM model. For TLS they address common measurement errors, while for BIM model, the associated LoA and its standard deviation are extracted.

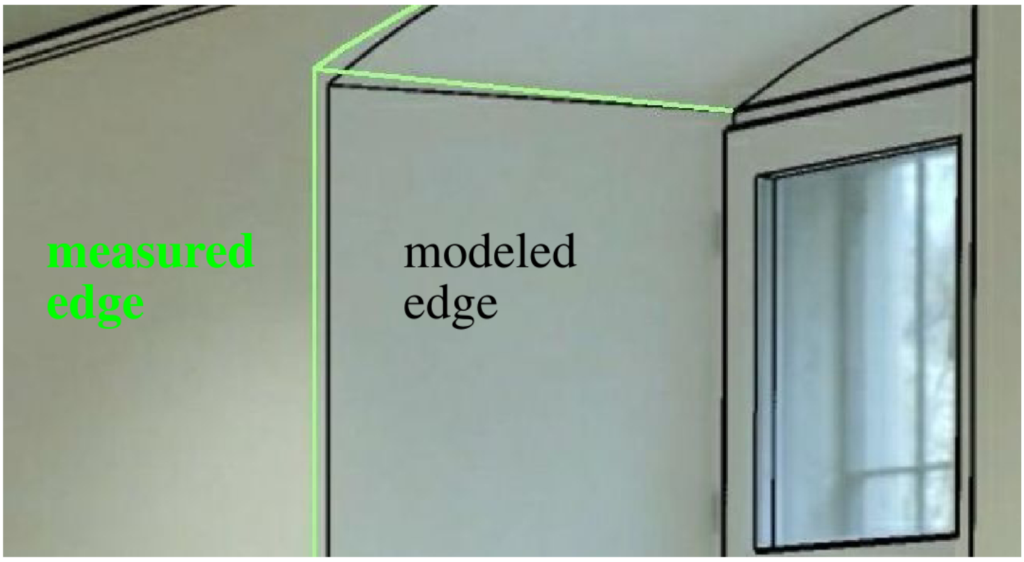


Figure 2.22 Measured edge exhibits some deviations to the modeled edge. The uncertainty estimation of both the 3D model and measurement allows us to decide whether it is a real deviation (change) or inaccuracy (no change). Source: [Meyer et al., 2021].

However, BIM models are distinct from semantic 3D city models, especially concerning the geometrical representation; since BIM components are typically generated using volumetric geometries, semantic 3D city models are frequently represented by outer-observable surfaces [Kolbe & Donaubauer, 2021].

2.5 3D semantic segmentation

The sensor-acquired 3D point clouds represent the geometry of the captured environment. However, to create a machine-readable input, each 3D point shall have an assigned meaning. In this Section, works related to semantic segmentation are described, detection, and classification on 3D point clouds focusing on deep and machine learning solutions, which recently have gained wide photogrammetry and computer vision communities' attention [Griffiths & Boehm, 2019a; Li et al., 2020].

Deep learning is a specialized form of machine learning that focuses on using deep neural networks to automatically learn scene representations from data. While it has achieved impressive results in various complex tasks, traditional machine learning still proves efficient in many practical scenarios where data may not be as abundant or when the interpretability of the model's decisions is crucial. The choice between deep learning and other machine learning methods depends on the specific problem and the available resources and data.

2.5.1 Random forests approach

Random forests is an example of a machine learning algorithm relying on hand-crafted features. In the seminal work Weinmann et al. [2013] propose and analyze key geometric features for point cloud segmentation training; Selected geometric features are visualized in Figure 2.23. Such features rely on the characteristic of point clouds for specific objects within a local neighborhood; for example, the planarity feature encodes wall-like surfaces while omnivariance encodes specific 3D distribution of points in the neighbourhood [Grilli & Remondino, 2020]. The common geometric features are:

$$\text{Linearity: } L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_1} \tag{2.19}$$

$$\text{Planarity: } P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1} \tag{2.20}$$

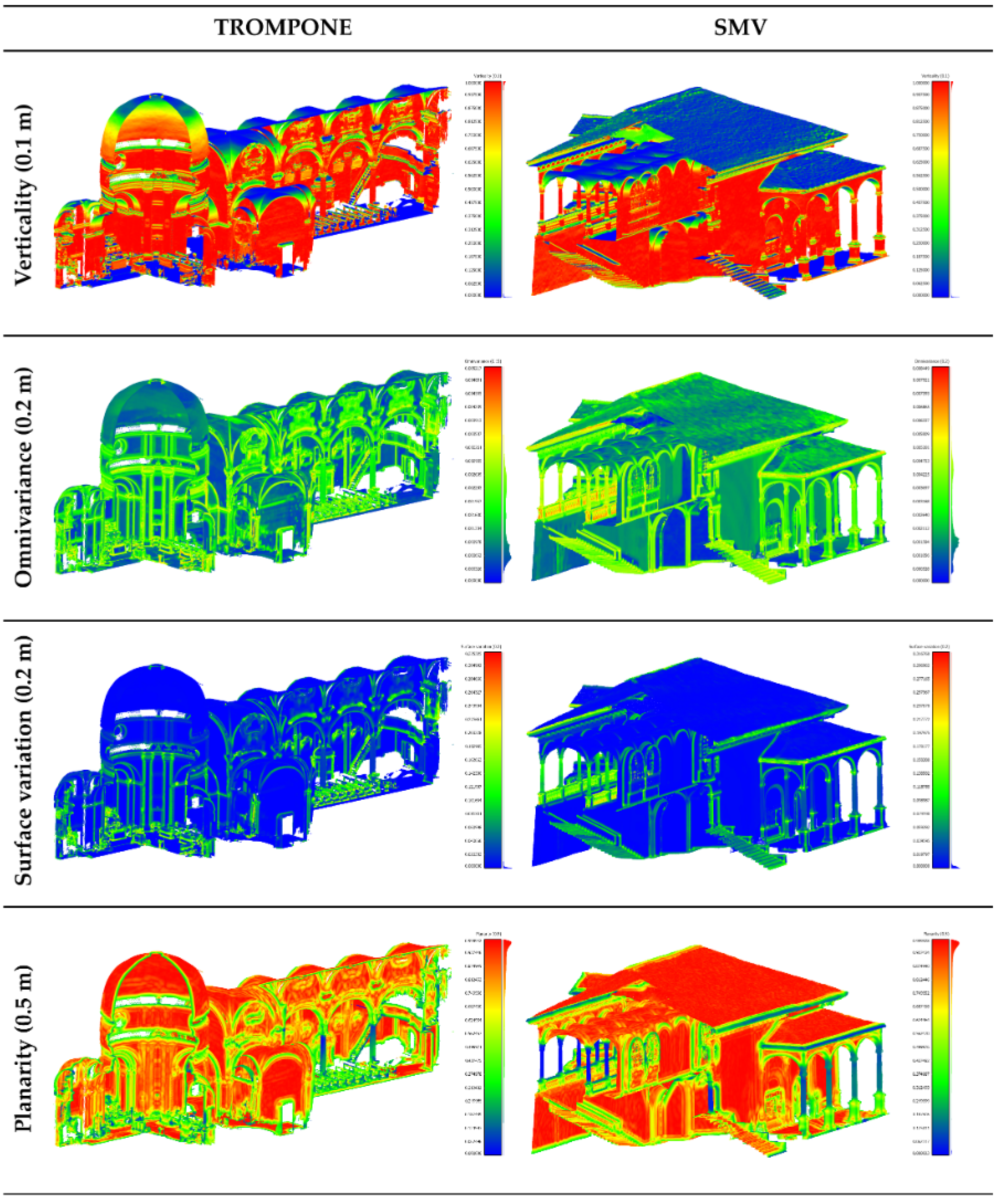


Figure 2.23 Selected geometric features at different radii, calculated for cultural heritage buildings (Trompone and Sacro Monte Varallo, Italy). Source: [Grilli & Remondino, 2020].

$$\text{Scatter: } S_\lambda = \frac{\lambda_3}{\lambda_1} \tag{2.21}$$

$$\text{Omnivariance: } O_\lambda = \sqrt[3]{\lambda_1 \lambda_2 \lambda_3} \tag{2.22}$$

$$\text{Anisotropy: } A_\lambda = \frac{\lambda_1 - \lambda_3}{\lambda_1} \tag{2.23}$$

$$\text{Eigenentropy: } E_\lambda = \sum_{i=1}^{3} \lambda_i \ln(\lambda_i) \tag{2.24}$$

$$\text{Sum of eigenvalues: } \Sigma_\lambda = \lambda_1 + \lambda_2 + \lambda_3 \tag{2.25}$$

$$\text{Curvature: } C_\lambda = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \tag{2.26}$$

$$\text{Verticality: } V = 1 - n_Z \tag{2.27}$$

$$\text{Density: } D = \frac{k+1}{\frac{4}{3}\pi r_{k\text{-NN}}^3} \tag{2.28}$$

As Grilli & Remondino [2020] show, such geometric features provide promising results in the context of cultural heritage building classification. They identify a set of features supporting such a task for specific labeled point cloud classes (floor, façade, column, arch, vault, window, molding, drainpipe, other). They utilize the concept of random forests to achieve that and conduct tests on six historical scenes located in Europe. They provide metrics Precision, Recall, and F1-Score, mostly oscillating around 0.8 for all but window and door classes where it is around 0.5, indicating limitations of the proposed approach.

2.5.2 View-based models

View-based models are a group of early deep learning 3D segmentation algorithms exploiting advancements of 2D segmentation. Especially in the domain of CNN where architectures are already robust and obtain high-accuracy results. Moreover, some pre-trained models can be utilized. To use this knowledge, the 3D representation of the point cloud is transformed into 2D images mimicking acquisition from various positions of a depicted object. One of the key advantages in comparison to other model types is that the multiview strategy can be applied in order to vote for the final prediction. Essentially, it combines knowledge from various views to obtain the most probable vote and minimize false assignments. Therefore, the view-based models are more efficient and have many datasets to train on already established as well as well-developed architectures. On the other hand, the projection from 3D to 2D space is a loss of information. Furthermore, multiviews may cause unnecessary information redundancy (Figure 2.24).

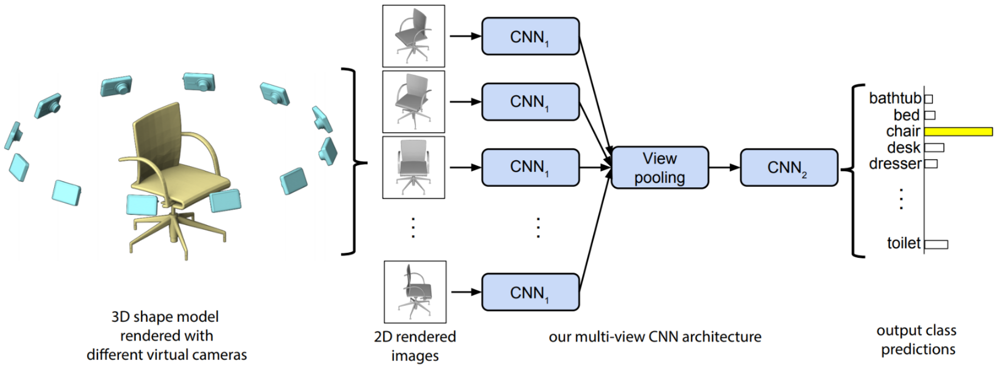


Figure 2.24 View-based model on an example of MVCNN [Su et al., 2015]. The 2D images generated with virtual cameras are acquired and then processed by independent CNN. These extract feature image feature maps from different cameras and then aggregated into a single feature map that describes a 3D object. This aggregated feature map is further passed into a second CNN for the classification. Source: [Su et al., 2015].

Multiview CNN (MVCNN) designed by Su et al. [2015] consists of two models simulating camera acquisition from 12 and 80 views. The views are learned separately and then fused, where a final voting is performed. In comparison to volumetric representation, the approach is much more efficient. The downside of the method is the lack of information preservation of the whole object. MVCNN-MultiRes presented by Qi et al. [2016] is an enhancement of the previously mentioned model. The main difference in this approach is a changed type of projection from 3D to 2D views. Also, multiresolution 3D filtering is used to enable more accurate 3D object capturing as well as sphere rendering, which makes the model more robust than its predecessor. An alternative approach is introduced by Dai & Nießner [2018] in the 3DMV network, which introduces the idea of creating a joint architecture combining images and geometry. The images are first mapped onto a 3D feature from a volumetric grid. Then, a multiview strategy is applied to extract meaningful parts. The model has shown promising results in 3D object classification. Yet, the model needs two sources of data, which increases the computational cost of the architecture. On the other hand, the RotationNet solution is based on the assumption that a potential observer should recognize the object even from a partial dataset [Kanezaki et al., 2018]. From the multiview images of an object, a set of photos is selected where a prediction of an object category is performed based on rotation. Then, the best pose is picked to maximize the likelihood of the object's category. The image viewpoints are estimated to predict the pose of the object. The model achieves results similar to MVCNN but with a smaller number of images. One of the architecture's limitations is that each image should have a fixed viewpoint.

2.5.3 Voxel-based models

Voxel-based models represent a deep learning approach where the input point cloud scene is transformed into a voxel representation. According to the study of Xu et al. [2021], the voxel structure in the construction industry research is applied at many stages of data processing, such as pre-processing (15%), registration (10%), modeling (28%), but mainly to segmentation and classification (47%). Voxels are derived from 3D point cloud datasets, with a certain size typically related to the input data resolution and describe the data distribution in the 3D space. Owing to that operation, convolutions and pooling operations can be applied as they operate on ordered datasets. However, this approach does not exploit 3D geometrical patterns and surfaces. Another downside of the voxel-based models is ineffective data representation, which stores not only occupied spaces but also unoccupied ones. Also, to set an appropriate grid cell size, one needs to have prior knowledge about the dataset and recognize possible disruptions in spatial relations resulting from voxels that are too large or too small. Moreover, the computational requirements increase cubically with the resolution of the dataset, limiting its application for large outdoor scenes. The advancements that minimize those disadvantages are octree-based grids, which define cubes of adaptive size to create a hierarchical data structure that decomposes 3D point cloud dataset to multiple leaf's nodes, i.e., voxels (Figure 2.25); This highly reduces the computational cost of algorithms.

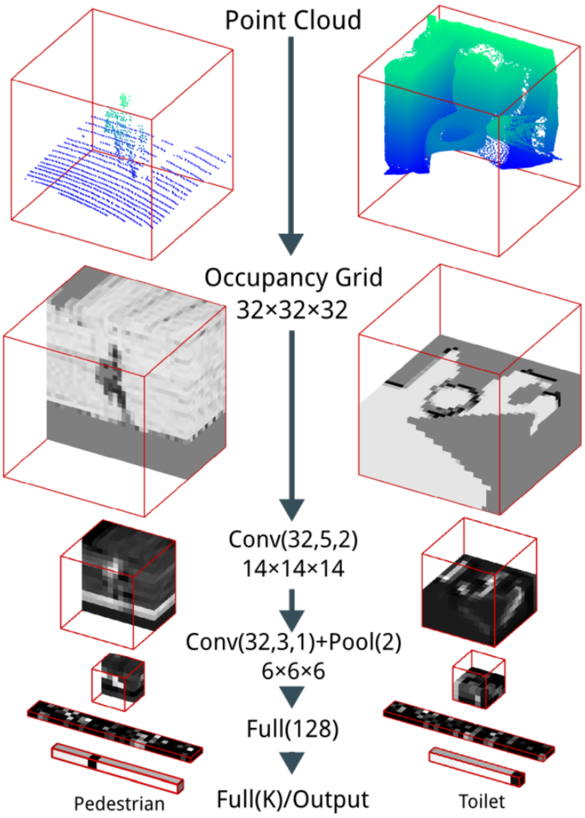


Figure 2.25 VoxNet is a 3D CNN architecture utilizing Convolutional (Conv), Pooling (Pool), and Fully Connected (Full) layers. Conv(f, d, s) involves f filters of size d and stride s. Pool(m) represents pooling with an area of m. Full(n) denotes a fully connected layer with n outputs. Source: [Maturana & Scherer, 2015].

3-D ShapeNet is one of the pioneering frameworks utilizing this data representation [Wu et al., 2015]. The method strongly exploits CNN and performs very well given low-resolution voxels. The scalability of the solution stands as a problem because of cubically increasing computational costs. Thus, it limits its application to large-scale or dense point clouds. The VoxNet model introduced by Maturana & Scherer [2015] introduces lighter architecture. However, there exist grids containing no meaningful information, which again increases the computational cost. Yet another approach is introduced by Wu et al. [2016] in the 3D-GAN network. They use a framework fusing GAN (general adversarial network) and volumetric convolutional networks. The main advantage of the model is unsupervised training in 3D object recognition, which yields a reliable shape descriptor. On the other hand, capturing detailed features is strongly dependent on the density of the input point cloud, which proves the method credible only on evenly distributed datasets. OctNet presented by Riegler et al. [2017] utilizes octree-based data representation, creating unbalanced octrees based on the input points density. The model learns the structure of the tree and occupancy values.

2.5.4 Graph-based Models

Yet another idea exploits graph representation to perform network training on point clouds. 3D point clouds are encoded as a type of non-Euclidean dataset where each node of a graph is a point of a point cloud, and edges show relationships between points in the point cloud. Neural networks used in graphs enable the propagation of node states till the balance is reached. Additionally, CNN can be applied, which enables immediate operations in the spectral and non-spectral domain on neighboring nodes. One of the most important characteristics of graph-based models is the exploitation of neighboring geometric relations between points, which in turn enables the extraction of local patterns. On the other hand, utilization of the relations is a challenging task. Furthermore, one needs to define an adaptive operator for varying sizes of neighborhoods. Also, the graph structure has to be tailored to CNNs needs.

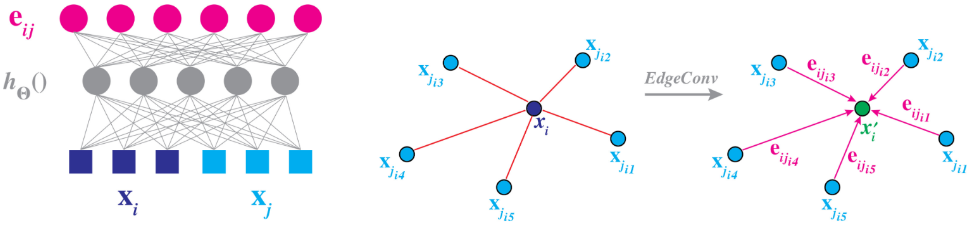


Figure 2.26 The left image illustrates the computation of an edge feature, e_{ij}, from a point pair, x_i and x_j; h_θ represents a fully connected layer, and its learnable parameters are the associated weights. The central and right images demonstrate the EdgeConv operation. The output of EdgeConv is obtained by aggregating the edge features related to all the edges that originate from each connected vertex. Source: [Wang et al., 2019].

One of the examples of such an approach is SyncSpecCNN, where the backbone is exploiting shared coefficients, multiscale graph analyzing, and sharing information among distinct but related graphs [Yi et al., 2017]. The convolutions are performed in the spectral domain. The model proved to be efficient in 3D parts segmentation. However, it is limited in terms of computational efficiency, missing local edges, and is highly dependent on once-learned coefficients (basis-dependent). Simonovsky & Komodakis [2017] utilize spatial domain to construct respective convolutions in their edge-conditioned convolution (ECC) network. The ECC model achieves scalability and efficiency through the learning of dynamic local patterns of neighboring points. However, the model is performing well only on small datasets as it is computationally costly. Dynamic graph CNN (DGCNN) presented by Wang et al. [2019], similarly to ECC, exploits spatial domain characteristics to perform convolutions. The novelty in the approach is the dynamic update of the fixed graph size (Figure 2.26). Thus, the DGCNN adapts better to the data size than the predecessor ECC. The architecture performs the calculation of joint scores of local and global features. DGCNN yields distinctive edges that have a fixed size that limits their utilization in different scales and resolutions of input data. The attention mechanism is also proposed for graphs. For example, Graph Attention Networks (GAT) proposed by Veličković et al. [2017] overcome the limitation of isotropy of input features for aforementioned graph-based models, owing to the attention mechanism. The attention enables steering an algorithm to the most important parts of the graph rather than treating each node's group equally. The main goal of this method is to trigger the self-attention mechanism, which obtains a hidden representation of each node by varying weight values.

2.5.5 Point-based models

All the previous methods require circumventing the unstructured nature of 3D point clouds to enable efficient learning. Yet, the intuition suggests capitalizing on the raw structure and starting network training directly on unordered points: Point-based models operate directly on an unordered point cloud structure. Therefore, they preserve the raw spatial local structure and 3D data representation. This trait limits the operation costs only to captured extents. One of the main challenges of those models is solving permutation problems. Achieving the permutation invariance in the unordered set of points results in the rejection of the spatial relationship between points and the influence of neighbors. This characteristic is perceived as a fundamental problem that limits the detection of local features.

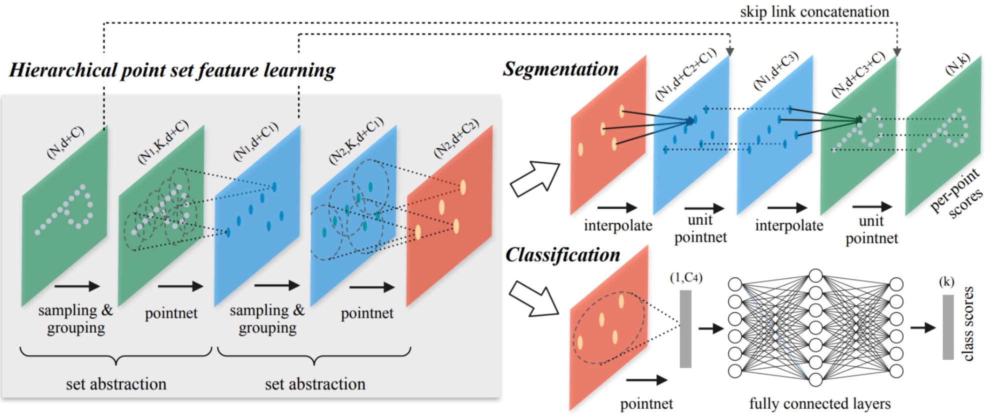


Figure 2.27 The PointNet++ architecture introduces hierarchical feature learning to capture features at multiple scales from 3D point clouds. This is achieved by using sampling and grouping layers, which define and extract neighborhoods of points at different sizes. These point neighborhoods are then passed through mini-PointNet architectures to extract informative features. Source: [Qi et al., 2017b].

The PointNet presented by Qi et al. [2017a] and its successor PointNet++ proposed by Qi et al. [2017b] are seminal works in direct training of 3D point clouds. PointNet proves to be a reliable architecture, which, however, is limited in preserving local features. This is rectified by PointNet++, which addresses the problem of local patterns and complex scenes (Figure 2.27). Albeit the PointNet++ architecture outperforms its predecessor, the local spatial relation between neighboring points still poses a challenge. PointCNN of Li et al. [2018] exploits convolutions to solve permutation and transformation problems. The local correlations are created, which create a hierarchical CNN network architecture. However, this approach does not use correlations of varying geometric features. KPConv is based on the concept of kernel point convolution, which extends the traditional convolution operation to irregular and unstructured 3D point clouds [Thomas et al., 2019]. Another approach, underscored by the pioneering Point Transformer network, is the usage of the self-attention mechanism [Zhao et al., 2021]. It is based on the Transformer architecture originally proposed for natural language processing tasks. The Point Transformer applies self-attention mechanisms to unordered point clouds, allowing it to capture long-range dependencies and contextual information effectively.

2.5.6 Outdoor 3D point cloud segmentation benchmark datasets

In data-driven approaches, such as deep learning (DL) and machine learning (ML), the high-quality data is pivotal. Benchmarks allow for the comparison of different algorithms' efficiency and yield training data.

The 3D semantic segmentation deep learning methods rarely focus on building or building parts segmentation. Dwellings are frequently annotated as buildings or as man-made structures without distinguishing for miscellaneous building parts, such as windows or doors. Only a few benchmarks split a building into constituent parts, and even then, subgroups describe buildings as façades, which is rather a syntactical difference than semantics; as anyway, that kind of benchmarks primarily cover only lower parts of buildings envelope (i.e., up to first floor).

As depicted in Table 2.4, to the best of my knowledge, only one outdoor benchmark, Architectural Cultural Heritage (ArCH), released by Matrone et al. [2020] has classes that fully capture greater details than a building envelope, such as stairs, gate, or façade ledge. Yet, the dataset is application-specific and concentrates on classifying distinct landmark buildings, which may prove not scaleable for the usual dwellings. There are 17 annotated and ten non-annotated scenes acquired through TLS, UAV-based, and terrestrial photogrammetry techniques. The classes are labeled according to IFC and CityGML classes; the semantic classes are shown in Table 2.1.

Table 2.1 Cultural-heritage-relevant classes in the ArCH dataset

No.	Architecture elements
1.	Arch
2.	Column
3.	Moldings
4.	Floor
5.	Door and window
6.	Wall
7.	Stairs
8.	Vault
9.	Roof
10.	Other

In the remainder of this Section, I present the primarily urban-related benchmarks, which, after some alteration, might be used as complementary benchmarks for 3D façade semantic segmentation; The complete analyzed list is shown in Table 2.4.

Zhu et al. [2020] introduce TUM-MLS-2016 containing 40M annotated points in an urban area. Each vector point contains X, Y, Z, and Intensity. Also, annotated instances of individual objects, and labeled points in a sequence of 360-degree scans are provided. The depicted area is a university campus located in Munich, Germany. Thus, the dataset depicts Central European urban architecture. Furthermore, most of the labeled points correspond to the building class (37%, 57%, and 74% in each subsection), which eventually estimates less than 20M points representing building class, which is almost 50% of the whole point-wise labeled benchmark. The creators of the dataset emphasize that the dataset has been created to obtain complete coverage of building façades, which is not the case for other available benchmarks. Ultimately, eight (and unclassified) classes are present in the benchmark (see Table 2.2). Similar eight (and unlabeled) classes are used in the semantic3D.net benchmark [Hackel et al., 2017], which comprises a large dataset of static (TLS) scans. The total dataset contains around 4BN points acquired in 30 non-overlapping scenes located in urban, suburban, and rural areas. The scenes are located in Central Europe and thus represent typical architecture for this region. One of the most represented classes in the dataset is building. Around 300M are designated for training, while more than 600M are for testing. Each vector point contains X, Y, Z, Intensity, and RGB values.

The pioneering work in benchmarking terrestrial and urban point clouds is the Oakland 3D dataset released by Munoz et al. [2009]. Each vector point contains X, Y, Z value; The dataset was acquired by MLS at a university campus in Oakland, USA, representing the North American architecture style. However, it contains merely 1.6M labeled points in the dataset with highly underrepresented 44 classes: The labels are comparatively detailed to other datasets; however, the point cloud is rather sparse. Besides

Table 2.2 Urban-relevant classes in the TUM-MLS-2016 dataset

No.	Urban elements
1.	Man-made terrain
2.	Natural terrain
3.	High vegetation
4.	Low vegetation
5.	Buildings
6.	Hardscape
7.	Scanning artefacts
8.	Vehicles
9.	Unclassified

Table 2.3 Early works on façade segmentation comprise sparse point clouds, on an example of the Oakland 2009 dataset

No.	Façade classes and its point count
1.	Façade (ca. 100 000 points)
2.	Wall (ca. 10 000 points)
3.	Stairs (ca. 500 points)
4.	Fence (ca. 800 points)
5.	Gate (ca. 100 points)
6.	Ceiling (ca. 700 points)
7.	Façade ledge (ca. 1100 points)
8.	Column (ca. 1100 points)

having classes such as crosswalk light, paved road, and foliage, it has classes directly or indirectly related to buildings (in total approximately 100,000 points), as shown in Table 2.3. Besides the aforementioned datasets that serve general purposes of detection or semantic segmentation tasks, there are datasets tailored for testing algorithms of façade segmentation, which are primarily image-oriented, thereby limiting testing of 3D point cloud façade segmentation algorithms. Nevertheless, such datasets are frequently a complementary source of evaluation; as such an excerpt about them is shown in the reminder of this section.

The most comprehensive image-driven work to date is introduced by Tyleček & Šára [2013], who present the CMP Façade Database comprising 606 images of annotated building façades. The façades are acquired in various cities around the world and encompass many architectural styles. The semantic classes are shown in Table 2.5. Another influential work entitled ICG Graz50 shown by Riemenschneider et al. [2012] contains 50 images with annotation of façades in Classicism, Biedermeier, Historicism, Art Nouveau, and post-modern architectural styles. The representation of single façade is built from about 30 perspective images and projected to a single orthographic image each. The geometry should be piecewise planar, but due to approximations, artifacts are present; The semantic classes are shown in Table 2.6. Also worth mentioning is the seminal eTRIMS database [Korc & Förstner, 2009] containing 60 non-orthorectified images of façades and their vicinity. Therefore, the classes include also non-façade-related objects, such as cars (see Table 2.7).

Other datasets follow a similar pattern in class distribution and selection. For example, a dataset comprising façades is located in Paris, France: the Paris Art Deco Façades Dataset presented by [Gadde et al., 2016], which contains 79 images of buildings in Art Deco style. The ETHZ CVL RueMonge 2014 dataset [Riemenschneider et al., 2014] extends the concept of sole images in the database to also accommodate image-based meshes. The dataset consists of 60 buildings and 428 respective images covering 700

Table 2.4 Point cloud benchmark datasets with a potential for façade and urban segmentation methods testing.

Name	Year	Sensor	Scalar fields	World	# points	# Classes	Façade-level classes?
Oakland 3D [Munoz et al., 2009]	2009	MLS	X,Y,Z	real	1.6 M	44	~
ETH PRS [Lande, 2012]	2012	TLS	X,Y,Z,I	real	✗	0	✗
Sydney Urban Objects Dataset [De Deuge et al., 2013]	2013	MLS	X,Y,Z,I	real	✗	26	✗
Paris-rue-Madame database [Serna et al., 2014]	2014	MLS	X,Y,Z,I	real	20 M	27	~
iQumulus [Vallet et al., 2015]	2015	MLS	X,Y,Z, I	real	12 M	8	✗
TUM-MLS-2016 [Zhu et al., 2020]	2016	MLS	X,Y,Z,I	real	1.7 BN	9	✗
semantic3D.net [Hackel et al., 2017]	2017	TLS	X,Y,Z, I, RGB	real	4 BN	9	✗
Paris-Lille-3D [Roynard et al., 2018]	2018	MLS	X,Y,Z,I	real	143 M	50	✗
SynthCity [Griffiths & Boehm, 2019b]	2019	MLS	X,Y,Z, RGB,N	synthetic	368 M	9	✗
A2D2 [Geyer et al., 2020]	2020	MLS	X,Y,Z,I	real	387 M	38	✗
ArCH [Matrone et al., 2020]	2020	TLS/MLS/UAV/TP	X,Y,Z, RGB, N	real	136 M	10	✓
Toronto-3D [Tan et al., 2020]	2020	MLS	X,Y,Z,I, RGB	real	78 M	8	✗
Whu-TLS [Dong et al., 2020]	2020	TLS	X,Y,Z,I, RGB	real	551 M	0	✗
BIMAGE Datasets [Blaser et al., 2021]	2021	MLS	X,Y,Z, RGB	real	840 M	0	✗
KITTI-360 [Liao et al., 2021]	2021	MLS	X,Y,Z, RGB	real	1 BN	19	✗
Paris-CARLA-3D [Deschaud et al., 2021]	2021	MLS	X,Y,Z,I,RGB	real+synthetic	60 + 700 M	23	✗

Table 2.5 Façade-relevant classes in the image-based benchmark CMP

No.	Façade-relevant classes
1.	Façade
2.	Molding
3.	Cornice
4.	Pillar
5.	Window
6.	Door
7.	Sill
8.	Blind
9.	Balcony
10.	Shop
11.	Deco
12.	Background

Table 2.6 Façade-relevant classes in the image-based benchmark ICG Graz50

No.	Façade-relevant classes
1.	Door
2.	Window
3.	Sky
4.	Wall
5.	Shop
6.	Balcony
7.	Roof

Table 2.7 Façade-relevant classes in the image-based benchmark eTRIMS

No.	Façade-relevant classes
1.	Building
2.	Car
3.	Door
4.	Pavement
5.	Road
6.	Sky
7.	Vegetation
8.	Window

meters along Rue Monge Street in Paris, annotated with pixel-level labels for 2D (image) and 3D (mesh) representation.

2.6 3D semantic segmentation for 3D façade reconstruction

This section is devoted to approaches of 3D semantic segmentation leading to 3D façade reconstruction with the focus on façade elements; It is created on the basis of my first-author publication: Wysocki et al. [2022b].

Substantial research effort has been devoted to detecting and reconstructing façade elements [Musialski et al., 2013]. Methods typically focus on such façade elements as windows [Tuttas & Stilla, 2013; Aijazi, 2014; Becker, 2011; Iwaszczuk et al., 2011], windows and doors [Ripperda, 2010; Riemenschneider et al., 2012], and balconies [Fan et al., 2021]. Since many assumptions in reconstruction stem from the input modality, the remainder of this Section is organized by dividing it into modality-groups.

2.6.1 Laser-based

Tuttas & Stilla [2013] provide insights into both windows detection and reconstruction with sparse oblique airborne laser scanning (ALS) point clouds. The presented approach utilizes the so-called *voyeur effect*: the phenomenon of laser pulses passing through windows and reflecting behind a façade (see Figure 2.28). Based on that assumption, indoor points can be projected to the known (building model) or found plane (mixture of region growing and RANSAC algorithms). The projection takes into account the incidence angle as well as a threshold for indoor points. The windows are reconstructed as 2D rectangular vector objects. Another work of Tuttas & Stilla [2012] also utilizes *voyeur effect* in order to detect and reconstruct a window within a façade. Here, the approach assumes that the center of a window is known (obtained from indoor points). Then, the Quadrant-Based Search is applied to find the maximum ranges of windows. The probabilistic density function (PDF) searches for the most probable edges not only for singular edges but also considering schematic localization of windows and thus windows' vector lines on a façade. Hoegner & Gleixner [2022] pursue the idea of using *voyeur effect* specifically for MLS point clouds. Besides ray intersections, they also analyze empty regions in point clouds. Due to the methods' assumption that each visible opening is a window, they do not distinguish between other openings, such as doors or underpasses. Also, their performance is limited in the presence of laser-impenetrable window elements, such as window blinds.

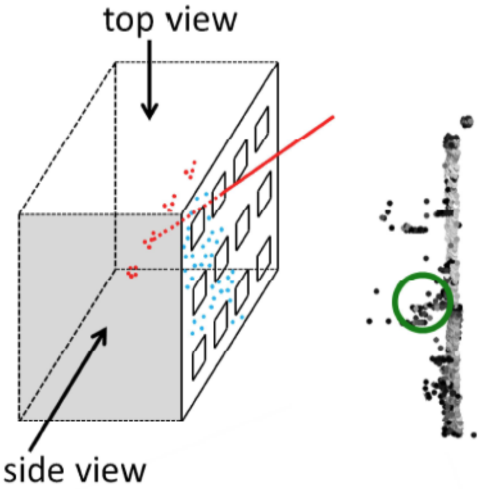


Figure 2.28 Intersection of laser rays with building planes exploited for the reconstruction of windows as bounding boxes. Source: [Tuttas & Stilla, 2013].

Aijazi [2014] propose an alternative approach to project 3D points to 2D plane, which results in a problem simplification. The windows are then found based on an evaluation of watertightness, assuming that windows are holes in a surface. Then, the parameters of windows are calculated based on symmetrical and temporal correspondences in successive passages. Firstly, the kernel searches for symmetrically corresponding objects in rows and columns and repeats the step for previous passages.

Another research direction is explored by Fan et al. [2021]. They present a method based on Gestalt and architectural form principles for the reconstruction of windows, doors, and balconies. The clusters are formed on the basis of Gestalt principles, while the principle of architectural form supervises understanding. The method is closely related to previous works based on façade grammars. The optimal layout graph model is found using a mixture of segmentation (RANSAC), likelihood and probability, and finally, the Bayesian framework.

Work relating to underpasses has been undertaken by Gargoum et al. [2018], where they assess the vertical clearance of overhead assets such as bridges. Their solution is based on LiDAR point clouds acquired in MLS campaigns on highways. The method first detects overhang objects as possible candidates and classifies them into either bridge or non-bridge structures. The selection is based on the assumption that a point measured above a vehicle trajectory suggests an existing overhang asset; a bridge represents a denser region than a non-bridge asset.

2.6.2 Laser- and camera-based

Becker [2011]; Becker & Haala [2007] focus on the façades (windows and doors) reconstruction based on a cell decomposition. They assume that the primitives detected on the part of a building can yield information about the overall building's structure. Thus, the generation of the whole building's façade is possible. The first proposal is made based on cell decomposition, and then, to enhance uncertain window position and size, the neighboring windows adjust their width and height. However, it has been observed that even further enhancements might be added using images. For instance, sashes (crossbars) can be found. Due to the acquisition geometry, not all of the crossbars can be found, which is again adjusted by neighboring windows.

A similar assumption about the organized layout of façade elements is followed by Ripperda [2010] who focus on window and door detection. The grammar which subdivides façade into smaller parts is introduced. The derivation process is controlled by rjMCMC, as shown in Figure 2.29. Distributions for the algorithm are taken from around 400 manually digitalized façades' photos. The rjMCMC also allows to include shape changes of a façade. Probability scoring functions are based on depth and color.

2.6.3 Camera-based

Early learning-based façade segmentation methods [Szeliski, 2010; Tyleček & Šára, 2013; Gadde et al., 2016; Riemenschneider et al., 2012] typically rely on ubiquity of 2D image façade segmentation datasets and represent façade elements as 2D objects [Musialski et al., 2013]. While achieving excellent performance, these methods require additional processing to reconstruct the façade in 3D [Pang & Biljecki, 2022].

rjMCMC has also found its application in solely image-based façade understanding, not only in hybrid solutions described in the preceding section (subsection 2.6.2) [Mayer & Reznik, 2007]. Unlike standard MCMC, rjMCMC allows for dynamic adjustments in the number of parameters, facilitating both the introduction and removal of objects and relationships, a feature dubbed reversible. According to Mayer [2008], there are two most important features of rjMCMC in the context of object extraction: Firstly, it accommodates known and unknown uncertainty regarding the existence and quantity of objects and relationships in the modeling process; Secondly, it allows for distribution sampling, which in turn unlocks the possibility of simulating objects and their relationship, eliminating the need for data-based analysis.

Iwaszczuk et al. [2011] focus on a window detection. They utilize textures created from infrared (IR) images. The local dynamic threshold and morphological operations operation serve to detect a window candidate. Then, they assume that a window is rectangular and appears on a homogeneous background

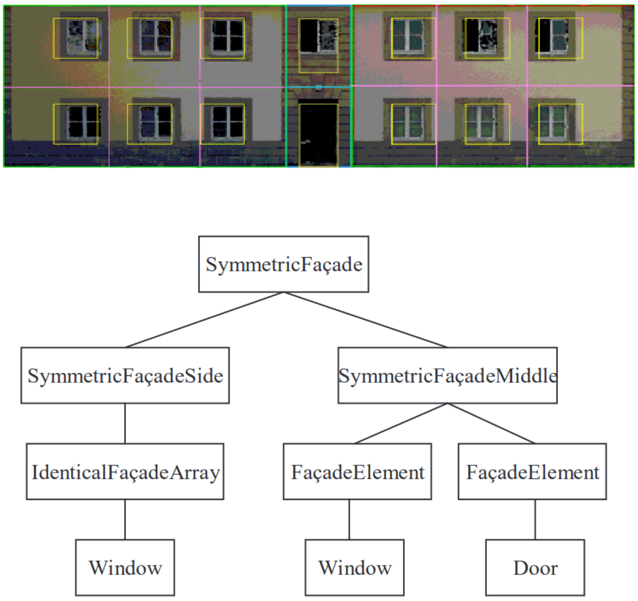


Figure 2.29 Exemplary derivation tree and the respective façade-based on rjMCMC. Adopted from: [Brenner & Ripperda, 2006].

that is distinct from the wall. Based on that assumption, a window model is proposed, which serves later as an adaptive mask that searches for window edges by analyzing intensity changes. Riemenschneider et al. [2012] focus on windows and doors detection. Their solution is based on the Cocke-Younger-Kasami (CYK) algorithm, which iterates over all elements of the grammar. They fuse low-level classifiers and mid-level object detectors. Also, they tailor CYK-based algorithms to support symmetric and repeating structures and introduce an irregular lattice model.

Recent works utilize efficient 2D image-based neural networks to identify façade elements in images and then project them onto 3D point clouds or their derivatives, such as 3D models [Huang et al., 2020; Pantoja-Rosero et al., 2022; Pang & Biljecki, 2022; Hensel et al., 2019]; the general approach is depicted in Figure 2.30. The seminal work of Nan & Wonka [2017] has accelerated this direction of research by providing generic and reliable out-of-the-box LoD2 building reconstruction open-source algorithm and software called PolyFit. Such 3D prior reduces task complexity and allows for photo-to-model projection. However, the method requires a 3D point cloud sufficiently covering the full object shape.

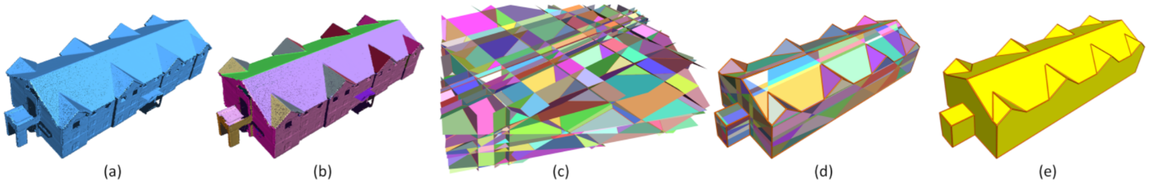


Figure 2.30 Footprint-free LoD2 reconstruction proposed by Nan & Wonka [2017] accelerates refinement-driven studies enabling out-of-box geometric prior reconstruction (see Figure 2.31). Proposed workflow: a) Point cloud b) Candidate planar segments (RANSAC) c) Proposed intersected faces d) Faces selected by solving optimization problem e) Reconstructed LoD2 models. Source: [Huang et al., 2020].

An ideal example following this strategy is the one of Huang et al. [2020], who propose a method employing FC-DenseNet56 [Jégou et al., 2017], trained with ortho-rectified façade images, to recognize façade openings. The labels are projected onto the LoD2 building model, which is reconstructed from a drone-based photogrammetric point cloud. The projected window and door labels are approximated to bounding boxes, which cut openings in LoD2 solids, thereby upgrading 3D models to LoD3. A similar approach is proposed by Pantoja-Rosero et al. [2022] where they explicitly use PolyFit [Nan & Wonka, 2017] for geometric prior extraction.

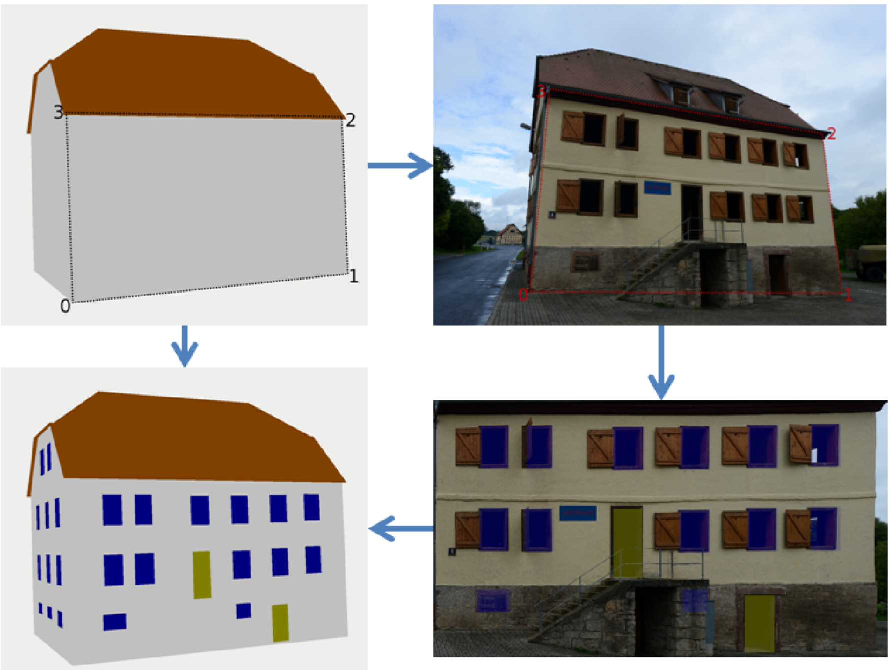


Figure 2.31 Recent trend in the refinement approach: LoD2 building model planes created using image-based point cloud (up-left corner) serve as projection targets for LoD3 features detected in images. Source: [Huang et al., 2020].

An alternative approach is shown by Hensel et al. [2019]; Zhang et al. [2019], where they follow the aforementioned workflow, but instead of image projections, they use already existing texturized 3D building models. Hensel et al. [2019] additionally accommodate for texture inaccuracies; they rectify the alignment of found opening bounding boxes by solving a mixed integer linear optimization problem (MILP). Wang et al. [2023] also defines the opening alignment as an optimization problem. In their case, they employ binary integer programming to align identified objects on texturized meshes. Worth noting that it is a new research direction utilizing deep learning networks to solve multidimensional optimization of building completion. For instance, Pang & Biljecki [2022] use a single-image reconstruction technique supported by LoD1 building models as priors to obtain a highly-detailed building model as a mesh.

3 Fusing 3D point clouds and 3D building models

Any processing of multimodal data requires a reliable fusion. In the case of geospatial data, such a framework is frequently established by data geo-registration and harmonization. In this chapter, a cloud-to-model fusion method based on publications [Wysocki et al., 2021a] and [Wysocki et al., 2021b] is discussed. The method exploits the uncertainties to determine the matching of point clouds and 3D building models.

3.1 Geometric model primitives for point cloud registration with uncertainty

The approach integrates MLS point clouds (point clouds) and semantic 3D building models (geometric model primitives), specifically addressing the challenge of uncertainty. The process begins with the determination of metadata and expert beliefs based on the input datasets. Subsequently, prior estimations are performed. The workflow then progresses to alignment and matching techniques, ultimately leading to the application of BayNet for identity estimation.

3.1.1 Uncertainty assignment

The initial and crucial stage in the data fusion process involves achieving a satisfactory alignment of diverse datasets. In this context, the data uncertainties refer to errors in the absolute 3D registration of the modalities. When dealing with façade-based registration, the uncertainty is quantified based on the absolute 3D point accuracy of footprints' and point clouds. Extracting this essential information is accomplished from the available metadata. Nevertheless, it is important to note that the metadata often lacks completeness, making it necessary to incorporate the expertise and judgment of skilled operators to overcome this limitation. In the matching context, coregistration accuracy is not the sole determinant of uncertainty. One such factor is LoD [Gröger et al., 2012], which specifies levels of richness of object details but only implies 3D point accuracy. Also, the reliability of the data used for semantic 3D city model creation may differ. For instance, for OpenStreetMap (OSM)-based city models, the credibility level should be approached differently depending on the location [Biljecki, 2020]. In practical situations, this confidence in the data and generalization level may be intertwined with coregistration deviations, and the resulting errors should be considered throughout the workflow. Therefore, the approach permits the incorporation of other uncertainty factors, as illustrated in Figure 3.1.

As illustrated in Figure 3.1, firstly, the alignment of semantic 3D models and MLS point clouds is performed. This operation consists of several sub-steps. The workflow starts with estimating CI based on the introduced parameters and data. Then, the feature coverage analysis is performed to select targets eligible for the matrix estimation that ultimately rectifies global position accuracy. As a result, the position is altered, and associated errors are passed to the matching process. Several pivotal uncertainties are defined for the MLS point clouds and semantic 3D models. The expert should incorporate the known error (e_1) of MLS point cloud global registration. The error is expressed as an anticipated maximal deviation from the true value. The certainty of the anticipated deviation is calculated based on the confidence level (CL_1) introduced by an operator. The upper bound and confidence level serve to find the standard deviation for the MLS measurement (σ_1). The same idea is employed to reference buildings. Also, building error is the assumed maximal discrepancy between the true value and a model (e_2). Confidence level (CL_2) expresses the belief in the error prior information. The found CI results in the estimation of the

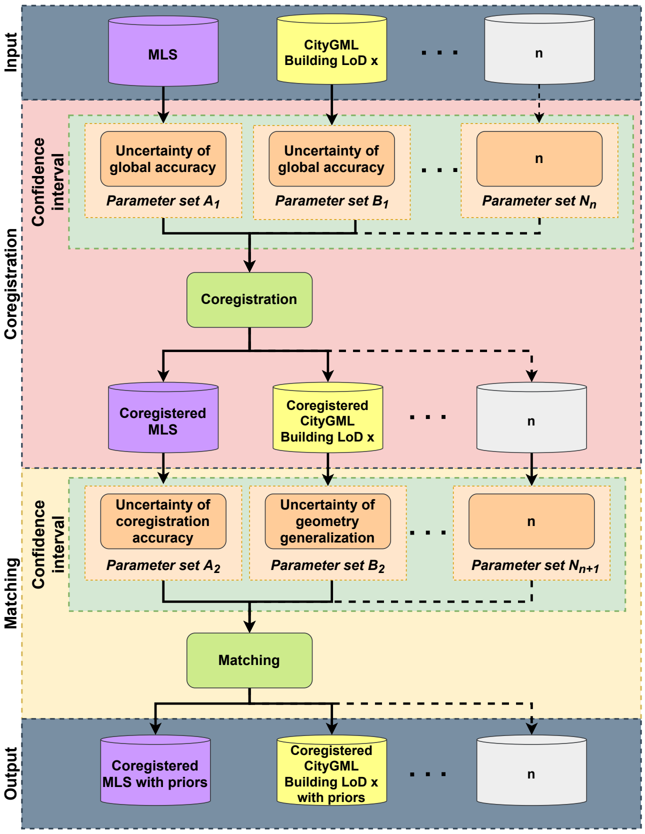


Figure 3.1 Estimation of priors: point clouds (purple), vector datasets (yellow), and defined uncertainties (orange) are the core elements of the process that is adaptable to new insights (gray).

standard deviation for the semantic 3D building model (σ_2). The total standard deviation (σ) is found using the formula [Suveg & Vosselman, 2000]:

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2} \tag{3.1}$$

The final upper and lower bounds are found based on $[\mu - 2\sigma, \mu + 2\sigma]$. This statement is correct for the inaccuracies assumed to represent Gaussian distribution and overestimate the bounds by operating in the L1 norm. The rationale of the presented estimations derives from the work of [Suveg & Vosselman, 2000]. The bounds might be perceived as 3D boxes in the discretized space as shown in Figure 3.2.

3.1.2 Coregistration

When confidence bounds are obtained, estimating the coverage of the building's features by MLS point clouds is feasible. The first estimation determining adequately covered building features is a density measure. It calculates the points per building feature within the introduced 3D boxes. Then, a threshold rejecting 3D boxes not reaching percentile value p_{val} is executed. The second tier measurement is the uniformity estimation introduced to neglect 3D boxes with small and densely populated concentrations of points. This measure is represented by a ratio $r \in [0, 1]$ of the MLS point cloud volume within the 3D box (v_1) to the total volume of the 3D box (v_2).

$$r = \frac{v_1}{v_2} \tag{3.2}$$

The estimated ratio should exceed a value of r_t to pass a threshold for further processing of a particular 3D box, whereby the value can be determined based on the expected measurement density of a mobile sensor. Within the 3D box for MLS point clouds, a RANSAC algorithm is applied to find a vertical-like plane within the shrunken area to maximize the correct correspondence [Wysocki et al., 2021a; Schnabel et al., 2007], see Figure 3.3.

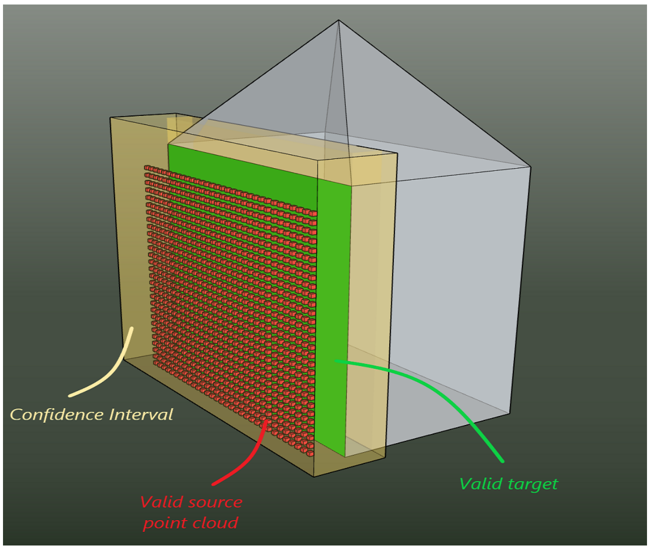


Figure 3.2 CI (yellow) discretized in L1 norm for a valid LoD 2 building model wall (green) and respective MLS point cloud (red).

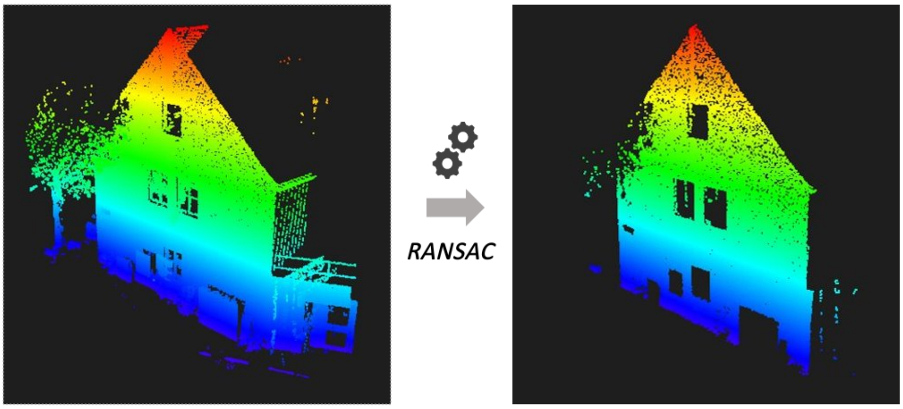


Figure 3.3 RANSAC applied within the established confidence interval; note that the confidence interval pre-filters possible outliers, limiting point cloud to wall-defined confidence area.

The previous steps yield two point clouds representing valid 3D building features (target) and respective MLS point cloud (source), as illustrated in Figure 3.2. Since the solution requires no MLS campaign trajectory, an alternative solution is proposed to find valid normal directions. The normal values are calculated for target and source point clouds in a homogeneous local coordinate system originating in the center of a scene. This step reduces large float numbers and computational time and assures consistent normal directions.

The established region of interest is now tailored for the ICP point-to-plane algorithm [Rusinkiewicz & Levoy, 2001]. It is because the *plane* is a target set, which, in this case, is represented by (sampled) planar surfaces of semantic 3D building models. On the other hand, *point* is an MLS point cloud filtered to represent planar elements of a 3D building model. Additionally, the uniform voxelization determined by voxel spacing v_s for both target and source point clouds diminishes the effect of uneven point cloud distribution. It also adapts to changing sizes of input 3D boxes. As the terrain represented by point clouds is uneven, contrary to the bottom edges of 3D models geometries, this might cause false Z-coordinate rectification. To avoid that, the height rectification is performed by estimating mode (probability mass function max. value of coordinate z of discrete random variable Z, ($z = argmax_{zi}P(Z = z_i)$) height within a 3D box of the introduced CI.

If X is a discrete random variable, the mode is the value x at which the probability mass function takes its maximum value (i.e., x=argmaxxi P(X = xi)).

Since the point clouds are assumed to be coarsely registered, the initialization matrix is represented by the identity matrix so as not to alter the initial position of the point cloud. The maximal corresponding distance corresponds to the estimated CI. The convergence criteria are met if the fitness score or relative root mean square error (RMSE) reaches $fitness_t$ and performs maximal $iterations_t$ iterations. The result of registration is a matrix applied back to the whole object in question and the associated error.

3.1.3 Matching

The matching extensively uses advances of the coregistration process. The matching adapts CIs and discretizes them again to the 3D boxes (as depicted in Figure 3.2). Since the raw ultimate misalignment is curbed, the inaccuracies are lower ,and the confidence level is higher,. The quantitative measure, RMSE, of the error is inherited from the coregistration process and is directly incorporated. New confidence borders (the 3D boxes) delineate matching bounds. The CIs serve as transmitters of the semantics and geometry between 3D building models and respective point clouds. Such 3D boxes of rich semantics, geometry, and incorporated uncertainty serve as priors for the BayNet.

3.1.4 Fusion decision based on Bayesian network

The data fusion method that considers uncertainties must quantify them throughout the process. This work incorporates the BayNet that explicitly maps uncertainties into the workflow. The designed BayNet consists of five nodes with two target ones for estimation of occupied spaces of 3D building models; The network is visualized in Figure 3.4. The nodes have two mutually-exclusive states *occupied* and *unoccupied*: The fusion couples datasets confirm one of the states, neglecting the unknown state and assuming full visibility. For two states, a joint probability distribution is calculated $P(X, Y)$, where CPT pre-scribes weights of calculating the probability of each node for each combination of its parent nodes' states. This design defines causal relationships between the nodes and results in belief estimation, as visualized using edges in Figure 3.4. To calculate a probability for a selected state, the marginalization process is used [Kjaerulff & Madsen, 2008; Stritih et al., 2020].

For such compiled networks, the obtained priors in the matching step serve as a basis for the BayNet. Such data is referred to as soft evidence as all priors contain explicit uncertainties expressed by CIs. However, the presented strategy also allows for qualitative input or hard evidence that has underlying full certainty. This trait makes the network flexible for new insights and pieces of evidence. The input uncertainty is propagated through the inference process [Stritih et al., 2020], which also estimates a posterior probability distribution (PPD) for each node in the network. This trait yields the expected state and related uncertainty for target nodes.

As shown in Figure 3.4, the first designed node in the network is *Occupied spaces for building envelope*. This node takes as input MLS point cloud and semantic 3D building data at a given global registration error with associated confidence and generalization range of 3D models, respectively. This node propagates towards the possible position and associated elements of a building. Roof and façade can have unmodeled surfaces such as dorms for roof and balconies for façade elements. Thus, further separation of building elements based on semantics leads to split for elements belonging to a wall (*Occupied spaces for wall and its elements*) or a roof (*Occupied spaces for roof and its elements*). This second tier of nodes narrows the possibilities of false associations. Both nodes lead to separate *Occupied spaces for building walls* and *Occupied spaces for building roofs* target nodes of the BayNet. The additional raw geometries of the semantic 3D model are added to the target nodes as soft evidence. This time, each piece of evidence has only a global registration error (without generalization) and a belief to check whether the explicitly modeled geometries are confirmed in reality. If there is no confirmation, then there is a high probability that an MLS observation is missing or there is no building feature present (P_{low}). On the contrary, if there is a high probability of the occupation, the elements are confirmed and probably do not require refinements and shall be fused (P_{high}). The in-between probability measure indicates discrepancies between a model and MLS measurements that should trigger further application-specific investigations ($P_{moderate}$). Therefore,

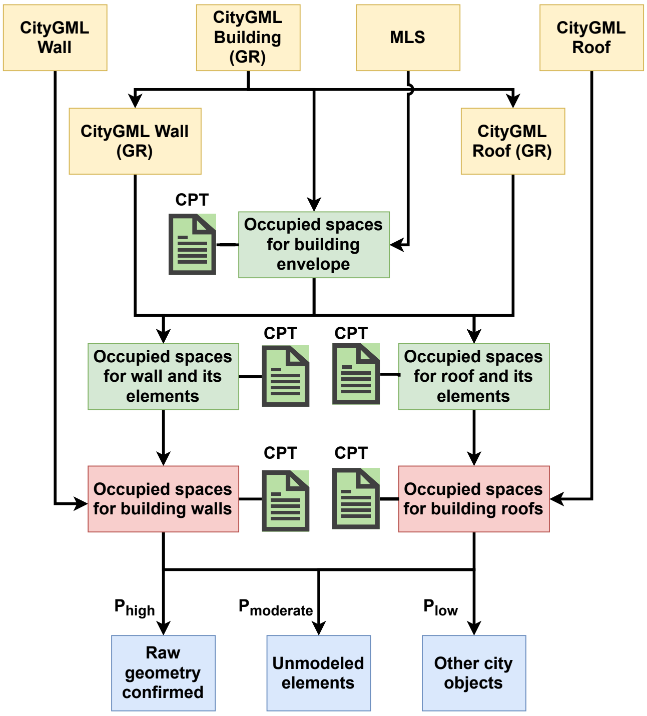

CityGML Wall
CityGML Building (GR)
MLS
CityGML Roof
CityGML Wall (GR)
CityGML Roof (GR)
CPT
Occupied spaces for building envelope
CPT
CPT
Occupied spaces for wall and its elements
Occupied spaces for roof and its elements
CPT
CPT
Occupied spaces for building walls
Occupied spaces for building roofs
P_high
P_moderate
P_low
Raw geometry confirmed
Unmodeled elements
Other city objects

Figure 3.4 The BayNet: Target nodes (red), soft evidence (yellow), and nodes (green) with CPT. GR stands for a generalized range of city model with associated uncertainty.

the distinction betweentween confirmed, unmodeled, and other city objects is based on the evaluated end probabilities. As illustrated in Figure 3.5, the final fusion shall be piece-wise, for example, voxel-to-voxel.

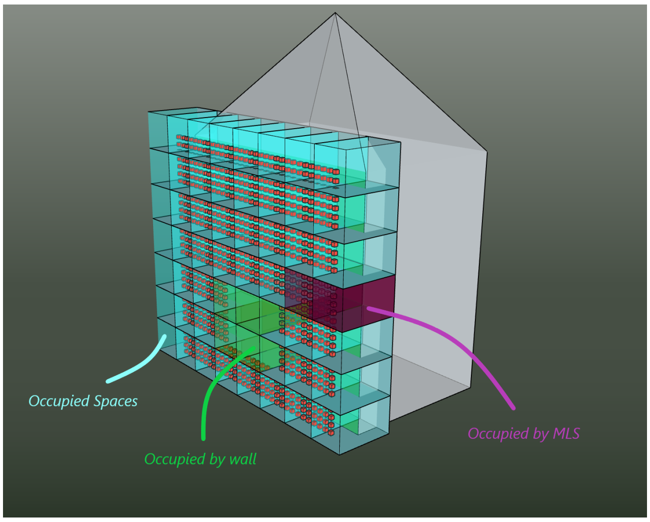


Figure 3.5 Visualization of voxelized CI space: Confirmed raw geometries with P_{high} (blue), unmodeled elements with $P_{moderate}$ (purple), and P_{low} for other city objects (green).

4 3D semantic segmentation of façade elements

As discussed in section 2.5, much research has been devoted to developing semantic segmentation algorithms. Yet, the semantic segmentation at the façade level remains underexplored. This chapter elaborates on methods developed within the scope of the doctorate pertinent to this topic, where the emphasis is put on the backbone method of uncertainty-driven ray-to-model analysis (section 4.2). The methods are primarily described based on Wysocki et al. [2022a]; Tan et al. [2023]; Wysocki et al. [2023b, 2022b]. These two sections are preceded by introducing identified façade-relevant classes (section 4.1), shown in Wysocki et al. [2022c].

4.1 Identifying façade-relevant classes

Each developed segmentation algorithm necessitates ground-truth data to validate the performance. As shown in section 2.5, there is a lack of façade-level point cloud benchmarks available as well as identified façade-relevant classes. Within the scope of this work, such a benchmark with the pertinent classes has been created. Based on the conducted research (section 2.5) and seeing the potential of already created urban-related point cloud benchmarks, a method reducing the need for creating new benchmark datasets is introduced: This reduction is achieved by enriching existing benchmarks with façade-related semantics.

Table 4.1 The proposed classes for point cloud benchmark datasets to facilitate testing of façade segmentation methods.

Index	Class	CityGML (building-related class)	Description
1	wall	WallSurface	Walls excluding any decorative elements
2	window	Window	Windows excluding any decorative elements
3	door	Door	Including garage doors
4	balcony	BuildingInstallation	Excluding pillars and other supportive structures
5	molding	BuildingInstallation	Decorative static elements adhering to a building (e.g., cornices)
6	deco	BuildingInstallation	Decorative elements mounted to a building (e.g., flags, gargoyles, lights)
7	column	BuildingInstallation	Excluding cornices (cornice - molding class)
8	arch	BuildingInstallation	Only surfaces oriented downwards
9	stairs	BuildingInstallation	Stairs excluding support structures (e.g., poles)
10	ground surface	GroundSurface	Any other ground surfaces inside a building envelope
11	terrain	Relief	Any other ground surfaces outside a building envelope (e.g., sidewalks)
12	roof	RoofSurface	Any surfaces relating to a roof structure (incl. dormers)
13	blinds	BuildingInstallation	Window closures open or closed
14	interior	InteriorWallSurface, CeilingSurface, FloorSurface	Measurements that reflect in a building
15	other	-	Any other elements

15 classes are proposed for façade segmentation, inspired by the approach of Matrone et al. [2020], which uses CityGML, Industry Foundation Classes (IFC), and Art and Architecture Thesaurus (AAT). While the number of classes is greater than that of Matrone et al., consistency and backward compatibility are maintained, i.e., it is possible to merge the classes. Moreover, the 15 classes might be aggregated into eight or fewer classes using the concept of level of façade generalization (LoFG), e.g., in the case of imbalanced data (see Figure 4.2). To facilitate both segmentation and reconstruction tasks, the classes are also consistent with the modeling guidelines of CityGML LoD3 building models [Gröger et al., 2012; Special Interest Group 3D, 2020]. The classes are shown in Table 4.1, with their names and indices, a respective building-related CityGML class, and a brief description. Extending point cloud benchmark datasets by adding new ground-truth classes inevitably necessitates manual work to be performed by

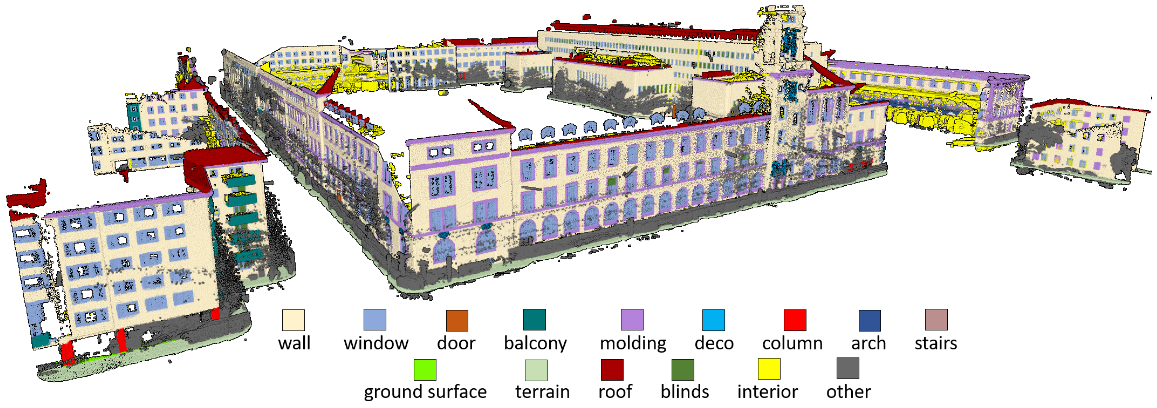


Figure 4.1 Identified 15 classes shown on an example of a point cloud depicting a set of buildings [Wysocki et al., 2024b].

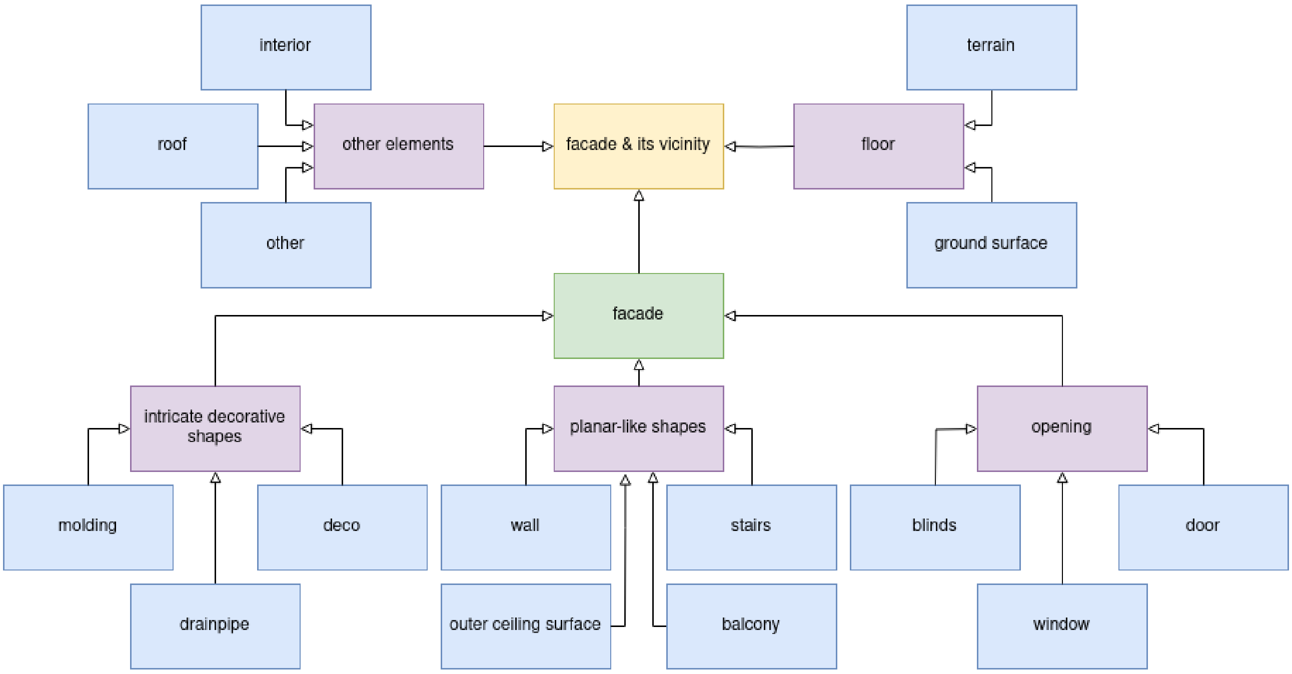
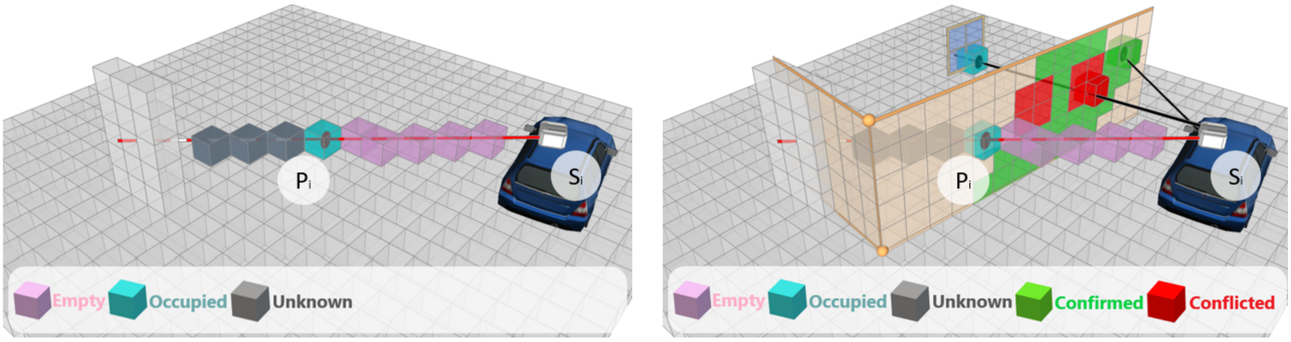


Figure 4.2 Level of Façade Generalization (LoFG): Primary 15 classes (blue) representing LoFG 4 are designed for aggregation into less detailed LoFG 3 (purple), which is more practical for the currently available imbalanced datasets; The LoFG 2 (green) allows to aggregate elements solely describing façade, whereas LoFG 1 presents a complete façade representation with its adjacent elements. Such hierarchical representation offers flexibility to the downstream tasks and addresses the frequent case of imbalanced datasets for the training of neural networks.

trained annotators. To minimize the effort involved, supporting algorithms can be used to pre-cluster point clouds [Zhu et al., 2020]. In the case of the proposed method, the central aspect is to first cluster objects that belong to the façade and its immediate vicinity and neglect all other objects. To this end, this method utilizes point clouds that are georeferenced to clip-out buildings: The position obtained from the global CRS, point clouds are superimposed on GIS datasets; This, in turn, allows creating buffers around building footprints extracted from vector GIS datasets, e.g., CityGML building models or OSM buildings. This ensures the rejection of a significant proportion of the point clouds and clusters the building-related points per building object while addressing global point positioning inaccuracies [Wysocki et al., 2021b]. Alternatively, when point clouds are not georeferenced, or GIS datasets are unavailable, existing benchmark points, annotated as buildings, can be used as a pre-cluster for façade-related points. If the aforementioned cases are not satisfied, the façade must be extracted manually, or else clustering algorithms must be used, similar to [Zhu et al., 2020].



(a) Ray casting of laser observations

(b) Rays analyzed with 3D model

Figure 4.3 Visibility analysis using laser scanning observations and 3D models on a voxel grid. The ray is traced from the sensor position s_i to the hit point p_i. a) The voxel is: *empty* if the ray traverses it; *occupied* when it contains p_i; *unknown* if unmeasured; b) the proceeding states are derived when analyzed with model: *Confirmed* when *occupied* voxel intersects with vector plane; and *conflicted* when the plane intersects with an *empty* voxel [Wysocki et al., 2022a]. The states are derived from accumulated observations in a probabilistic manner.

4.2 Conflict analysis for semantic façade segmentation

In this section, an analysis pertinent to the 3D façade-level segmentation is presented, namely the ray-to-building conflict analysis. By analyzing the laser ray trajectory, the relation between an existing low LoD building model and point clouds is established. The method is complemented by the 3D reconstruction presented in subsection 5.2.1.

4.2.1 Uncertainty assignment

Confidence intervals (CI) are introduced to quantify multimodal uncertainties. The CI is estimated based on the confidence level (CL), with its associated z value (z), standard deviation (σ), and mean (μ).

The CI for façades is estimated using $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Assuming Gaussian distribution and operating in the L1 norm, the maximum upper and lower bounds are given by $[\mu_i - 2\sigma_i, \mu_i + 2\sigma_i]$ [Suveg & Vosselman, 2000]. σ_1 describes the location uncertainty of MLS point clouds, while σ_2 addresses the uncertainty of semantic 3D building walls. The assumptions are made for the point cloud global registration error e_1 and for the global location error of building model walls e_2. The operator's belief regarding the deviation from the true value is quantified using the confidence levels CL_1 and CL_2 for point clouds and building model walls, respectively. Both confidence levels CL_i are bound to the respective z_i value, while μ_i divided by z_i is an estimate of the standard deviation σ_i value [Hazra, 2017].

4.2.2 Visibility analysis concerning uncertainties

Automatic visibility analysis is conducted to identify missing model elements. As shown in Figure 4.3, an occupancy grid is introduced to analyze the multimodal data of the semantic 3D building models and the MLS observations. The grid is an octree structure encompassing a volume of interest, in which 3D voxels are the octree's leaves. The 3D voxels are used to search for conflicts between the semantic 3D building models and the MLS measurements, where the voxels' size v_s is chosen according to the expected uncertainties of the 3D building model and the MLS point cloud.

A measurement by the laser scanner is emitted from the position of the sensor s_i, oriented by the vector r_i, and pointing toward the reflective position $p_i = s_i + r_i$ (Figure 4.3a). Voxels that cover p_i are labeled as *occupied* (blue), those between s_i and r_i as *empty* (pink), and those behind p_i and traversed by the elongated ray as *unknown* (gray). Repeated voxels' observations are considered in a probabilistic fashion, as shown by Moravec & Elfes [1985]. Probabilities are assigned using log-odds notation and clamping policy [Hornung et al., 2013; Tuttas et al., 2015].

$$L(n|z_{1:i}) = max(min(L(n|z_{1:i-1}) + L(n|z_i), l_{max}), l_{min}) \tag{4.1}$$

where

$$L(n) = log[\frac{P(n)}{1 - P(n)}] \tag{4.2}$$

$P(n)$ denotes prior probability, whereas the values of l_{min} and l_{max} are used to define the clamping thresholds of the log odd-values $L(n)$, also called log-odds bounds. Faces are inserted into the occupancy grid (Figure 4.3b) to enable a comparison between the point cloud and vector model. Each inserted face has an uncertainty given by the estimated upper bound of the CI and the associated CL. The upper CI defines a range of façade position deviations, while the CL indicates its associated belief. Finally, each voxel is represented by the position, size, and probabilities relating to the model and measurements.

The façades are subjected to a piece-wise comparison by analyzing the state of the voxels (Figure 4.3b): Voxels determined by laser observations as having the state *occupied* and that are occupied by the intersection of façades are labeled as *confirmed* (green); while voxels labeled as *empty* by the laser observations and which intersect with façades are labeled as *conflicted* (red).

A texture map is defined for each façade-surface of the building model. It has a cell spacing c_s, following the projection of the voxel grid to the plane. Each pixel in the texture map is labeled in relation to the voxel's state as *confirmed* or *conflicted*; the areas of the façade-surface uncovered by the MLS observations are labeled as *unknown* (grey). Such a composed map is referred to as a conflict map. The next processing step depends on the ratio $r_c \in [0, 1]$ of the number of pixels with the state *conflicted* a_1 to the total surface area of façade-surface a_2, including *unknown* and *confirmed* parts.

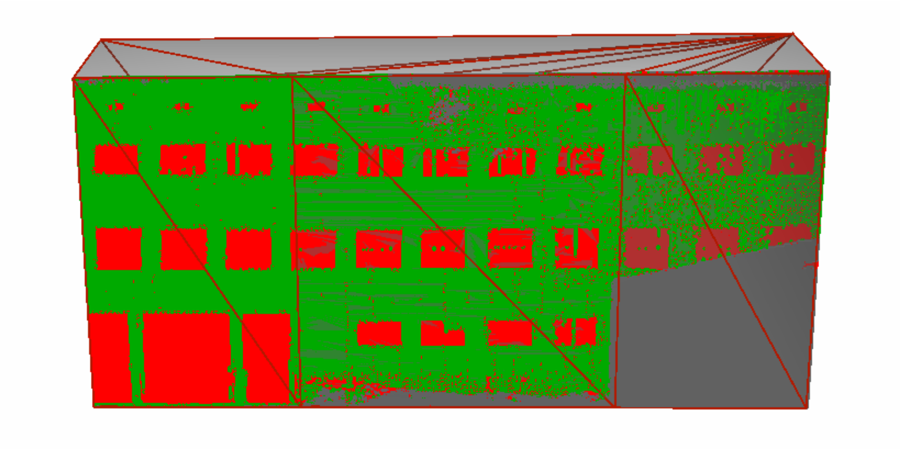


Figure 4.4 Example of a texture map showing the states *confirmed*, *conflicted*, and *unknown*.

$$r_c = \frac{a_1}{a_2} \tag{4.3}$$

If r_c is smaller than $r_{c_{min}}$, the modeled façade-surface is deemed to be correct, which means that any refinement is superfluous. If r_c is greater than $r_{c_{max}}$, the modeled façade is significantly erroneous and a refinement is unviable. The modeled façade is therefore only considered for the refinement when its ratio r_c is: $r_c \in [r_{c_{min}}, r_{c_{max}}]$. The *conflicted* pixels mark the probability-quantified location of façade openings, such as windows, doors, and underpasses, as shown in Figure 4.4.

4.3 Probabilistic conflict analysis supported by model knowledge

The previous subsection 4.2.2 shows how to derive conflicts at the surface of the building models. However, the semantics there is absent, as there is a limited distinction between the type of a façade element.

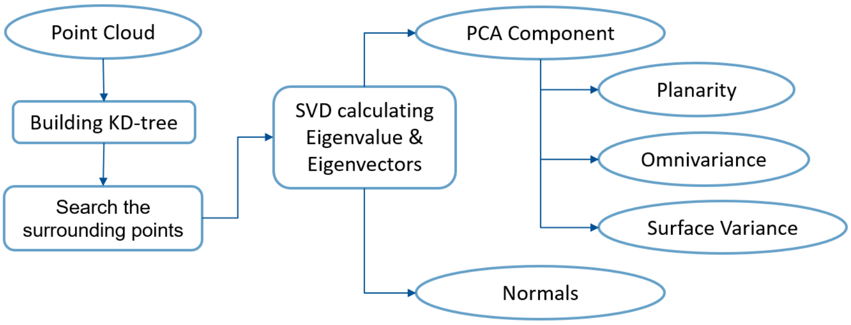


Figure 4.5 Workflow of geometric features extraction on an example of a few geometric features.

This section introduces a method based on model knowledge, which enables deriving semantics of the identified conflicts; this method is complemented by a façade element library, which extends the concept of the final 3D modeling (described in section 4.3).

4.3.1 3D façade semantic segmentation at point cloud level

There have been many studies demonstrating the effectiveness of geometric features in point cloud classification, which can be viewed as a detailed semantic interpretation of local 3D structures (section 2.5). According to the study of Grilli & Remondino [2020], $planarity$, $omnivariance$, and $surface\ variation$ are efficient in the classification of building point cloud data applying Random Forest. Here, the following geometric features are utilized: *height of the points, roughness, volume density, verticality, omnivariance, planarity*, and *surface variation*. The last three mentioned features are based on the normalized eigenvalues λ_i $(\lambda_1 > \lambda_2 > \lambda_3)$, which are derived from the 3D point coordinates within a considered spherical neighborhood r_i [Weinmann et al., 2013; Grilli & Remondino, 2020]. To find such surrounding points for each point, a k-d tree is used for nearest neighbor queries. Then, singular value decomposition (SVD) calculates the respective structure tensor, which directly provides information on the surrounding points' distribution and structure. In this way, the eigenvalues λ_1, λ_2, λ_3, as well as the corresponding eigenvectors e_1, e_2, e_3 are derived from the covariance matrix. In addition, the eigenvalues with $\lambda_1 > \lambda_2 > \lambda_3$, known as the components of principal component analysis (PCA), are used for further measures of other covariance features. Besides, the second vector of eigenvectors, which can describe the general direction perpendicular to the curve defined by surrounding points, is also included. Figure 4.5 illustrates the pipeline of the geometric features calculation. As shown in Figure 4.6, the proposed method leverages not only the potential of geometric features but combines it with the effectiveness of DL networks. The early fusion of the geometric features into the point-based models allows for enhanced input training knowledge beyond spatial X, Y, and Z encoding.

To add geometric features into networks, the internal structure of the neural networks requires altering. For point-based networks, such as PointNet, features learned in this network are calculated by convolution on the coordinate of each point and then gathered globally into a general single layer. Hence, the input expands the dimension of the coordinate into $3 + N_f$, where N_f is the number of pre-computed geometric features. As for hierarchically structured networks with set abstraction layers, such as PointNet++, features from different scales of metric spaces are collected, consisting of the sample layer, grouping layer, and the PointNet layer. The former two layers generate regions on different scales using point cloud coordinates, while the PointNet layer is used to gather features from a group of points in different regions created in the former two layers. For this reason, the input dimension is increased for the PointNet layer to $N \times K \times (d + N_f)$, where N is the number of points, K is the number of regions after grouping, d is the dimension of coordinates, N_f is the dimension of external geometric features. A similar approach is replicated for the Point Transformer self-attention network. An example of such an early-fusion approach is illustrated in Figure 4.7.

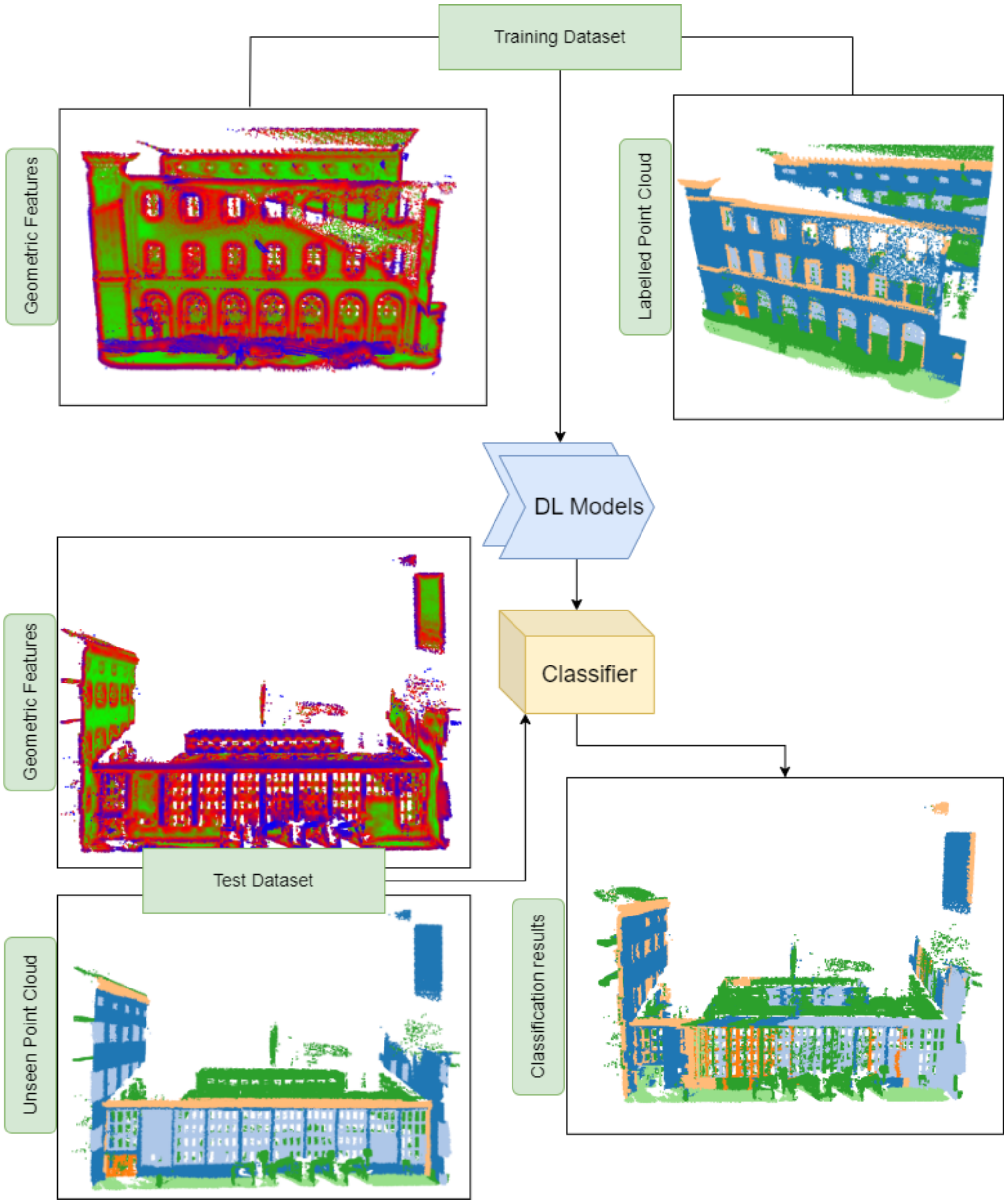


Figure 4.6 Point cloud classification with geometric features process for unseen datasets.

The goal of semantic segmentation is to divide a point cloud into several subsets based on the semantics of the points to facilitate 3D semantic reconstruction. As shown in Figure 4.8, and considering the introduced LoFG 3 (see Figure 4.2), eight relevant classes for façade segmentation tasks are considered: *arch* (dark blue), *column* (red), *molding* (purple), *floor* (green), *door* (brown), *window* (blue), *wall* (beige), and *other* (gray).

The final softmax output layer enables obtaining an output vector of probabilities for each predicted class, which becomes fundamental for running the conflict classification approach. The points are projected onto the 3D model façade-surface, forming a texture map layer with labels corresponding to the classes, as shown in the example of windows (orange) in Figure 4.9. The texture cell spacing follows the size of a conflict map to ensure compatibility.

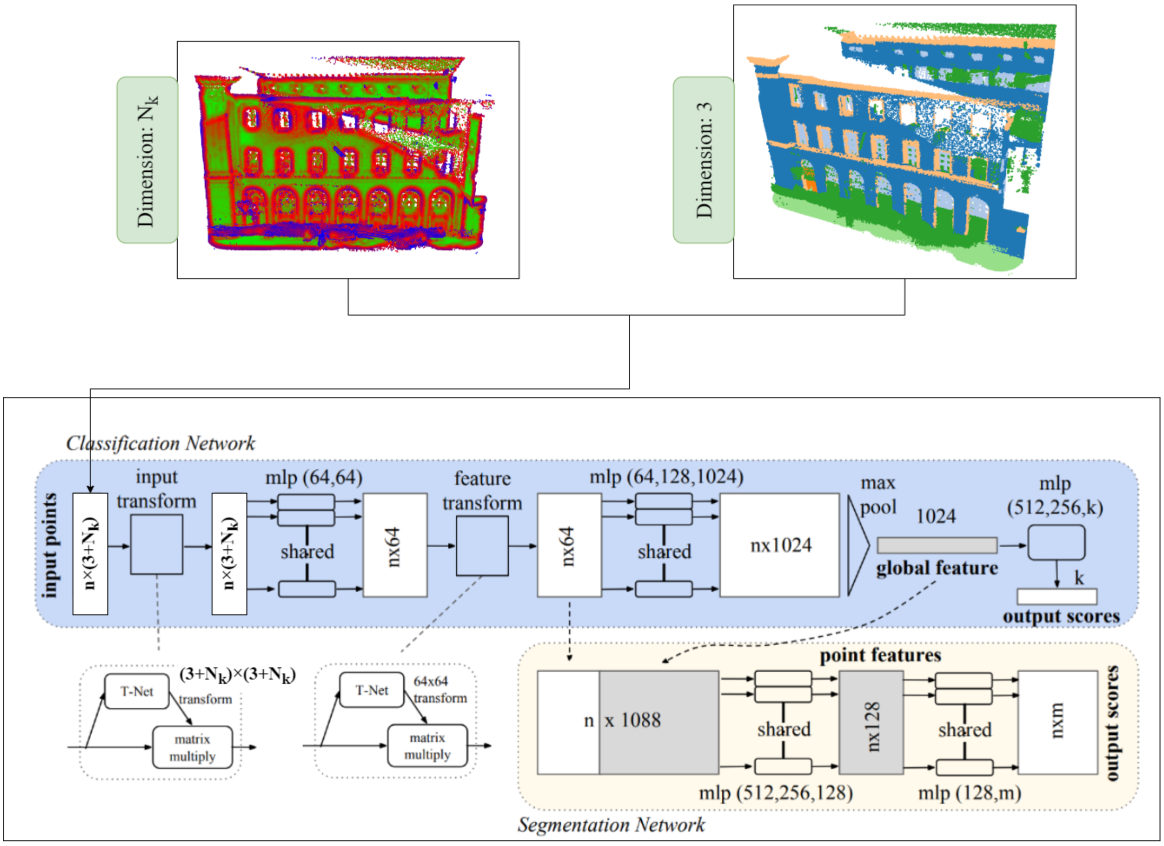


Figure 4.7 Proposed early-fusion on an example of the point-based PointNet network. The lower part of Figure stems from: [Qi et al., 2017a].

4.3.2 Probabilistic classification: the Bayesian approach

Visibility analysis yields conflict map textures on façades, whereas point cloud segmentation provides semantics. Here, to identify façade openings, a BayNet is used. As shown in Figure 4.10, the designed BayNet comprises one target (red), two input (yellow), one decision (blue), and two output nodes (green). Each directed link represents a causal relationship between the X and the Y nodes. The CPT prescribes weights for each state and node combination (gray). The target, *opening* state is calculated using the joint probability distribution $P(X, Y)$ and the CPT. The marginalization process is used to calculate the probability of the target node Y being in the *opening* state y. The process sums conditional probabilities of the states x stemming from parent nodes X [Stritih et al., 2020]. Since the network consists of texture layers with state probabilities, the data evidence represents the so-called soft evidence [Stritih et al., 2020]. In an inference process, soft evidence is added to update the joint probability distribution. This process provides the most likely node states by estimating the posterior probability distribution (PPD).

Pixel classes from the *conflict map* and *semantic map* textures form clusters if they have a neighbor in any of the eight directions of the pixel. The co-occurring *conflicted*, *window*, and *door* cluster classes lead to a high probability of unmodeled openings. This output is used for further opening 3D modeling and is back-projected onto segmented point clouds as either the *window* or *door* class. On the other hand, co-occurring *confirmed*, *window*, and *door* clusters lead to a low probability of existing openings. The low probability P_{low} and the high probability P_{high} labels are assigned to clusters based on the probability threshold P_t: $P_{high} > P_t >= P_{low}$.

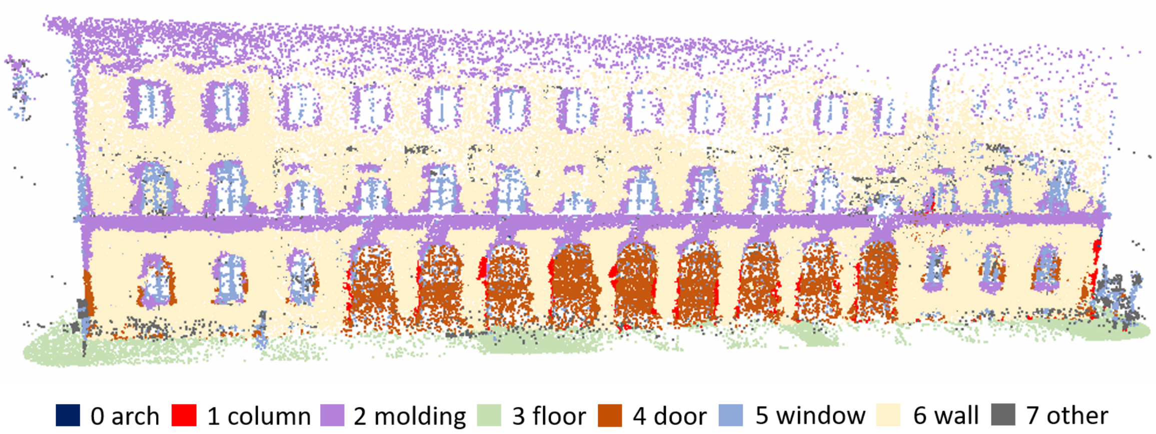


Figure 4.8 Exemplary 3D façade semantic segmentation result.

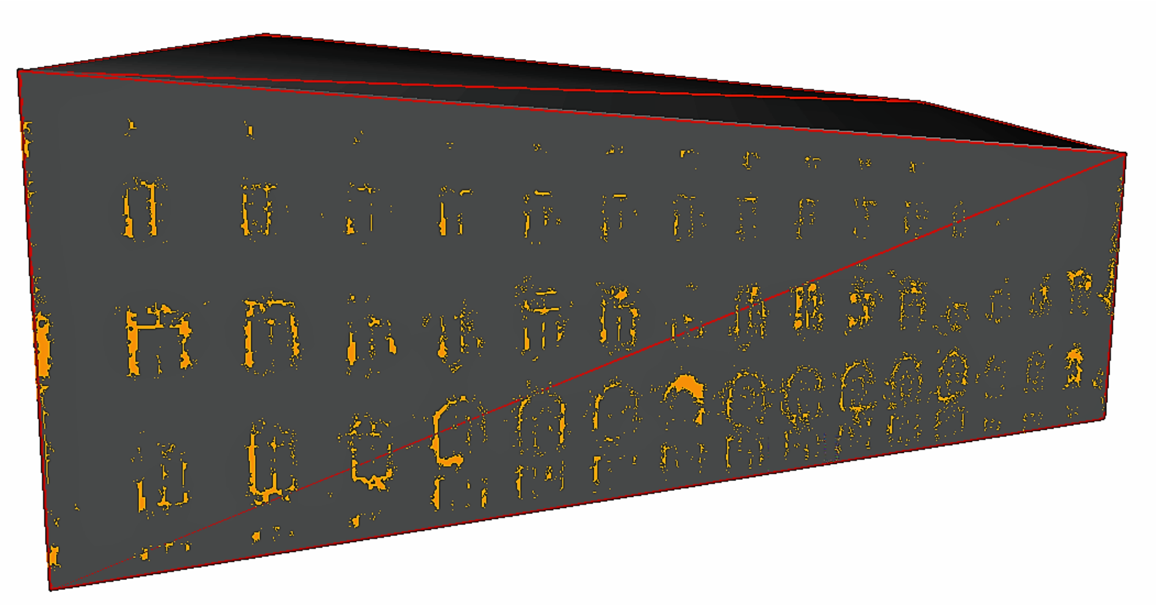


Figure 4.9 Texture layer showing the class *window* projected onto a façade of 3D model.

4.4 Probabilistic conflict analysis supported by model knowledge and images

As shown in the previous sections, the first task is to generate a ray-based conflicts probability map consisting of three states (*conflicted*, *confirmed*, and *unknown*), analyzing the visibility of the laser scanner in conjunction with 3D building models. However, this map is limited to the laser field-of-view and does not provide façade-specific semantics. To address this limitation, additional two probability maps are introduced derived from point clouds and images: The former is generated by a modified Point Transformer network [Zhao et al., 2021; Wysocki et al., 2022a] (top branch), while the latter is produced using Mask-RCNN [He et al., 2017] (bottom branch), as illustrated in Figure 4.11. These are then fused as three probability maps via a Bayesian network, resulting in a target probability map that represents the occurrence of openings and their associated probability score. The method is complemented by the semantic reconstruction presented in subsection 5.2.1.

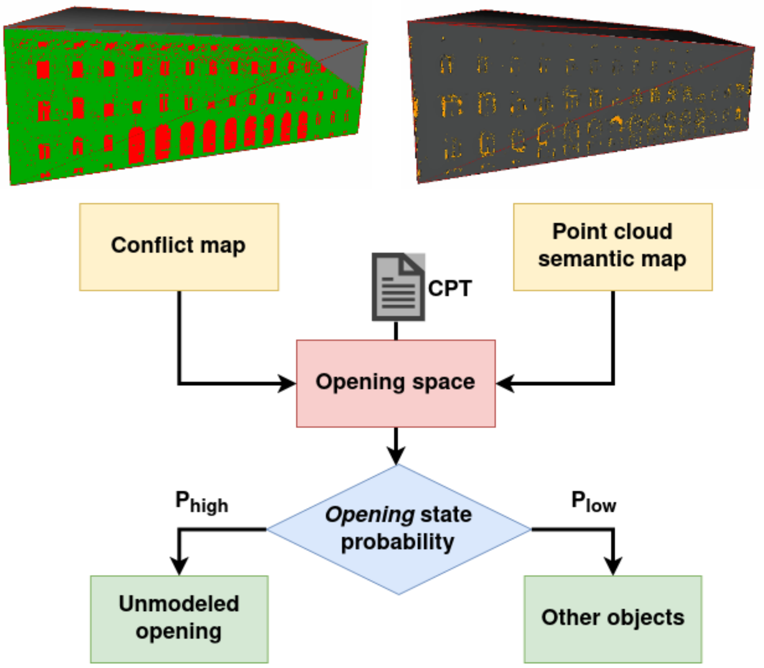


Figure 4.10 Input nodes (yellow) and CPT estimate the probability of opening space (red) in BayNet: if (blue) the probability is high, openings are unmodeled; otherwise, areas indicate other objects (green).

4.4.1 Visibility analysis concerning uncertainties

The ray tracing is performed on a 3D voxel grid to determine areas that are measured by a laser scanner and analyze them with a 3D building model (Figure 4.3). The total grid size adapts to the input data owing to the utilized octree structure with leaves represented by 3D voxels of size v_s dependent on the relative accuracy of the scanner.

As shown in Figure 4.3a, the laser rays are traced from sensor position s_i, using orientation vector r_i, to hit point $p_i = s_i + r_i$. The approach leverages the MLS trait of multiple laser observations z_i to decide upon the laser occupancy states (i.e., *empty*, *occupied*, and *unknown*) and includes the respective occupancy probability score. The states' update mechanism uses prior probability $P(n)$, current estimate $L(n|z_i)$, and preceding estimate $L(n|z_{1:i-1})$ to calculate and assign the final state. The mechanism is controlled by log-odd values $L(n)$ along with clamping thresholds l_{min} and l_{max} [Hornung et al., 2013; Wysocki et al., 2022b,a]:

$$L(n|z_{1:i}) = max(min(L(n|z_{1:i-1}) + L(n|z_i), l_{max}), l_{min}) \tag{4.4}$$

where

$$L(n) = log[\frac{P(n)}{1 - P(n)}] \tag{4.5}$$

As illustrated in Figure 4.3a, in the visibility analysis process of laser observations, voxels encompassing p_i are deemed as *occupied* (light-blue), those traversed by a ray as *empty* (pink), and unmeasured as *unknown* (gray). Then, as shown in Figure 4.3b, to assign further voxel states, occupancy voxels and building models are analyzed: Voxels are *confirmed* (green) when *occupied* voxels intersect with the building surface and are *conflicted* (red) when a ray traverses a building surface and reflects inside a building. The final probability estimate, however, also concerns 3D model uncertainties.

Specifically, the method addresses uncertainties of global positioning accuracy of building model surfaces and point clouds along the ray. In contrast to subsection 4.2.1, here the uncertainty is considered in the L2 norm, without the discretization into the L1 norm (Figure 4.12). Let us assume that the probability distribution of global positioning accuracy of a building surface $P(A)$ is described by the Gaussian distribution $\mathcal{N}(\mu_1, \sigma_1)$, where μ_1 and σ_1 are the mean and standard deviation of the Gaussian distribution. Analogically, let us assume that the probability distribution of global positioning accuracy of a point in point

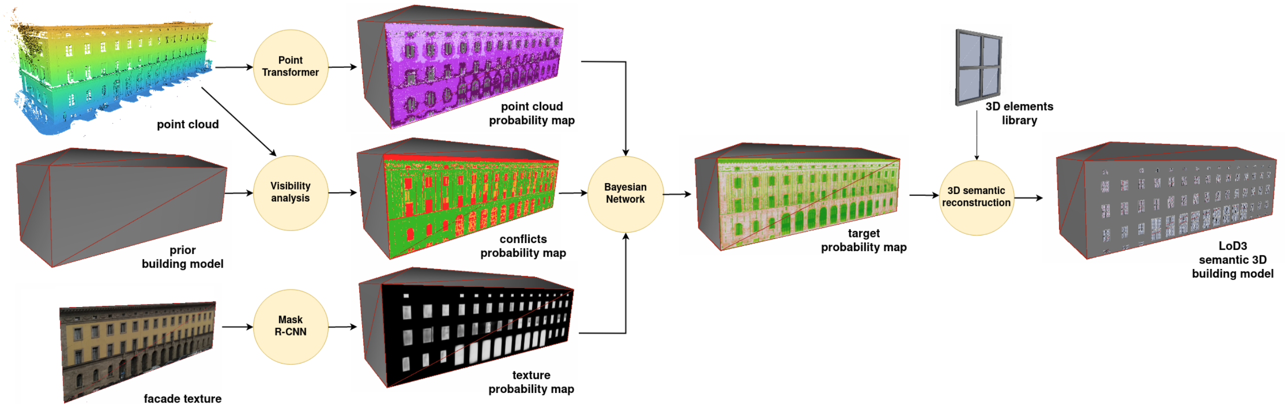


Figure 4.11 The workflow of the proposed method consists of three parallel branches: The first is generating the point cloud probability map based on a modified Point Transformer network (top); the second is producing a conflicts probability map from the visibility of the laser scanner in conjunction with a 3D building model (middle); and the third is using Mask-RCNN to obtain a texture probability map from 2D images. Then, three probability maps are fused with a Bayesian network to obtain final façade-level segmentation, enabling a CityGML-compliant LoD3 building model reconstruction (described later in chapter 5).

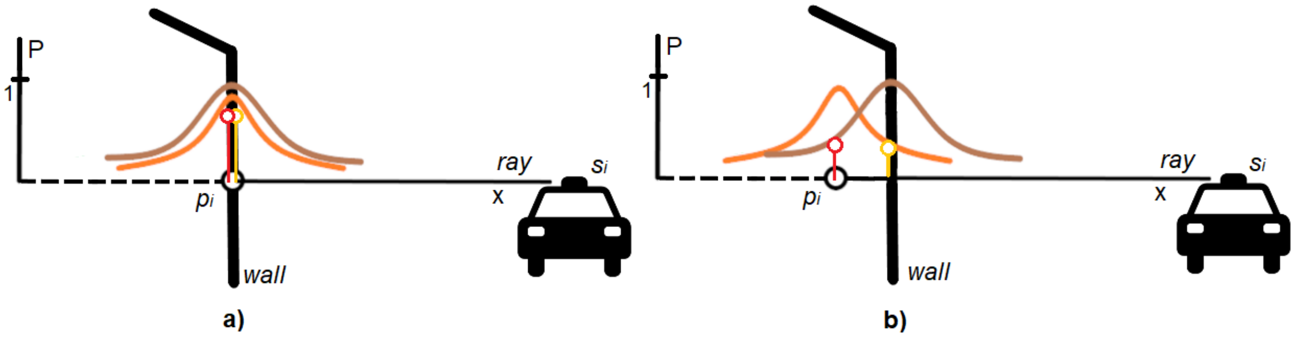


Figure 4.12 Points ray tracing from the sensor position s_i to the hit point p_i. Probability distributions of 3D building wall accuracy P_A (brown) and of point accuracy P_B (orange) considering sensor position s_i and hit point p_i along the ray, with red pole $P_A(x = x_B)$ and yellow pole $P_B(x = x_A)$. a) Situation A: very high probability of surface confirmation, b) Situation B: low probability of surface confirmation, but high probability of missing object.

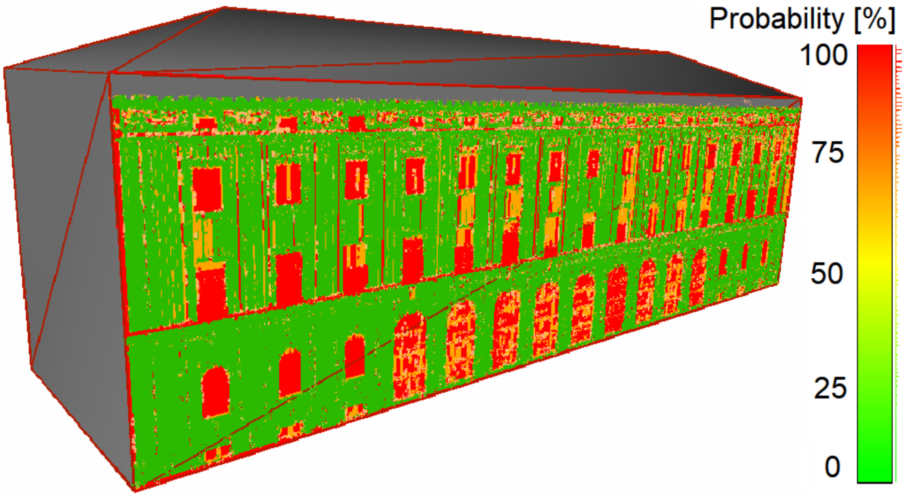


Figure 4.13 Exemplary *conflict probability map*: high probability pixels present high conflict probability, whereas low probability pixels show high confirmation probability.

cloud $P(B)$ is described by the Gaussian distribution $\mathcal{N}(\mu_2, \sigma_2)$. To estimate the probability of the confirmed $P_{confirmed}$ and conflicted $P_{conflicted}$ states of the voxel V_n, the joint probability distribution of two independent events $P(A)$ and $P(B)$ is used:

$$V_n = \left\{ \begin{matrix} P_{confirmed}(A, B) = P(A) * P(B) \\ P_{conflicted}(A, B) = 1 - P_{confirmed}(A, B) \end{matrix} \right\} \tag{4.6}$$

A *conflicts probability map* (Figure 4.13) is derived by projecting the vector-intersecting voxels to the vector plane, where the cell spacing is consistent with the voxel grid; each pixel receives probability values of the states *conflicted*, *confirmed*, and *unknown*, accordingly.

4.4.2 3D semantic segmentation on point clouds

The semantic segmentation of 3D point clouds is performed analogically to the previous subsection 4.3.1. Namely, the enhanced Point Transformer (PT) network [Wysocki et al., 2022a; Zhao et al., 2021] is applied. The enhancement involves fusing geometric features at the early training stage to increase 3D

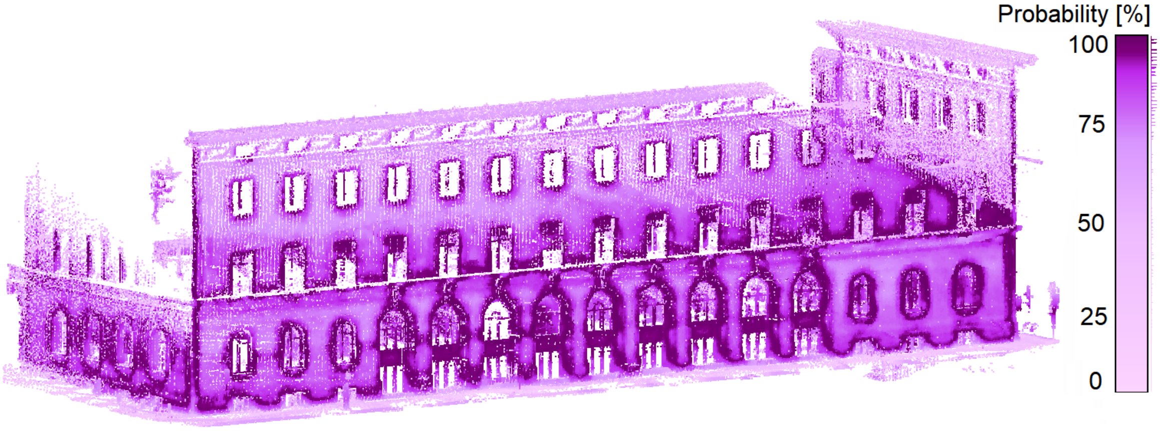


Figure 4.14 Exemplary results of the modified network: point cloud colors according to the probability vector of the class *window*.

façade segmentation performance [Matrone et al., 2020; Wysocki et al., 2022a]. In this work, seven geometric features are considered: *height of the points, roughness, volume density, verticality, omnivariance,* *planarity*, and *surface variation* [Weinmann et al., 2013; Grilli & Remondino, 2020; Wysocki et al., 2022a], which are calculated within an Euclidean neighborhood search radius r_i. Following the introduced LoFG 3 (see Figure 4.2), eight pertinent classes for the façade segmentation task are defined: *arch*, *column*, *molding*, *floor*, *door*, *window*, *wall*, and *other* [Wysocki et al., 2022a].

The final softmax layer of the modified PT network provides a per-point vector of probabilities of each class as an output (Figure 4.14). Notably, in contrast to subsection 4.3.1, this method do not discards points based on a probability threshold but consider each point and its class probability score for further processing, minimizing information loss. Finally, the *point cloud probability map* (Figure 3.4) is created by projecting the points onto the face of a building while preserving the probabilities and following the cell spacing of the *conflict probability map* (subsection 4.4.1).

4.4.3 2D semantic segmentation on images

As demonstrated by Hensel et al. [2019], Faster R-CNN [Ren et al., 2015] effectively identifies approximate façade openings positions. In the presented approach, Mask-RCNN [He et al., 2017] is applied, which builds upon the concept of Faster R-CNN and identifies probability masks within proposed bounding boxes. This trait allows obtaining more accurate instances that are not necessarily restricted to a rectangular shape. Analogically to the 3D semantic segmentation stage (subsection 4.4.2), the pixel-predicted probabilities are preserved. To generate the texture probability map (Figure 4.15), the pixels and their probabilities are projected onto the building face, aligning with the cell spacing of the previously introduced probability maps.

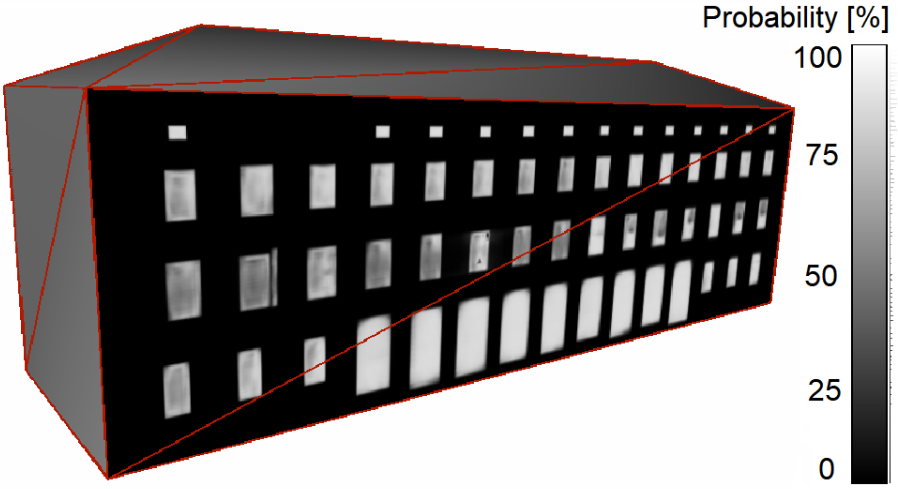


Figure 4.15 Exemplary *texture probability map*: high probability pixels stand for a high probability of opening.

4.4.4 Final segmentation by late-fusion with Bayesian network

To calculate the final shape, semantics, and probability score of opening instances, the multimodal probability maps are fused using a Bayesian network. The network quantifies uncertainties and assigns weights based on evidence when calculating the target probability map. Figure 4.16 shows the network architecture, including three input nodes for each probability map, to infer the probability of opening occurrence. The X and Y nodes exhibit a causal relationship, forming directed acyclic links. A CPT is employed to assign weights to combinations of each node and state. The target node estimates two mutually exclusive states: opening and non-opening. The probability of node Y (opening space) being in the state y (opening) is calculated using the marginalization process, which combines the conditional probabilities of the parent nodes' X states x (i.e., of point cloud probability, conflicts probability, texture probability maps) [Stritih et al., 2020; Wysocki et al., 2022b].

The probability maps serve as pieces of evidence updating the joint probability distribution $P(X, Y)$ of the compiled network. The inference mechanism performs the update and estimates the posterior probability distribution (PPD), which provides the states' probability [Stritih et al., 2020; Wysocki et al., 2022b]. In general, the network favors situations where there is a high probability of an opening occurring if at least two pieces of high-probability evidence co-occur; otherwise, it yields a low opening probability. For example, a very high *conflict* probability overlying high texture *opening* probability and medium point cloud *opening* probability should yield a high *opening* probability.

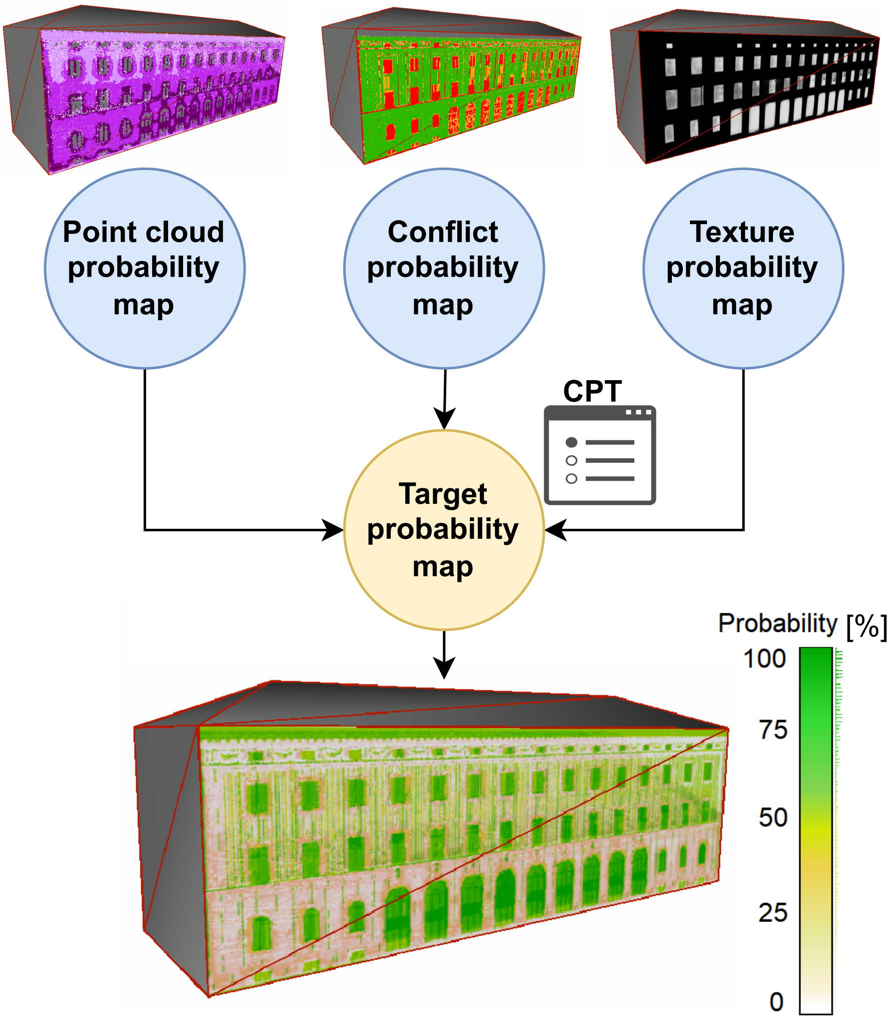


Figure 4.16 The Bayesian network architecture comprises three input nodes (blue), one target node (yellow), and a conditional probability table (CPT) with the assigned combinations' weights.

The Bayesian network outputs are the high probability clusters P_{high}, which have a neighbor in any of the eight directions of the pixel. To distinguish between different classes, the superior overlying per-pixel class probabilities are selected. The pixel-wise probability scores are then averaged per instance and kept for the final 3D model.

5 Enrichment of 3D models using façade elements

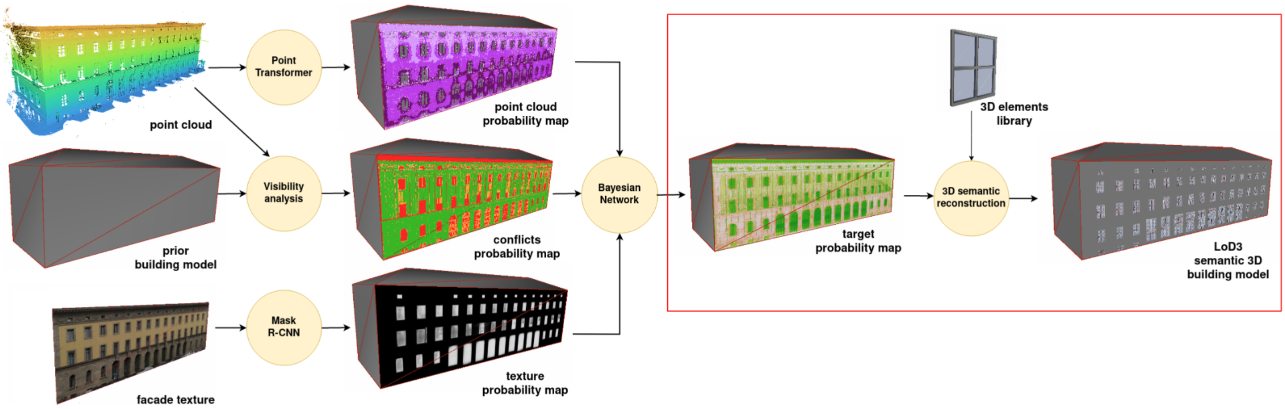


Figure 5.1 The semantic 3D reconstruction (red box) relies on the accuracy of the preceding segmentation methods. The adapted Figure 4.11, chapter 4.

This chapter is devoted to 3D semantic reconstruction methods within the refinement strategy. The methods exploit geometric priors and uncertainty while reconstructing the models at LoD3, maintaining and upgrading its semantics where required. As illustrated in Figure 5.1, reconstruction performance relies on the accuracy of preceding segmentation methods. Therefore, the reconstruction methods presented here employ the segmentation methods presented in chapter 4.

The probabilistic conflict analysis method is introduced in section 5.2, mainly based on the publication of Wysocki et al. [2022b]; the method extension by library and model knowledge is shown in section 4.3 and derives from Wysocki et al. [2022a]; Froech et al. [2023b]; whereas the image-driven extension is presented in section 4.4 drawn from Wysocki et al. [2023b].

5.1 Façade elements relevant to standardized semantic 3D building models

The refinement process involves altering the shape of the prior semantic model. This feature not only removes parts of unwanted geometry but also imposes changes on the semantic structure of a 3D model. This section presents guidelines for conducting such refinement.

As presented in Table 5.1, GroundSurface, RoofSurface, and WallSurface positions shall not be altered, as it will corrupt the overall geometry of the prior model since the street-level measurement typically covers only one or two sides of the building. Such change shall be only undertaken if the point cloud significantly covers the building's outer shell (e.g., see [Nan & Wonka, 2017]) and is of higher global accuracy than usually cadastre-derived WallSurface [Roschlaub & Batscheider, 2016].

Yet, even when the geometry is altered, the entity's identifiers shall not be changed, as external and internal links associated with city model objects exist. For the latter, as highlighted in Figure 5.2, the input hierarchical model of a LoD2 comprises parent-child relation (orange, id and parent_id). If the Building id (i) is changed, the WallSurface parent_id (i) is invalid, orphaning the WallSurface object. A WallSurface id (z) shall not cause any internal hierarchical disruption, but it might disrupt external links. For example, an identifier of a WallSurface can be linked to a report about its solar potential; in case the identifier is changed, this information will be lost.

Table 5.1 Analyzed point cloud and CityGML 2.0 classes relevant for the LoD3 reconstruction with new, proposed functions (green) for absent ones [Special Interest Group 3D, 2020].

Point cloud class	CityGML class	LoD	Function	Refinable	Confidence score
ground surface	GroundSurface	2, 3, 4	-	~	~
roof surface	RoofSurface	2, 3, 4	-	~	~
wall	WallSurface	2, 3, 4	-	~	✓
window	Window	3, 4	-	✓	✓
door	Door	3, 4	-	✓	✓
underpass	BuildingInstallation	3, 4	1002 underpass	✓	✓
balcony	BuildingInstallation	3, 4	1000 balcony	✓	✓
molding	BuildingInstallation	3, 4	1016 molding	✓	✓
deco	BuildingInstallation	3, 4	1017 deco	✓	✓
column	BuildingInstallation	3, 4	1011 column	✓	✓
arch	BuildingInstallation	3, 4	1008 arch	✓	✓
drainpipe	BuildingInstallation	3, 4	1018 drainpipe	✓	✓
stairs	BuildingInstallation	3, 4	1060 stairs	✓	✓
blinds	BuildingInstallation	3, 4	1019 blinds	✓	✓

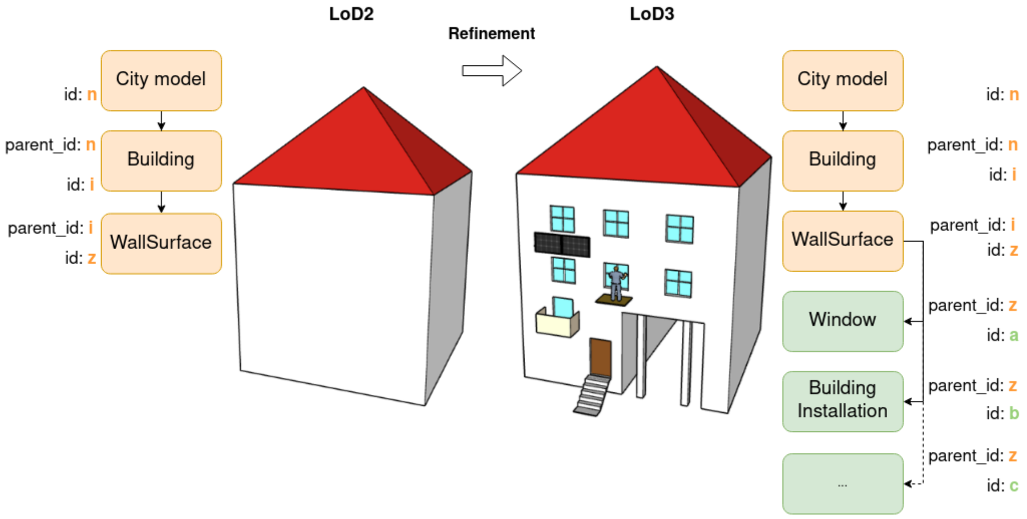


Figure 5.2 The refinement procedure on semantic 3D city model at LoD3.

As shown in green in Figure 5.2, each new object should be attached not only geometrically to a WallSurface but also by the corresponding parent_id (z). As such, each façade-element is assigned to one WallSurface, maintaining the model's consistency; Each new object shall possess a new unique object id (green a, b, c).

Regardless of the total building coverage, identified openings, such as underpasses, windows, and doors, shall cut out the WallSurface geometry where required, and 3D geometries should be fitted into this empty space. Although it is allowed to model an underpass at LoD2, in practice, it is rarely the case, primarily owing to the aerial perspective of data acquisition for LoD2 models [Dukai et al., 2020]. The underpasses are deemed as façade openings that substantially impact façade semantics and geometry; as such, they are classified as LoD3 features.

The notion of detection and reconstruction confidence is required since many applications nowadays rely not only on geometric accuracy but also on the associated confidence of the object's semantics. For example, map-based navigation of cars uses a voting process fusing multiple sensor detections to decide upon the next manoeuvre [Wilbers et al., 2019]. Therefore, retaining the uncertainty information and adding it as GenericAttribute called $confidence \in [0, 1]$ to the specific object is of great value. The enrichment of the given building model is also pursued by attaching the confidence probability map (subsection 4.2.2) as a texture to the respective WallSurface.

5.2 Semantic 3D reconstruction with geometric prior

As described in section 4.2, the façade segmentation method provides a geo-referenced conflict texture on the face of the model. Such conflicts provide cues for reconstruction, which should subsequently remove all detected conflicts. Simultaneously, the 3D reconstruction enriches the semantics of a building entity and its geometric accuracy by adding new 3D geometries into the 3D entity.

5.2.1 Extracting target shapes

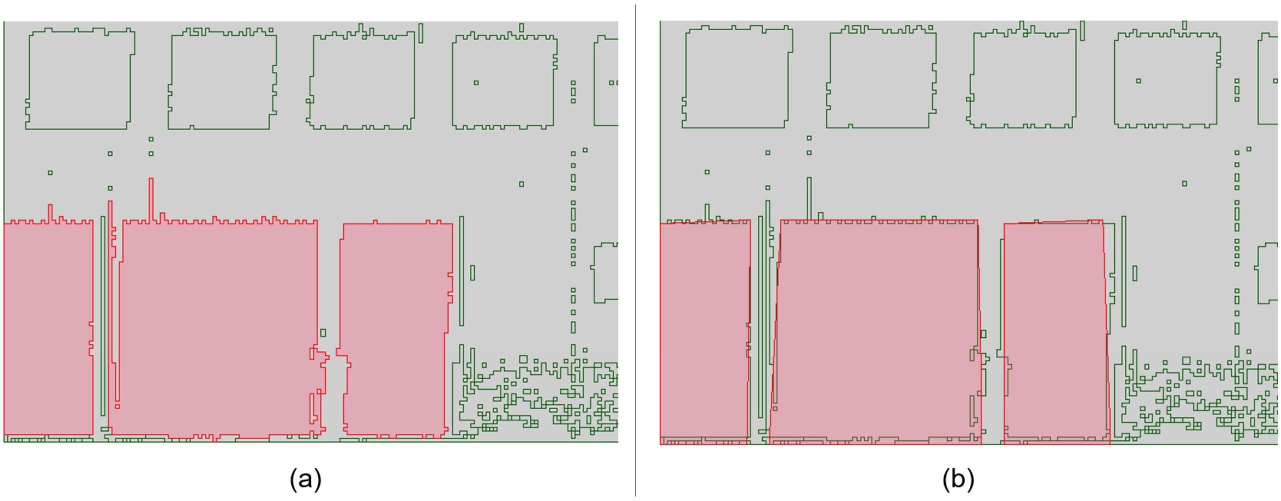


Figure 5.3 Extracting target shapes on textures: a) Identified blobs of interest, b) generalization of extracted shapes.

The generalized shape is used to create a 3D representation of the space to be deducted from the raw model. A façade element in a building is modeled using the CSG difference operation, in which the reconstructed space is subtracted from a raw 3D building geometry, as shown in Section 5.2.4.

As described in the segmentation methods chapter 4, the high probability clusters P_{high} are extracted from a Bayesian probability texture as opening shape candidates. Adding to existing shape indices [Basaraner & Cetinkaya, 2017], the completeness index is introduced, which measures the r_{cp} ratio of outer shape area to inner-holes area. The candidates are rejected if their area is smaller than the chosen area threshold value b_s and if their completeness index score r_{cp} is smaller than r_{cp_t}. An example of extracted shapes is illustrated in Figure 5.3a.

As shown in Figure 5.3a, the contour lines of the extracted blobs can display noise, spikes, and inclusions. Morphological opening operation is applied to minimize the effect of spiky and weakly connected contours. It then approximates the outer shapes using the modified Douglas-Peucker algorithm, which considers parameters related to distances d_1, d_2, and angles a_1 [Douglas & Peucker, 1973]. Subsequently, these shapes are generalized to minimum bounding boxes, for which a modified rectangularity index [Basaraner & Cetinkaya, 2017] is calculated. The modification considers the relation of the bounding box sides a to b, where outliers are rejected based on the upper PE_{up} and lower PE_{lo} percentiles of the index score. Any blobs that are not connected to the ground are extended toward the ground surface to overcome the so-called border effect, visible in Figure 5.3a. The final generalized shapes are shown in Figure 5.3b.

5.2.2 Matching shapes with façade object's library

The target shapes provide semantics and regions of interest for a particular façade element. However, the sub-type of the detected class is missing, e.g., the detection provides a window but does not specify whether it is a rectangular-, circular-, or oval-shaped window. Here, a matching approach is introduced, which leverages the accuracy of MLS point clouds and the ubiquity of high-quality 3D façade elements' libraries. Specifically, an enhanced Bag-of-Words (BoW) approach [Csurka et al., 2004] is employed to match measured façade elements with those from the library without the rectangular assumptions. Figure 5.4 provides an overview of the proposed method. The training initiates with 3D model preparation

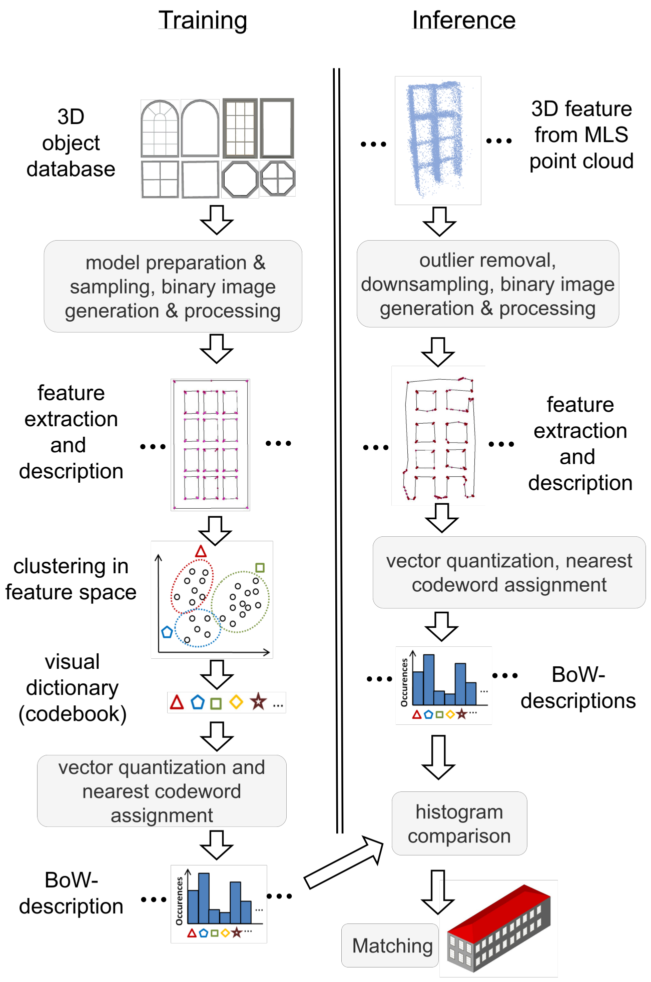

Training
Inference
3D object database
model preparation & sampling, binary image generation & processing
feature extraction and description
clustering in feature space
visual dictionary (codebook)
vector quantization and nearest codeword assignment
BoW-description
Occurences
3D feature from MLS point cloud
outlier removal, downsampling, binary image generation & processing
feature extraction and description
vector quantization, nearest codeword assignment
BoW-descriptions
histogram comparison
Matching

Figure 5.4 Matching detected façade elements with its representative sub-type from the 3D façade-object library.

and sampling, creating binary images from sampled point clouds. The method extracts and describes features from these images, which are then clustered to obtain a visual dictionary. Codewords are assigned to feature vectors by quantizing with Euclidean distance, and occurrences of each codeword provide model representations as bags of codewords. The codebook represents target point clouds during inference, and histogram distances are employed for comparison. The model with the closest histogram distance to a target point cloud is selected as the best match.

A key element in the proposed method is establishing correspondence between features extracted from MLS point clouds and CAD models representing the predefined library. The representation difference is highlighted in Figure 5.5, where the point-sampled-distance $d_{sampling}$ CAD model (a) is compared to an exemplary window acquired by MLS system (b).

Feature extraction

Point clouds are normalized after outlier removal and downsampling. To account for the decreasing point density with increasing height, an outlier removal is applied that depends on the average height of the objects of the MLS point cloud. Next, the point clouds are ortho-projected to a binary image, ensuring its frontal view. As illustrated in Figure 5.6, the standard image processing techniques are applied to enhance

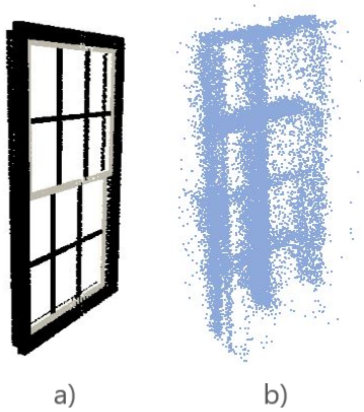


Figure 5.5 a) Point cloud sampled from a predefined library of CAD objects b) Point cloud acquired by MLS.

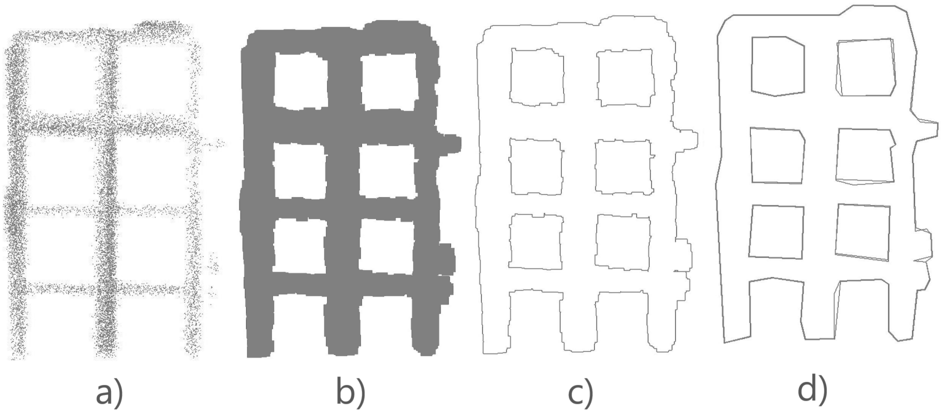


Figure 5.6 Image processing on an ortho-projected point cloud: a) projected image of a point cloud, b) dilated image, c) edge detection (Laplace) [Berzins, 1984], d) line simplification (Douglas-Peucker) [Douglas & Peucker, 1973].

the extraction of meaningful key points. Figure 5.7 a) shows that most identified keypoints are located at semantically meaningful positions.

The workflow utilizes Oriented FAST and Rotated BRIEF (ORB) [Rublee et al., 2011] descriptor as a key point point detector, which is based on the Binary Robust Independent Features (BRIEF) descriptor [Rublee et al., 2011; Calonder et al., 2012]. This descriptor is characterized by a resistance to noise and higher computational speed compared to other descriptors, such as Scale Invariant Feature Transform (SIFT) [Rublee et al., 2011; Lowe, 1999]. This method uses dense feature sampling as an alternative approach to interest point detection. This method is opposed to the concept of identifying and describing key points. In dense feature sampling, descriptors are sampled at points on a dense grid over the image, hence the name. Here, the ORB descriptor is sampled for each point in the dense grid. This approach allows extracting much information at the cost of higher computational intensity [Nowak et al., 2008].

Incorporating semi-global features

The semi-global information into the BoW approach is incorporated by using Histogram of Oriented Gradients (HOG) feature descriptor [Dalal & Triggs, 2005]. The fundamental concept of this descriptor is the investigation of the gradients and their orientation within the image on a dense grid. Normalized histograms of these gradients are established for each of these grid cells. The resulting one-dimensional vector characterizes the structure of the objects in the image [Dalal & Triggs, 2005].

Generally, shapes possess a limited number of features Bronstein et al. [2011]. This charachteristic poses difficulties in extracting large numbers of distinct features. Semi-global information can be used to

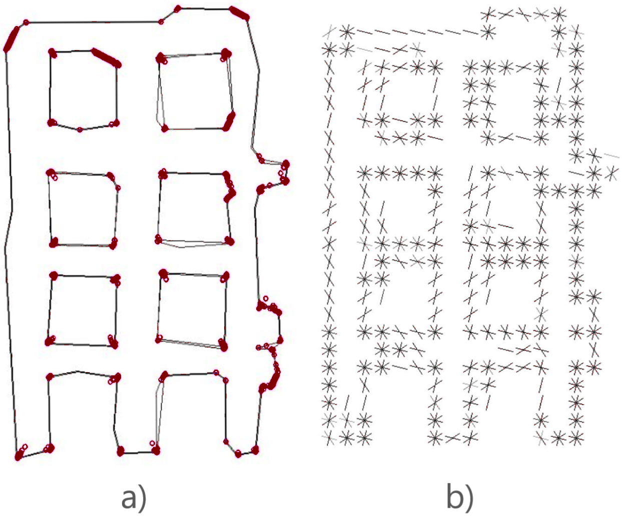


Figure 5.7 Examples for feature extraction: a) ORB-keypoints b) HOG-image.

mitigate the effects of this issue. However, semi-global and global features, cannot be directly integrated into the standard BoW method. Their (semi-) global uniqueness prevents the establishment of a frequency of occurrence. To overcome this issue, the proposed method incorporates HOG descriptors as semi-global information into the BoW approach. Figure 5.7 b) illustrates an example of the information obtained with HOG. A 2D diagram that displays the distribution of the gradients in the respective cell is shown for every cell in the image with a gradient. Figure 5.8 shows an overview of the proposed concept, where a structure similar to a histogram is introduced by considering each HOG descriptor variate as a separate histogram bin, with the value of the bin equaling the value of the corresponding variate. This structure comprises as many bins as HOG variates. The occurrence histogram of the BoW approach is concatenated with the histogram-similar structure constructed from the HOG-variates to a combined histogram.

Clustering and histogram similarity

The clustering is performed using the off-the-shelf K-Means clustering [Hartigan & Wong, 1979] algorithm in the BoW approach. Each descriptor variate is an axis in this clustering, which formulates the clustering problem as having the same dimensionality as the extracted feature descriptor. The setting of the number of clusters hyper-parameter n is critical for the performance of the BoW method since the meaningful assignment of data points to cluster centers depends on it [Kang & Yang, 2018].

The histogram similarity is measured to obtain the final match. To assess the similarity of the histograms within the presented BoW approach, the Pearson Chi-Square-Distance is used, measuring between the probability distributions of P and Q [Cha, 2008]:

$$D\chi^2(P,Q) = \sum_i \frac{(P(i) - Q(i))^2}{P(i)} \tag{5.1}$$

5.2.3 Prior model as solid and surface model

In the method, the assumption is made that the prior model at LoD1 or 2 is correct and represents a watertight, manifold, volumetric, solid structure, following the CSG modeling paradigm. This characteristic allows extracting the outer building shell, the boundary representations, also referred to as B-rep or mesh. The simplified differences between the representations are shown in Figure 5.9. Both representations are essential to the subsequent modeling process. On the one hand, the method capitalizes on solid's features of using Boolean operators in the CSG tree to create complex geometries; On the other hand,

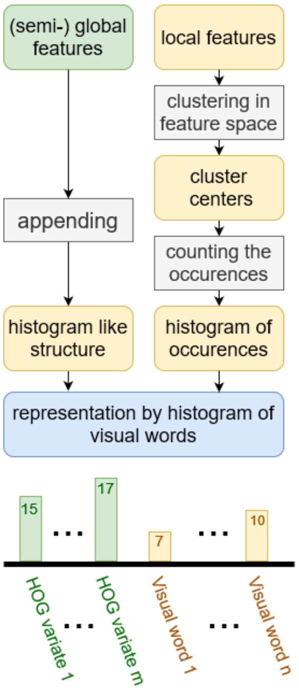


Figure 5.8 Incorporating semi-global features.

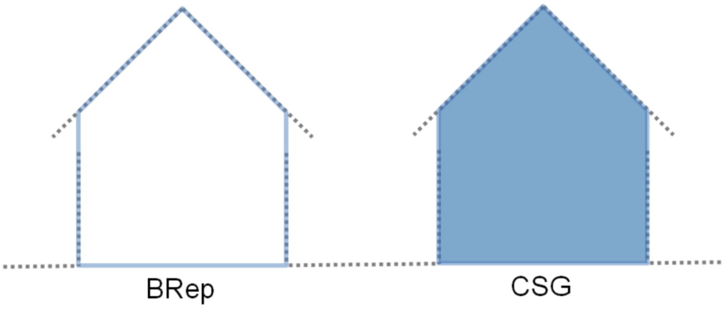


Figure 5.9 Simplified 3D building representation as B-rep and CSG geometry. Adapted from: [Huang et al., 2020].

the modeling paradigm of semantic 3D city models follows B-Rep, where the bounding surfaces describe volumetric extent [Kolbe & Donaubauer, 2021].

5.2.4 Enriching prior models by façade elements using boolean operators and predefined library

A 3D façade element is reconstructed based on the generalized shapes and their semantics. They are extruded in a perpendicular direction and at length measured from the investigated façade to the corresponding back-façade, given the detected class is *underpass*. The length is set for other openings as ratio parameter len_n.

This approach leads to the formation of prismatic objects (Figure 5.10), which represent the input solids for a CSG tree (Figure 5.11). It should be noted that the prismatic objects do not represent tangible objects but rather a free space. They are therefore first aggregated ($\cup$) and then subtracted ($-$) from a solid building geometry, as shown in Figure 5.11 [Wyvill & Kunii, 1985]. The subtraction remodels intersected polygons and closes inner-building gaps using face-intersecting vertices of the underpass space, thereby partitioning the polygons into smaller parts and closing gaps with polygon triangles. This approach is undertaken for large-size conflicts, such as underpasses.

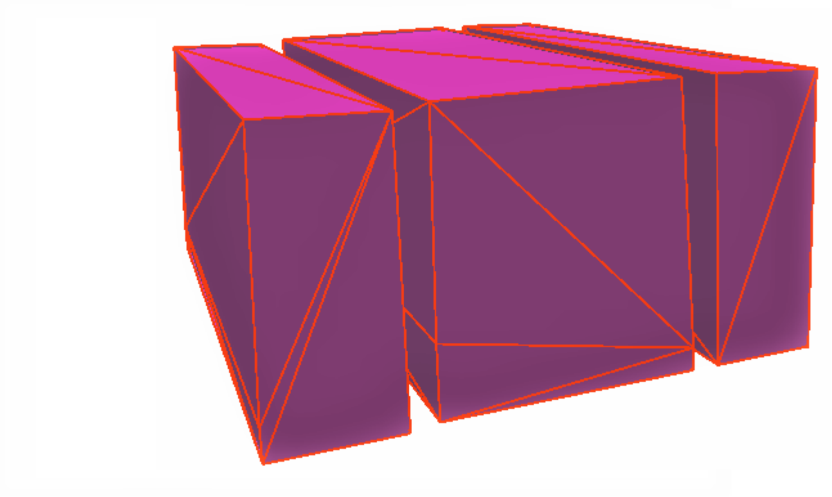


Figure 5.10 Prismatic 3D volumes used for subtraction: Example of an underpass subtraction space.

The cut shapes are not closed but filled with the fitted library object for the library-driven additions into raw models (Figure 5.12). Texture-identified objects and their shapes are used as fitting boundaries for 3D façade element models, which are loaded from a pre-defined library. The opening models' coordinate origin is erased and then placed in the bottom left corner of a model. The offset to global coordinates is calculated between the opening model origin and the corresponding instance bounding box's bottom left corner. After the shift, the rotation is performed as a difference between the façade's face and opening model orientation. As presented in Figure 5.14, aligned 3D models are scaled to fit bounding box boundaries; to avoid self-overlapping surfaces, the holes are cut in the input 3D model using the CSG operations. In contrast to the previously introduced CSG tree (Figure 5.11), the subtraction shape is not stored in the final model but rather removes the unnecessary parts of surfaces to avoid self-overlap for the fitted model. The final semantics are assigned following the identified class, the introduced function codes, and principles in Table 4.1, and Figure 5.2. The exemplary semantic reconstruction is shown in an example of an underpass in Figure 5.13.

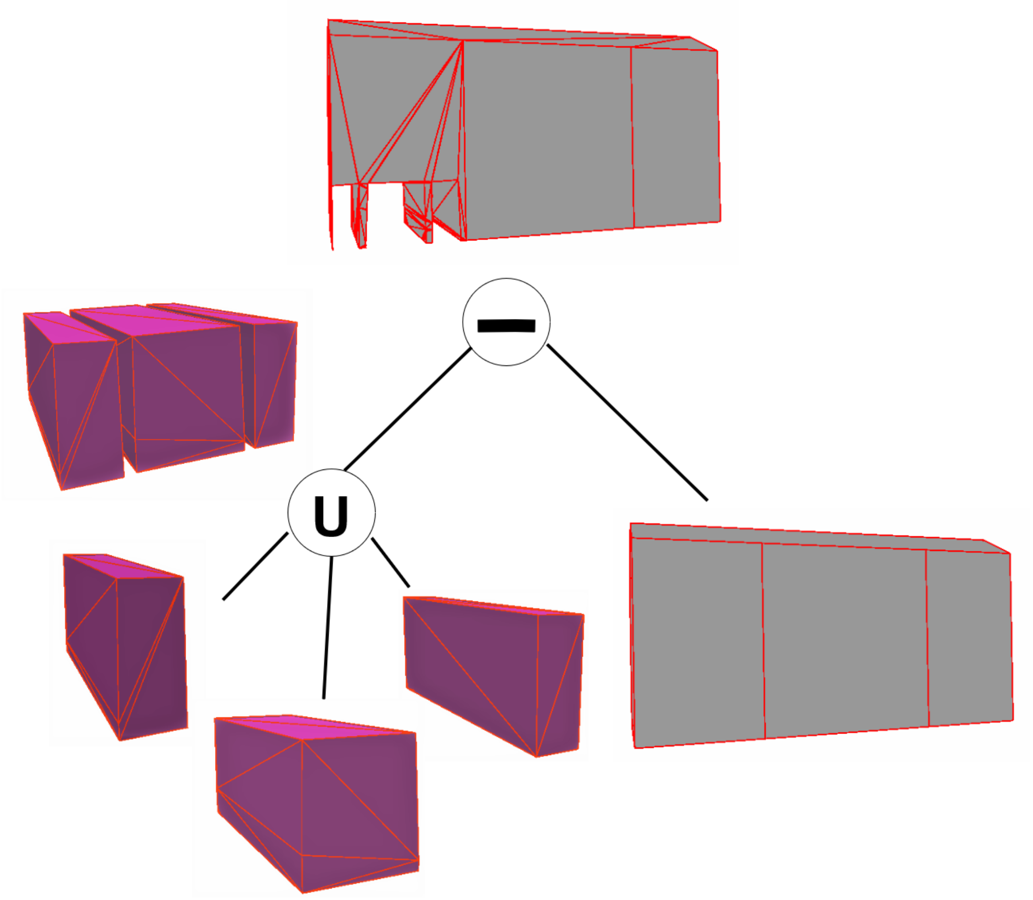


Figure 5.11 Subtraction CSG tree: Union of the subtraction volumes (∪) and its difference (−) to the solid volume of the building results in a refined building model.

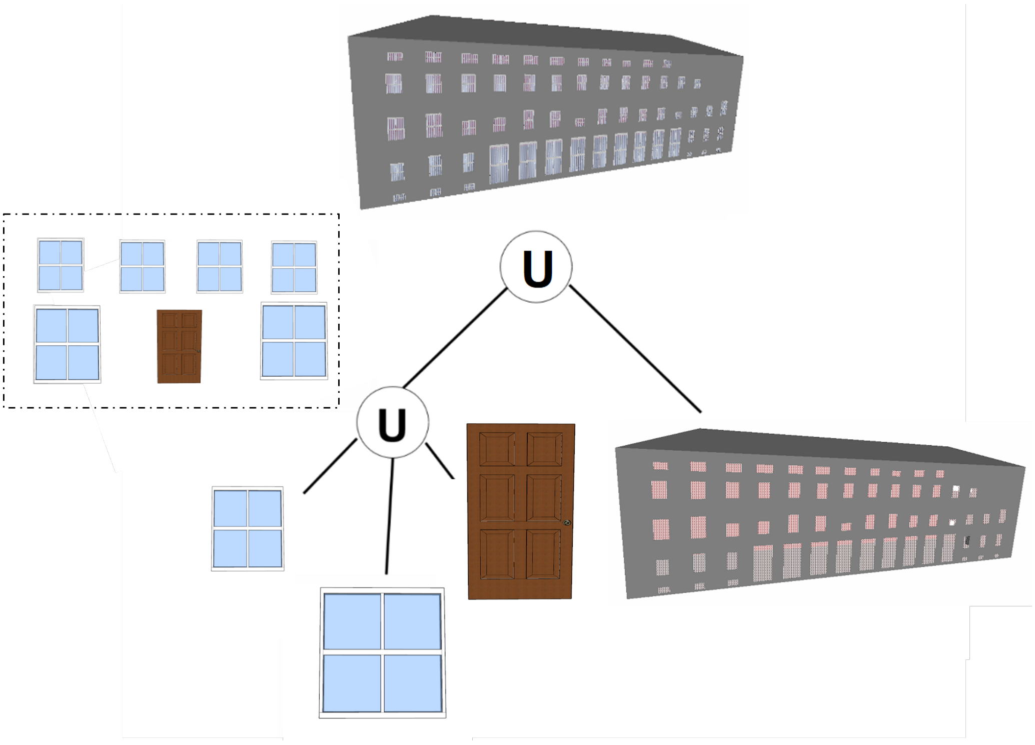


Figure 5.12 Addition CSG tree: Union of the pre-defined fitted models (∪) and its union (∪) to the hole-cut volume of the building results in a refined building model.

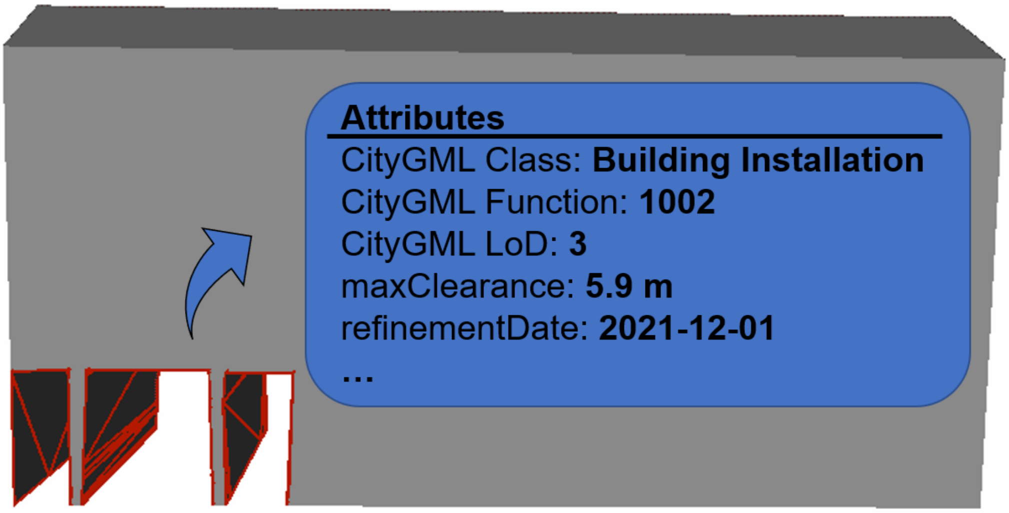


Figure 5.13 Exemplary enrichment in the underpass model: The semantically modeled underpass structure as a BuildingInstallation with the respective CityGML function: 1002.

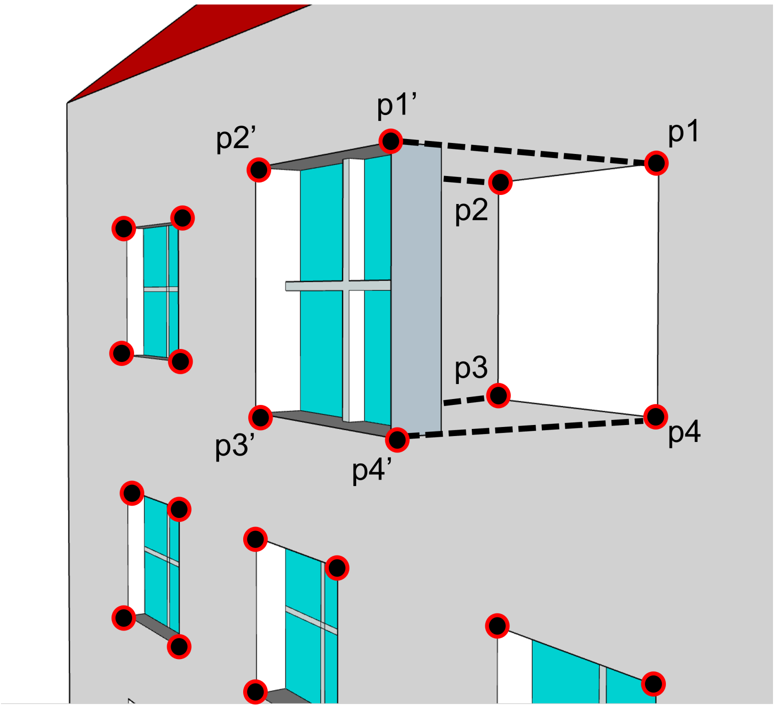


Figure 5.14 Exemplary junction points (red-black) of cut shapes as fitting points on an example of pre-defined 3D window models. The cut geometry is shown as hollow rectangle (p1, p2, p3, p4) whereas respective fitting points are on the pre-defined model corners (p1', p2', p3', p4').

6 Experiments

To validate the performance of the presented strategy, point cloud, semantic building models, and images were acquired. All the collected datasets used were open-sourced, with the exception of the proprietary MoFa 3D point cloud. The datasets are available under the open repository called TUM2TWIN[1] [Barbosa et al., 2023].

6.1 Experiments design

The experiments in this thesis are designed to validate the presented strategy and subsequent methods. Moreover, such experiments support answering the research questions and drawing conclusions from the presented work.

The fusion of point clouds and 3D models method (chapter 3) is tested by the comparison of the deviation rate between point clouds and models before and after performing the method. The quantitative measure used is Hausdorff distance, which complements the experiments by qualitative performance. Since the focus of the thesis is on MLS scanning, the experiments are perfomed on wall surfaces of city models, as roof surfaces are frequently not captured, disabling the reliable validation. These experiments utilize the TUM-MLS-2016 (subsection 6.2.1), MoFa 3D (subsection 6.2.3), and TUM-FAÇADE (subsection 6.2.2) datasets as a subset of MLS point clouds; The semantic 3D building models are encoded in the CityGML standard at LoD2 and LoD1, and represent the city of Munich (subsection 6.3.1) and Ingolstadt (subsection 6.3.2).

The method of 3D semantic segmentation of façade elements (chapter 4) is validated by analyzing the detection rate and accuracy of segmentation. The presented methods are compared to the baseline methods using the standard metrics in the field. These experiments utilize the TUM-MLS-2016 (subsection 6.2.1), MoFa 3D (subsection 6.2.3), and TUM-FAÇADE (subsection 6.2.2) datasets as a subset of MLS point clouds; The semantic 3D building models are encoded in the CityGML standard at LoD2 with textures and LoD3, representing the city of Munich (subsection 6.3.1).

The final experiments set concentrates on validating the 3D semantic reconstruction method (chapter 5). Here, the improvement rate is validated on an example of underpasses, where volume and surface gain are analyzed. The final reconstruction results on other opening types are analyzed against watertightness and deviations to the ground-truth LoD3 models and established mesh-driven reconstruction method. This experiments set utilize the TUM-MLS-2016 (subsection 6.2.1), MoFa 3D (subsection 6.2.3), and TUM-FAÇADE (subsection 6.2.2) datasets as a subset of MLS point clouds; The semantic 3D building models are encoded in the CityGML standard at LoD2 with textures and LoD3, representing the city of Munich (subsection 6.3.1). All experiment implementations are openly available under my repositories [2]. The tests were conducted on Intel Core i7-10750H CPU 2.60 GHz x 12, 16 GB RAM Ubuntu 20.04.4 LTS. The used programming languages included Python, C++, R, and software suite FME and open source MeshLab [Cignoni et al., 1998]. The primarily used libraries comprised PyTorch [Paszke et al., 2019], Open3D [Zhou et al., 2018], OctoMap [Hornung et al., 2013], PCL [Rusu & Cousins, 2011], and bnspatial [Masante, 2020].

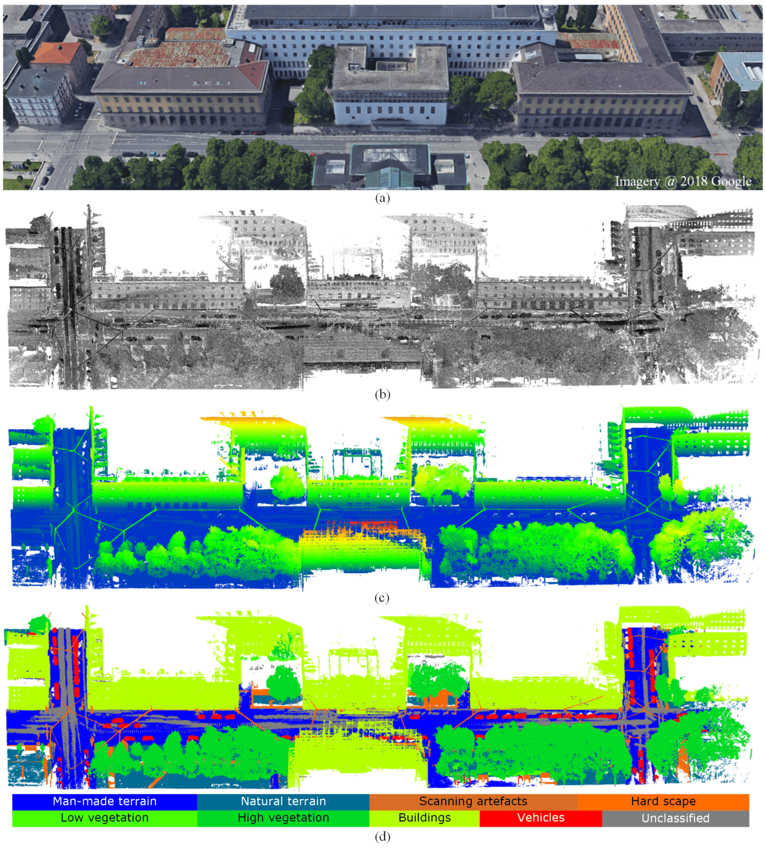
[1] https://tum2t.win/

[2] https://github.com/OloOcki

6.2 Point cloud data

6.2.1 TUM-MLS-2016

The point clouds in TUM-MLS-2016 were collected via obliquely mounted two Velodyne HDL-64E LiDAR sensors mounted on the Mobile Distributed Situation Awareness (MODISSA) platform. The Velodyne HDL-64E scanners were scanning at 10 Hz rotational frequency while acquiring 130,000 range measurements per rotation with a maximal distance of 120 m. This type of laser scanner possesses 64 laser rangefinders, which have a vertical field of view of 26.8° divided into 64 scan lines. The scanners were mounted on wedges with a 25° angle to the horizontal and rotated outwards at a 45° angle. The entire point cloud covered an urban area with an inner and outer yard of the Technical University of Munich main campus. It covers approximately $0.2km^2$, with around 1km along roadways. The scans were recorded synchronously with the position and orientation data of an Applanix POS LV 520, where the inertial navigation system was supported by the real-time kinematic (RTK) correction data of the German satellite positioning service (SAPOS), which ensured geo-referencing. The dataset includes more than 40 million annotated points with labels for eight classes of objects [Zhu et al., 2020]. The expected global accuracy is less than 50cm, while relative accuracy 5-10cm.



(a)

(b)

(c)

(d)

Figure 6.1 Part of the dataset showing the TUM main entrance from a) Google Earth, 2018 b) Point cloud colored by signal intensities c) Point cloud colored by height. d) Point clouds colored by eight classes. Source: [Zhu et al., 2020].

6.2.2 TUM-FAÇADE

The TUM-FAÇADE dataset is derived from the TUM-MLS-2016 point clouds, where the former enriches the latter in 17 façade-level semantic classes. Due to that, the expected accuracy correspond to these of TUM-MLS-2016. The dataset comprises 33 annotated facades totaling approximately 333 million façade-level labeled and geo-referenced points. As discussed in subsection 4.4.2, and due to the current data-classes-imbalance, LoFG3 are defined as pertinent for the experiments. Therefore, 17 TUM-FAÇADE's classes are combined into seven by merging: *molding* with *decoration*; *drainpipe* with *wall*, *outer ceiling surface* and *stairs*; *floor* with *terrain* and *ground surface*; *other* with *interior* and *roof*; *blinds* with *window*; whereas *door* remained intact [Wysocki et al., 2022c].

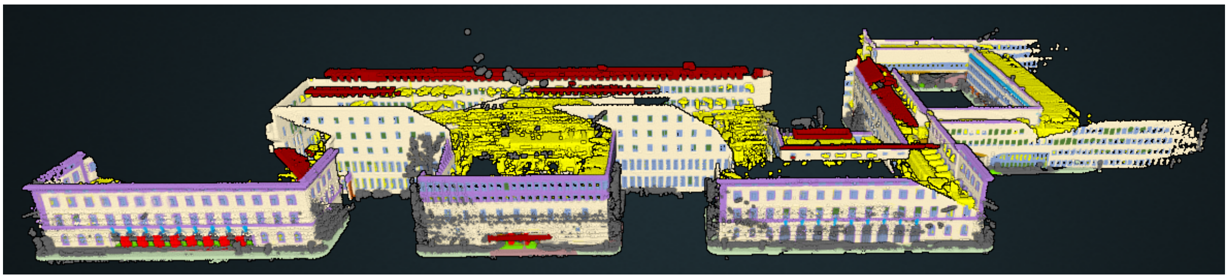


Figure 6.2 The TUM-FAÇADE dataset showing the TUM main entrance, colored by classes.

6.2.3 MoFa 3D

The MoFa 3D point clouds were acquired at the TUM campus in 2021 and covered approximately the same area as the TUM-MLS-2016 dataset. The additional area was acquired for the town center of Ingolstadt, Germany. The proprietary mobile MoSES (ger. Mobiles StraßenErfassungsSystem) platform was used to acquire the data. The system is equipped with two laser scanners and synchronized cameras. The point cloud was geo-referenced by a proprietary mobile mapping platform, supported by the German SAPOS RTK system [3D Mapping Solutions, 2023]. The expected global accuracy is less than 20cm, while relative accuracy 1-3cm.

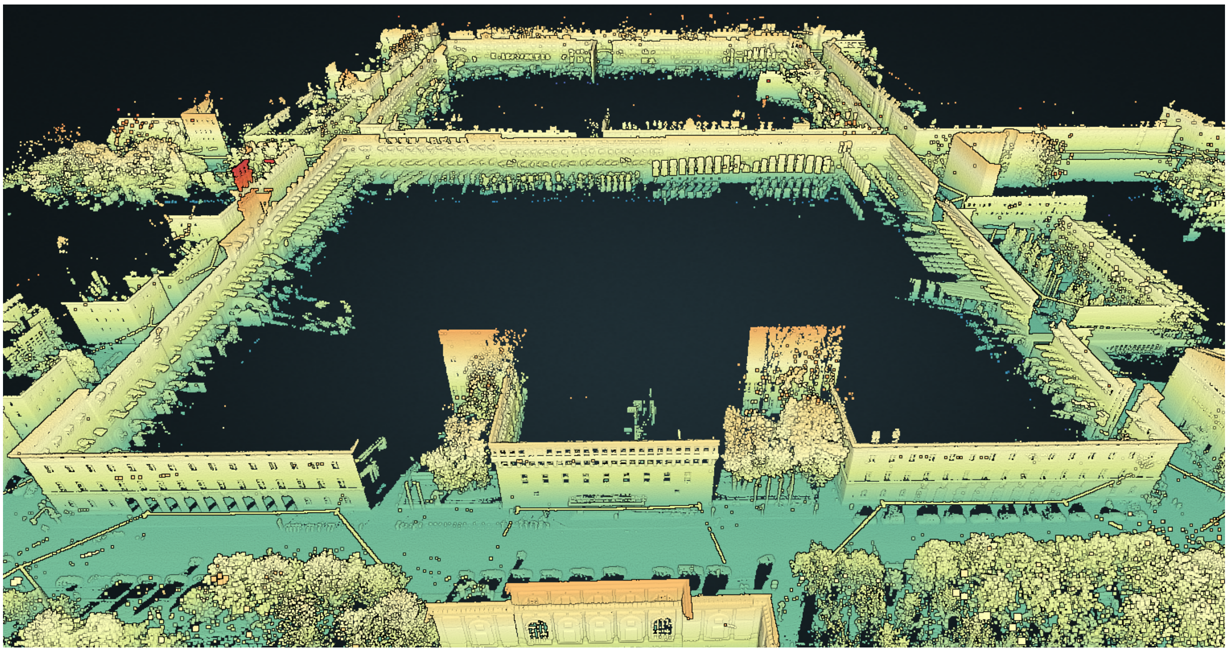


Figure 6.3 The MoFa 3D dataset showing the TUM main entrance, colored by height.

6.3 Semantic 3D building model data

6.3.1 CityGML models of Munich, Germany

The open data CityGML-compliant building priors at LoD2 were acquired from the state-wide open access portal of Bavaria, Germany [Vermessungsverwaltung, 2023], which were created using 2D cadastre footprints in combination with aerial observations [Roschlaub & Batscheider, 2016]; comparable results can be achieved with methods such as PolyFit [Nan & Wonka, 2017] or City3D [Huang et al., 2022]. The textures were acquired manually at an approximately 45° horizontal angle using a 13MP rear camera of a Xiaomi Redmi Note 5A smartphone and projected to the respective faces: this approach simulated terrestrial acquisition of a mobile mapping unit or street-view imagery where no ortho-rectifications were applied [Hou & Biljecki, 2022]. The selected building models were refined manually into LoD3 based on a combination of TUM-FAÇADE, MoFa3D, and textured LoD2 models. The so-called *building 23* is one of the prominent examples, as it has been commonly used as a validation object for various methods [Wysocki et al., 2022c, 2021c; Hoegner & Gleixner, 2022; Tuttas & Stilla, 2013; Wysocki et al., 2022a]. Expected global accuracy is 1-3cm (footprint) while relative accuracy of LoD3 is 1-3cm (facade elements), and around 25cm for LoD2.

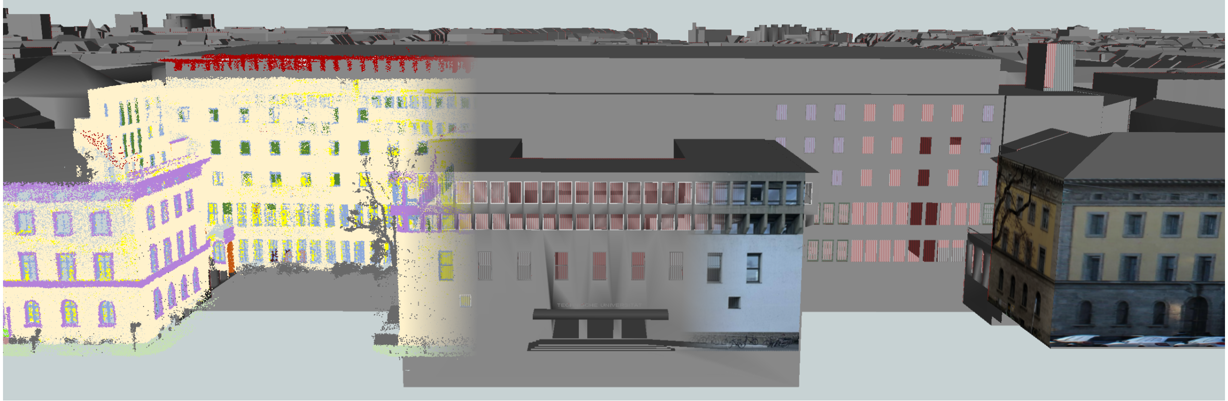


Figure 6.4 The manually created LoD3 building models, LoD2 textured building models, and TUM-FAÇADE point clouds representing the Technical University of Munich, Munich, Germany. Source: [Barbosa et al., 2023].

6.3.2 CityGML models of Ingolstadt, Germany

Analogically to the building models dataset of Munich (subsection 6.3.1), the Ingolstadt models were LoD2 CitGML-compliant and downloaded from the state-wide open access portal of Bavaria, Germany [Vermessungsverwaltung, 2023]. As such, they used the same source data and methods for creation as the one representing Munich. In contrast to the Munich-based models, these models had no textures. Additionally, HD-maps-derived LoD1 building models were used. Their conversion was performed using an open source software r:trån [Schwab et al., 2023a]. The validation dataset for the reconstruction was open data LoD3 building models, created on the basis of the MoFa 3D point cloud [Schwab et al., 2021]. Expected global accuracy is 1-3cm (footprint) while relative accuracy of LoD3 is 1-3cm (facade elements).

6.3.3 Façade elements library

The library comprises a set of acquired four CAD models from the SketchUp 3D Warehouse library and is complemented with the façade elements of LoD3 building models from the TUM2TWIN and Ingolstadt datasets [Trimble Inc., 2021]. Figure 6.6 shows the part of the library. The models were selected so that each window that is present in the TUM2TWIN dataset at least matches one of the models. Additional, out-of-distribution shapes were added, for example, a window that is of octagon-like shape as a model without a matching window in the dataset. A manual altering of the CAD models was performed to remove

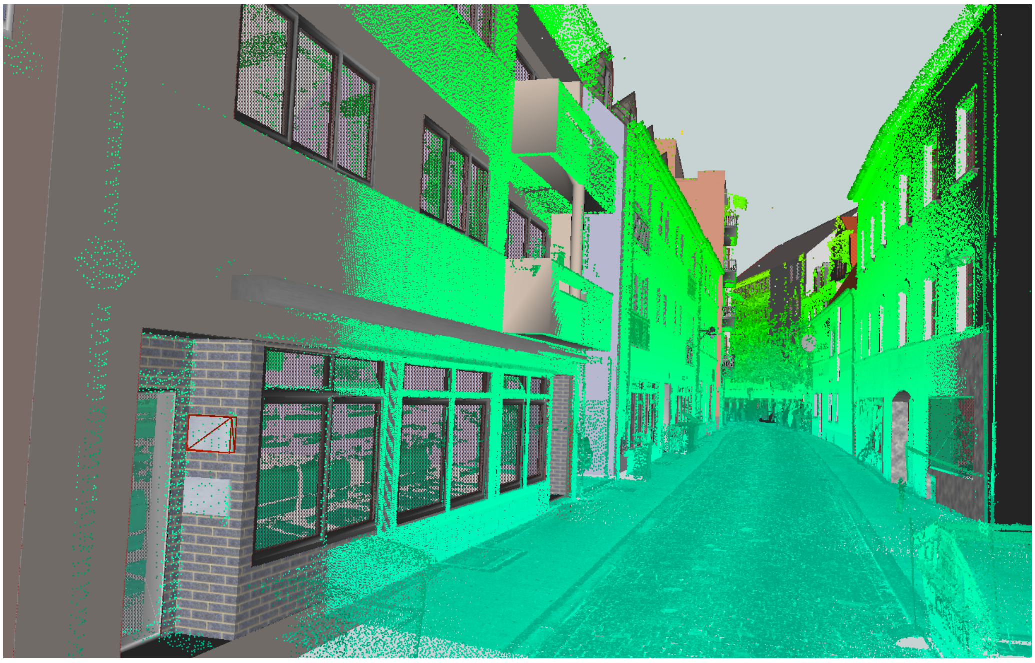


Figure 6.5 The manually created LoD3 building models of Ingolstadt overlaid with a MoFa 3D point cloud. Source: [Schwab et al., 2021].

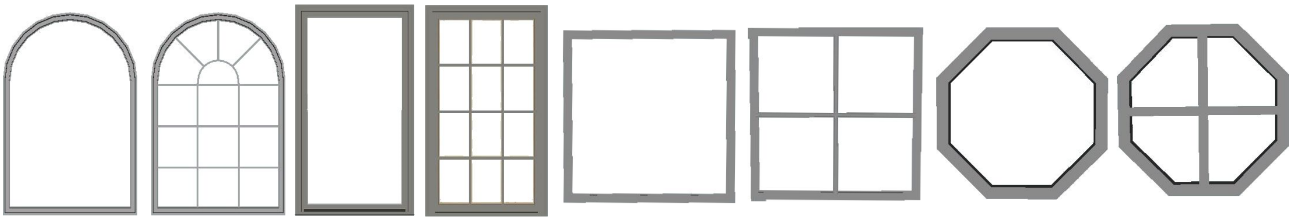


Figure 6.6 Simplified view of the different window types: A subset of the whole 3D façade elements library.

non-relevant parts and incorrect geometries or to add or remove window bars. In total, there were nine window types and one door type acquired.

6.4 Parameter settings

For the visibility analysis and uncertainties, the parameters were set to the size of voxels to $v_s = 0.1$ m and initialized with a uniform prior probability of $P = 0.5$ to perform the ray casting on an efficient octree structure [Hornung et al., 2013]; the standard [Hornung et al., 2013; Tuttas et al., 2015] clamping and log-odd values were used. Namely, clamping values were assigned as $l_{min} = -2$ and $l_{max} = 3.5$, corresponding to $P_{min} = 0.12$ and $P_{max} = 0.97$, respectively; while log-odd parameters were fixed to $l_{occ} = 0.85$ for *occupied* and $l_{emp} = -0.4$ for *empty* states, corresponding to $P_{occ} = 0.7$ and $P_{emp} = 0.4$, respectively. The uncertainty of building models and point clouds was assigned considering their reported global positioning accuracy. As such, the parameters of building models were set to $\mu_1 = 0$ and $\sigma_1 = 3$, while for the TUM-MLS-2016 and MF point clouds were set to $\mu_2 = 0$, $\sigma_2 = 2.85$ and to $\mu_2 = 0$, $\sigma_2 = 1.4$, respectively. The threshold ratio for the conflicted pixels area a_1 and the total surface area a_2 were set to $r_{c_{min}} = 0.1$ and $r_{c_{max}} = 0.6$.

For the semantic segmentation, the following parameters were used. For the modified PT data preprocessing, redundant points were removed within a 5 cm radius, which resulted in 10 million points; the point cloud was split into 70% training and 30% validation subsets. Regarding the geometric features, the optimal search radius d_i was set following Grilli et al. [2019]: As for the features *roughness, volume density,* *omnivariance, planarity*, and *surface variation* the radius was set to $d_i = 0.8\ m$; whereas for *verticality* to $d_i = 0.4\,m$. For the image segmentation, a pre-trained Mask-RCNN on the COCO dataset [Lin et al., 2014] was used. The inference was fine-tuned with 378 base images of the CMP façade database [Tyleček & Šára, 2013], where two classes were selected for training: *door* and *window* including *blinds*. As P_{high} pixels in the Bayesian network, the values higher than $P_{high} = 0.7$ were considered. To reject outliers, the modified rectangularity percentiles were fixed to $PE_{up} = 95$ and $PE_{low} = 5$.

7 Results

7.1 Fusion of 3D models and point clouds

The calculations in this section were performed on the semantic 3D building models of Ingolstadt and the TUM main campus described in the previous section 6.3, while the point cloud data comprised TUM-MLS-2016 and MoFa 3D point clouds described in the previous section 6.2.

7.1.1 Coregistration

Based on the obtained matrices, the ICP point-to-plane algorithm achieved a rectification of error of around 0.5 [m] to 0.04 [m]. This translated to new CI at rates $e_1 = 0.04$ [m], $CL_1 = 90\%$ for MLS point cloud and $e_2 = 2.00$ [m], $CL_2 = 95\%$ for generalization of CityGML building model. This estimated CI at 1.44 [m] for the possible range of building elements; this is depicted in Figure 7.1 as GR. Since the BayNet aims to find confirmation of the raw CityGML wall models, too, the CI for these geometries were obtained. Thereby, $e_1 = 0.04$ [m], $CL_1 = 90\%$ for MLS point cloud and $e_2 = 0.03$ [m], $CL_2 = 95\%$ for CityGML wall objects was estimated at 0.03 [m] CI.

This step allowed the creation of 3D boxes within which density and uniformity estimation were performed. Since the buildings in question had no major extruded parts, these tests were performed without vertical-like filtering implementation. Also, since the MLS acquisition geometry implies capturing of outer walls, the roof elements were assumed to be uncovered. This test rejected 15 wall segments as not adequately covered, and six passed both tiers of measurement.

7.1.2 Bayesian network performance

This BayNet was tailored to detect confirmation between raw CityGML wall geometry and MLS point clouds. As such, the utilized BayNet was a distilled version of the general concept shown in Figure 3.4. The BayNet was designed in GeNIe Modeler[1]. The previously obtained discretized CIs were the basis for soft spatial evidence incorporation. In the case of this test, these 3D boxes were further discretized to 3D vertical patches. Each patch had a size of 0.05 x 0.05 [m] and height derived from the respective CI 3D box. As shown in Figure 7.1, the network consisted of 4 evidence, two nodes, and one target node. While respective CPT are not shown in Figure 7.1, they are available under the attached repository[2]. The soft evidence explicitly incorporated probabilities derived from the CI calculation: These were propagated throughout the network to the target node. The network inference was performed in the R bnspatial package[2]. The probability for target node state *Occupied* (see Figure 7.1) was calculated using the bnspatial function. For the designed BayNet, probability thresholds were set accordingly: $P_{high} > 0.7$, $0.7 <= P_{moderate} => 0.3$, $P_{low} < 0.3$.

7.1.3 Fusion results

Each spatial output split by probability thresholds might trigger various further investigations. The coarse fusion of both datasets might be plausible for some applications (see Figure 7.2). The core fusion task, however, was performed based on the high confirmation rate. Thus, spatial objects with P_{high} were used for the coupling of MLS point clouds and CityGML models of facades. Each block is linked to a specific

[1] BayesFusion, LLC (http://www.bayesfusion.com/)

[2] https://cran.r-project.org/web/packages/bnspatial/vignettes/bnspatial.html

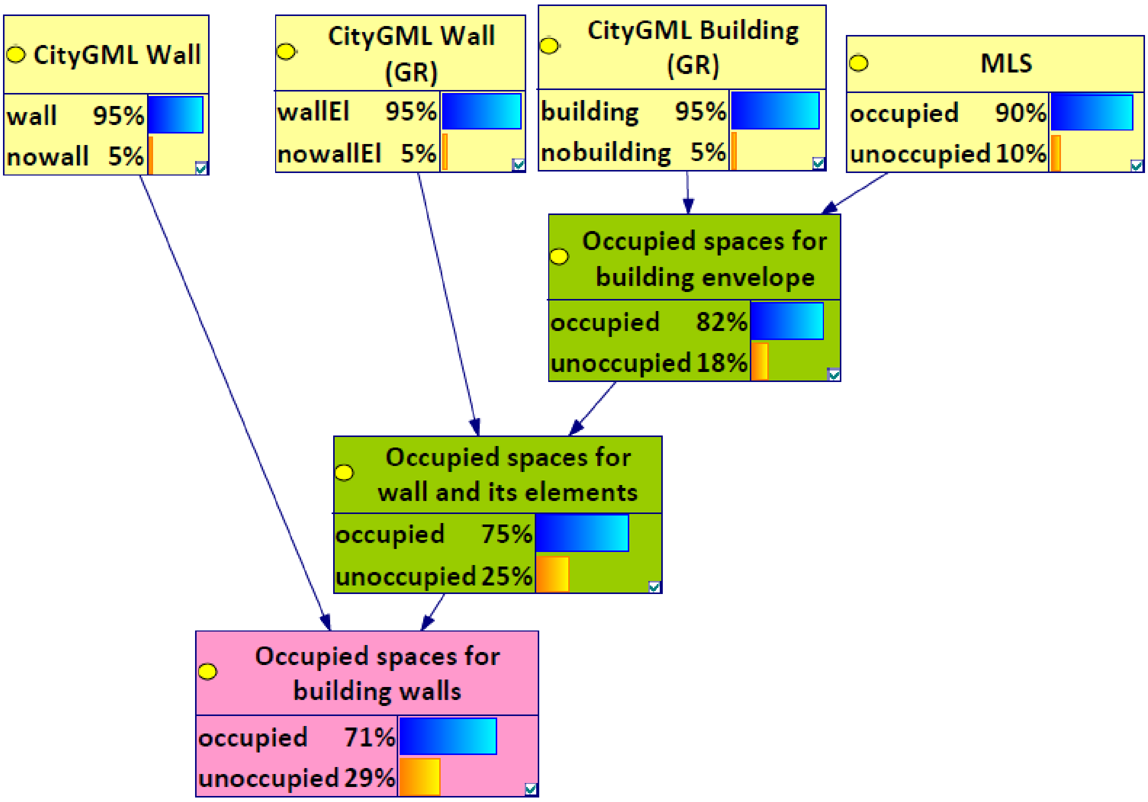


Figure 7.1 The BayNet with soft pieces of evidence (yellow), nodes (green), and target node (pink) and associated inference PPD scores. GR stands for a generalized range of city models with associated uncertainty.

raw model and point cloud part. Additionally, the blocks inherited calculated probabilities that can serve as metadata for further processing steps or a database update. As depicted in Figure 7.3 and in Figure 7.4, the confirmation of closely and densely aligned point clouds with raw 3D models was achieved. The vertical 3D patches were suitable for coarse wall confirmation but should be finer for other purposes. For instance, 3D vertical patches did not consider building openings. However, this could be mitigated by, for example, the utilization of grid voxels or octree structures of small size. The method achieved satisfactory results also regarding $P_{moderate}$ and P_{low} thresholds. As depicted in Figure 7.4, $P_{moderate}$ together with P_{high} might serve as a trigger for the façade refinement purposes. For instance, P_{high} patches were the areas not desired for the refinement. Whereas $P_{moderate}$ marked possibly unmodeled elements or strong deviations w.r.t. the raw model. Moreover, the inner points were marked as P_{low} probability to belong to the façade. This might mitigate their negative impact on the façade refinement process. The one-sided Hausdorff distance [Cignoni et al., 1998] was used to quantitatively measure fusion performance. The tests were conducted on a raw situation before applying the strategy and an end fusion using P_{high} areas. To compare deviations under constant conditions, the maximal deviation was set to 0.5 [m]. The results are shown in Table 7.1 and in Figure 7.5.

Fusion set	MLS vs. LoD2 façade		
	max	μ	RMS
P_{high}	0.5	0.04	0.06
Raw	0.5	0.20	0.23

Table 7.1 Comparison of raw and P_{high} fusion set deviations - all metrics given in meters.

To show efficacy of the method on other data, a set of experiments were conducted on Ingolstadt dataset, where uncertainties for building models were assumed, and for point clouds no uncertainties were taken

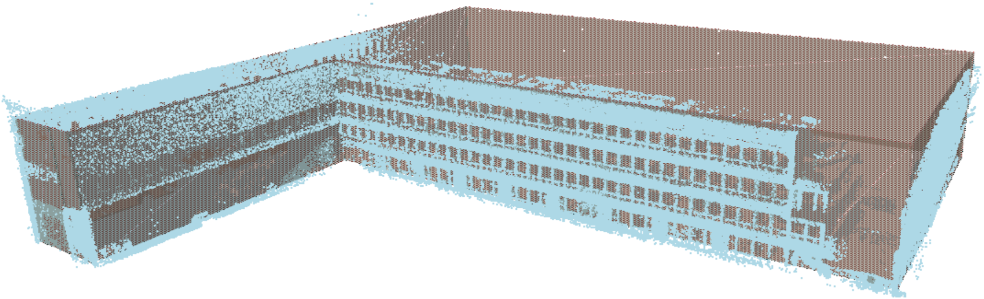


Figure 7.2 The raw CityGML façade geometry (gray) and coregistered MLS point cloud (blue) within CI for walls.

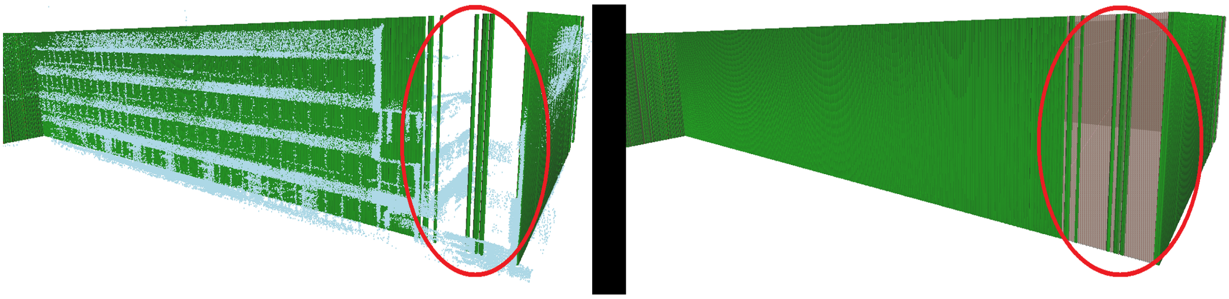


Figure 7.3 P_{high} blocks (green) indicate confirmed parts of the raw façade geometry (gray) by MLS point clouds (blue).

into account. The other experiment set was conducted on the Ingolstadt datasets, where the vertical-like objects consisted of 87 buildings in this test scenario. The coverage analysis has rejected 18 buildings from the reconstruction process. This accelerated the reconstruction process and avoided reconstruction errors. Furthermore, only those LoD1 walls were accepted for further reconstruction for which the corresponding LoD3 wall contained an *DataAvailable* attribute of *Sufficient* (except two on the periphery of the area). These attributes have been added by the creators of the LoD3 dataset and document the MLS point cloud coverage of the LoD3 buildings. Similar to the road segment experiment, the assumption of rigid boundaries has certain advantages and disadvantages that also apply to the buildings. For example, due to the rigid borders of the LoD1 input model, the modeling of walls of gable roof buildings present in LoD2 and 3 was prevented, as shown in Figure 7.6. On the other hand, an increased depiction of details on the building surface, such as windows and doors, can be observed. These are not present in LoD2 but LoD3 building models. Ultimately, the refined structure shows more geometric details and captures even small deviations compared to the generalized geometries of the LoD3 building model, as shown in Figure 7.6. Moreover, the additional building features not present and significantly distant from the searched plane in the input dataset, such as balconies (in case of LoD1), are not reconstructed. Also, objects adjacent to buildings, such as tree branches, can be misclassified as building parts. This case happens due to the assumption that the RANSAC algorithm should find one portion of inliers per building feature. However, this only occurs if the object is located within the respective accuracy range on the prolonged plane direction and within the plane margin introduced by the RANSAC fitting plane model. This can be extended by the introduction of another stopping criterion. Since the walls of the LoD1 building models are the subject of the refinement, this comparison reflects the deviations between the raw buildings and the reconstructed surfaces that shall be perceived as a gain of the method. The validation, however, is performed using the building models in LoD2 and LoD3. As shown in Table 7.2, the validation against LoD3 confirms that the refined structures at the highest octree level 12 have the highest quality w.r.t. the chosen measure.

The discrepancies encountered when comparing to the LoD2 models are due to the different measurement techniques. The outliers present in the *max* column of Table 7.2 are caused by falsely segmented points or balconies, as shown in Figure 7.7, where the histogram indicates that most faces deviate by about 0.2 meter distance.

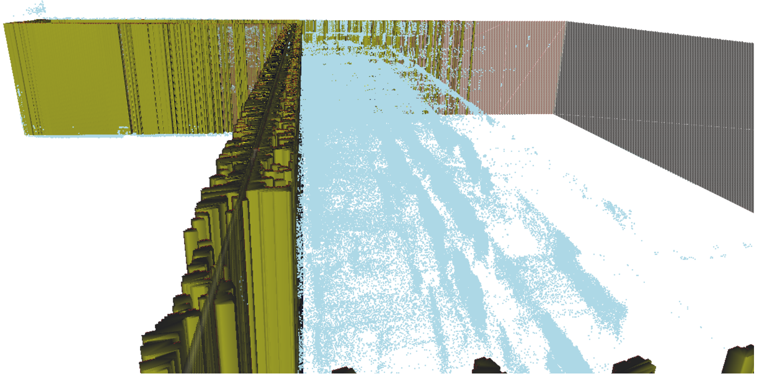


Figure 7.4 $P_{moderate}$ as an indicator of possible deviations between raw model (gray) and MLS observations (blue). Such deviations indicate the possibility of unmodeled surfaces.

Octree level	LoD1 (raw geom.)			LoD2 (ref. geom.)			LoD3 (ref. geom.)		
	max	μ	RMS	max	μ	RMS	max	μ	RMS
8	6.85	0.18	0.41	6.80	0.41	0.69	4.78	0.18	0.43
10	3.93	0.16	0.31	6.70	0.37	0.60	4.40	0.13	0.31
12	2.94	0.16	0.31	5.88	0.39	0.63	3.59	0.11	0.27

Table 7.2 Comparison of refined LoD1 building geometries with the geometries of LoD1-3 building models, with all metrics given in meters

7.2 3D semantic segmentation of façade elements

The calculations in this section were performed on the semantic 3D building models of the TUM main campus described in the previous section 6.3; while the point cloud data comprised TUM-FAÇADE, TUM-MLS-2016, and MoFa 3D point clouds described in the previous section 6.2.

7.2.1 Detection rate

The methods of Hoegner & Gleixner, 2022, [Hoegner & Gleixner, 2022] were tested on the three facades of the *building 23* at the TUM campus using the TUM-MLS-2016 data; thus the validation of the detection accuracy was conducted using the same setup and the manually modeled LoD3 building (Table 7.3). To show the ratio of the detection rate to the laser-covered rate, the metrics were introduced for all existing façade openings (AO) and only laser-measured façade openings (MO).

The multimodal fusion enabled a higher detection rate and still maintained a low false alarm rate. If compared to the Hoegner & Gleixner (H&G) Hoegner & Gleixner [2022] (ray-to-model) and CC Wysocki et al. [2022a] (ray-to-model with point cloud semantics) methods, Scan2LoD3 Wysocki et al. [2023b] (ray-to-model with point cloud and image semantics) achieved higher detection rate on the TUM dataset by 10% and 6%, respectively (Table 7.3 and Figure 7.14). The MoFa 3D (MF) map provided more accurate results (i.e., 91% of all openings correctly detected) owing to higher point cloud global accuracy and complete façade A coverage; also other maps complemented the MF's laser-observed openings, as exemplified by façade B (Table 7.3).

7.2.2 3D semantic segmentation

For analysis of different combinations of geometric features, the importance of different features is calculated with random forests (Figure 7.8). As shown in Figure 7.8, coordinate components x, y, z were the most powerful factors in random forests classification, which is an intuitive and expected result for any 3D segmentation approach. Features with a score over 0.05 were also considered influential in this part of the

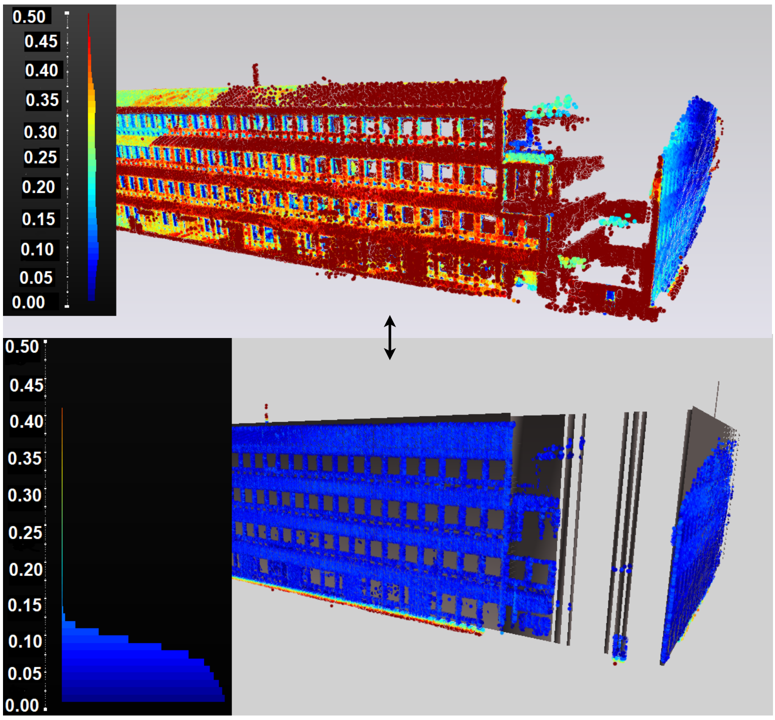


Figure 7.5 Raw (above) and P_{high} (below) fusion sets with deviations obtained by the Hausdorff metric given in meters.

experiment, including *surface variation*, *planarity*, PCA components, and the second dimension of the second eigenvector. The rationale behind these features selection is that they are ideal for facade semantic segmentation tasks. On the one hand, *planarity* encodes planar-like structures which often characterise human-built objects, such as facades. On the other hand, *surface variation* captures the opposite facade elements and noisy features. The standard PCA complement feature space by additional sublime feature differences. Based on this result, there were two different feature combinations selected; one selection included nine kinds of features as input: *planarity*, *surface variation*, *omnivariance*, three PCA components, and three dimensions of the second eigenvector, while for the other, *omnivariance*, the first dimension and third dimension of second eigenvector were removed due to their insignificant performance in random forests classification, only top six features in terms of importance including *surface variation*, *planarity*, PCA components and the second dimension of the second eigenvector were kept for comparison. These two selections of geometric features served as an additional input for DL models PointNet and PointNet++.

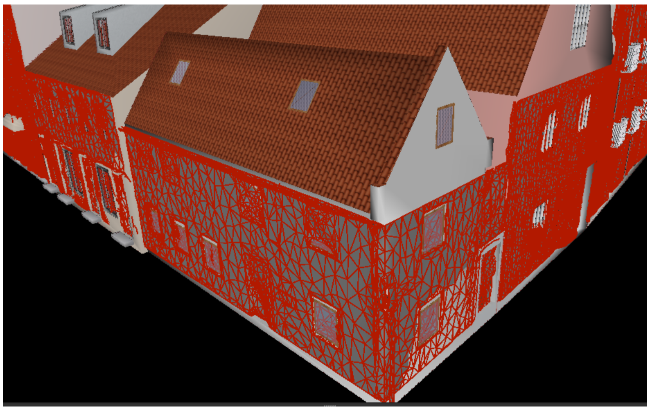


Figure 7.6 Fused point cloud represented by mesh (red edges) with LoD1 building models juxtaposed with LoD3 buildings models.

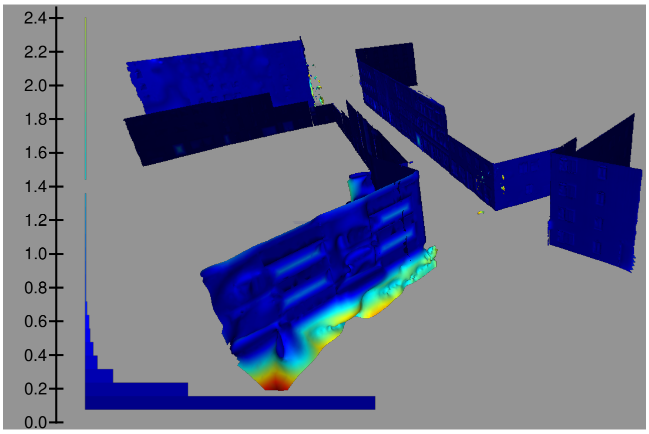


Figure 7.7 Deviations, obtained by Hausdorff metric given in meters, projected to refined structures at level 12 in comparison to LoD3 buildings.

	H&G				CC				Scan2LoD3 (TUM)				Scan2LoD3 (MF)			
	A	B	C	Tot	A	B	C	Tot	A	B	C	Tot	A	B	C	Tot
AO	66	17	20	**103**	66	17	20	**103**	66	17	20	**103**	66	17	20	**103**
MO	60	17	10	**87**	60	17	12	**87**	60	17	12	**89**	66	12	18	**96**
D	60	15	4	**75**	60	15	6	**81**	60	16	11	**87**	65	16	16	**97**
TP	60	12	4	**76**	60	15	5	**80**	60	16	11	**87**	65	14	15	**94**
FP	0	3	0	**3**	0	0	1	**1**	0	0	0	**0**	0	0	1	**3**
FN	6	5	16	**27**	6	2	15	**23**	6	1	9	**16**	1	3	5	**9**
DA	91	71	20	**74**	91	88	25	**78**	91	94	55	**84**	98	82	75	**91**
FA	0	0	0	**4**	0	0	17	**1**	0	0	0	**0**	0	12	6	**3**
DM	100	71	40	**87**	100	88	42	**90**	100	94	92	**98**	98	117	83	**98**

Table 7.3 Detection rate for all openings (DA) and laser-measured openings (DM) and the respective false alarm rate (FA) for facades A, B, and C (AO = all openings, MO = laser-measured openings, D = detections, TP = true positives, FP = false positives, FN = false negatives).

Validation of improved models for façade-level classification was based on the overall accuracy of different approaches. Except for comparison between with and without geometric features, different combinations of geometric features were also tested in this study. As shown in Table 7.4, the dataset labeled with 59 refers to performance on building No 4959459, and 23 means validation on building No 4959323. In addition to a dataset with only coordinates as input (XYZ), a dataset with nine kinds of geometric features (XYZ+9F) and six kinds of geometric features (XYZ+6F) were also evaluated. From statistical results in Table Table 7.4, it is clear that among all the various combinations, coordinates with six kinds of geometric features (XYZ+6F) as input performs better for PointNet++. The results of XYZ and XYZ+6F are shown for the qualitative analysis of influence of geometric features. The confusion matrix in Figure 7.11 represents comparison of classification results on different classes using PointNet++ with and without geometric features. Moreover, a visual examination (Figure 7.12 also shows how PointNet++ performs with or without geometric features on the two unseen buildings.

Further experiments on the building 23 were performed using the PT method with a combination of geometric features and the presented CC method back-projected to the point clouds. Semantic segmentation results were again validated on unseen ground-truth point clouds of the TUM-FAÇADE dataset. For evaluation, the following metrics were used: The overall accuracy (OA); F1 score per class; and average: precision (μP), recall (μR), F1 score (μF1), and intersection over union (μIoU). The *arch* and *column* classes were omitted in the validation, since they were absent in the ground-truth building. As shown in Table 7.5, for the baseline of the validation served the PT network [Zhao et al., 2021]. The presented feature-extended version of the PT network (PT+Ft.) served as an input for the proposed CC.

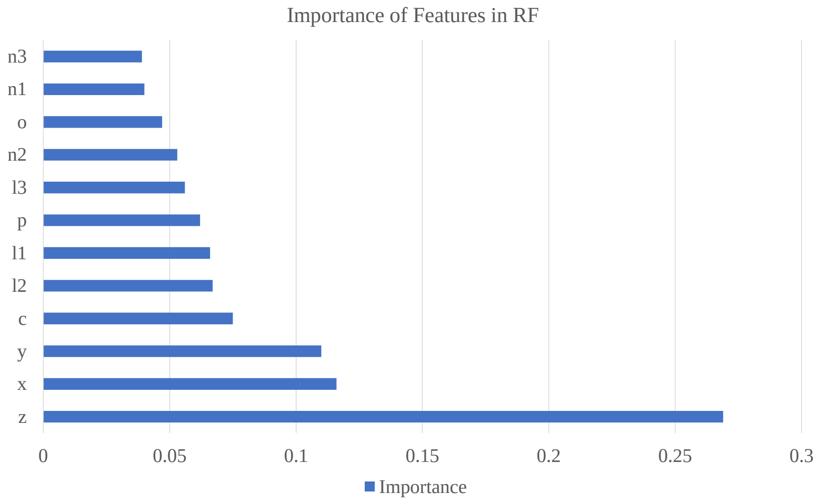


Figure 7.8 Feature importance ranking in random forests (symbols used: *x,y,z* - 3D point coordinates; *o* - *omnivariance*, *c* - *surface variation*, *p* - *planarity*; *l1, l2, l3* - eigenvalues of the PCA components $\lambda_1, \lambda_2, \lambda_3$, respectively; *n1, n2, n3* - normals).

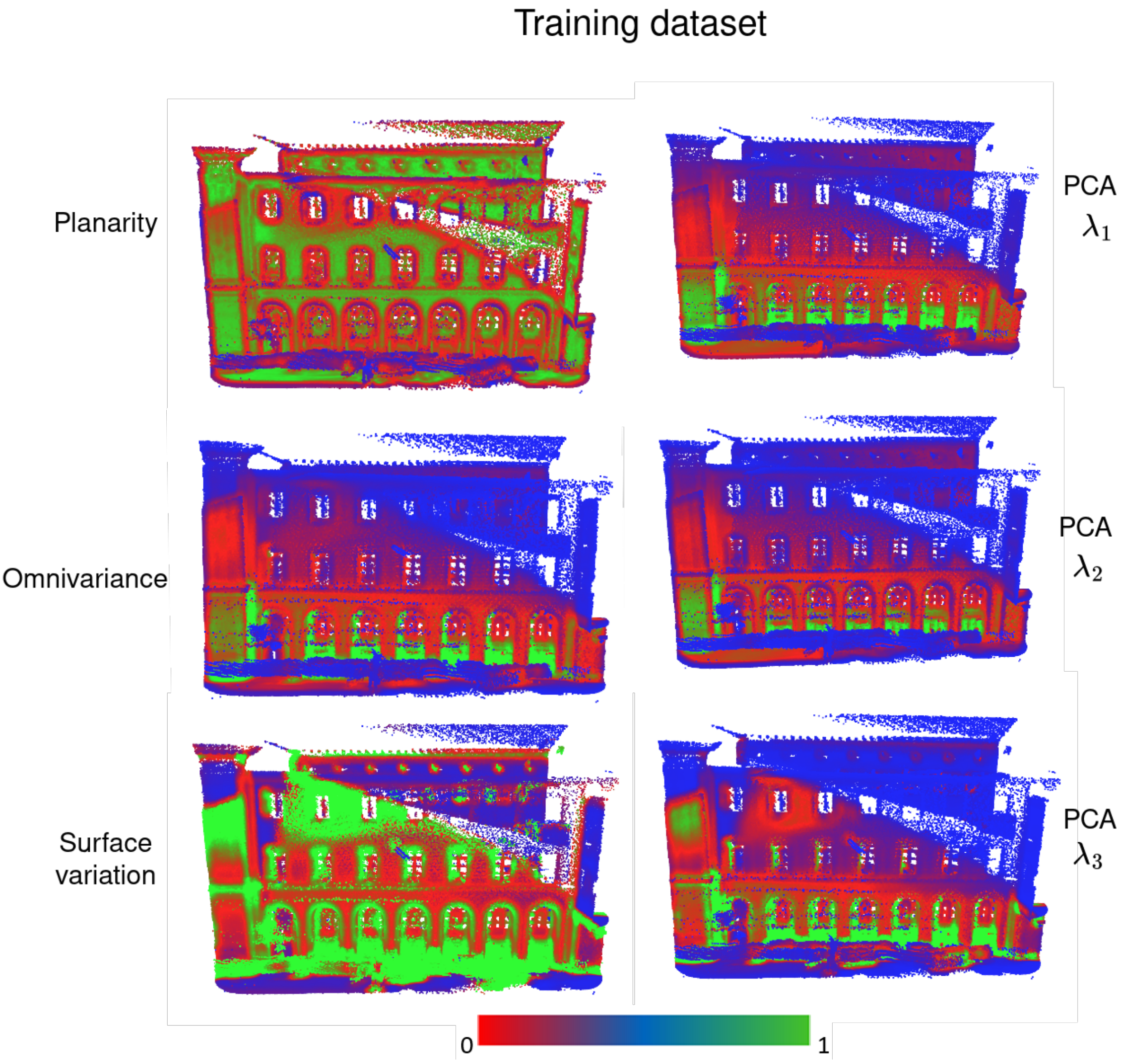


Figure 7.9 Covariance geometric features on an exemplary test set building No 4959322.

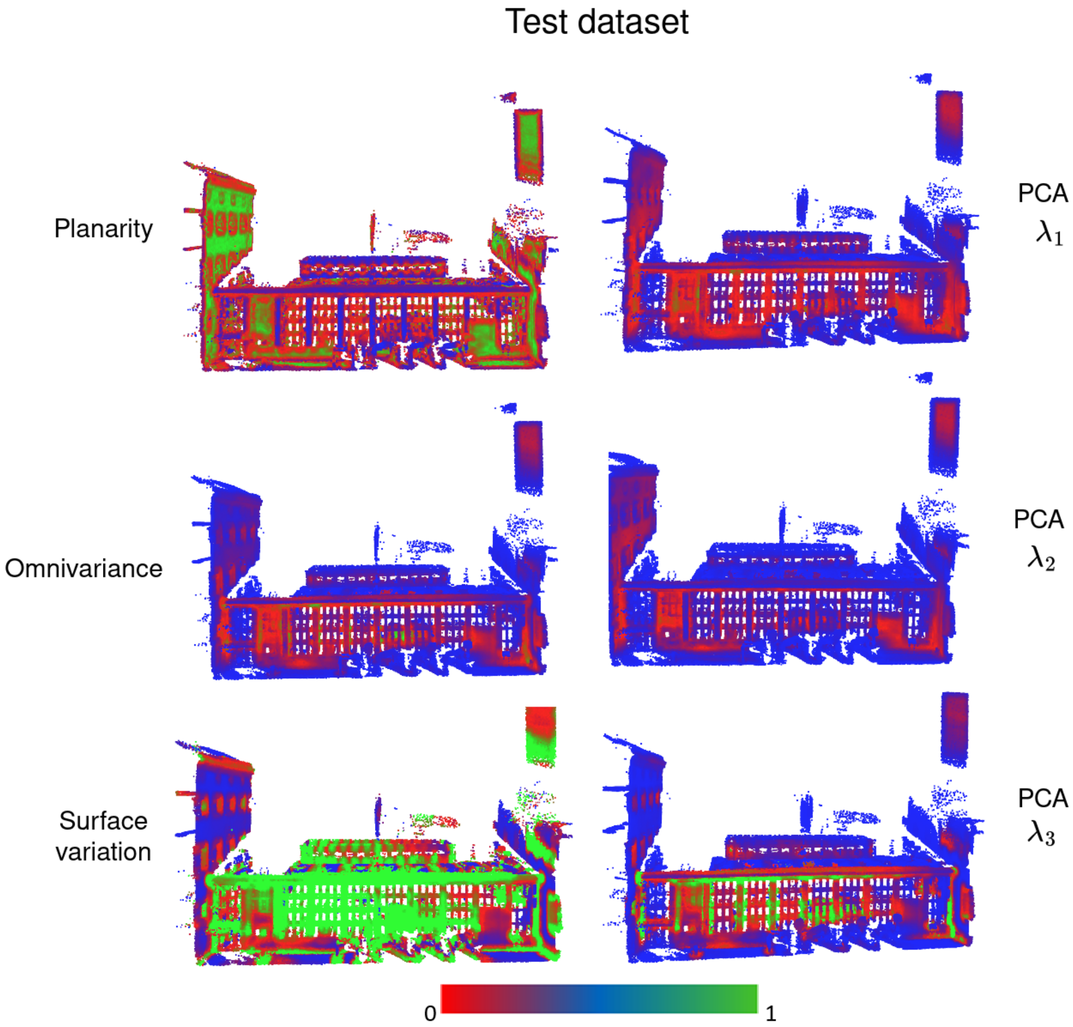


Figure 7.10 Covariance geometric features on an exemplary test set building No 4959459.

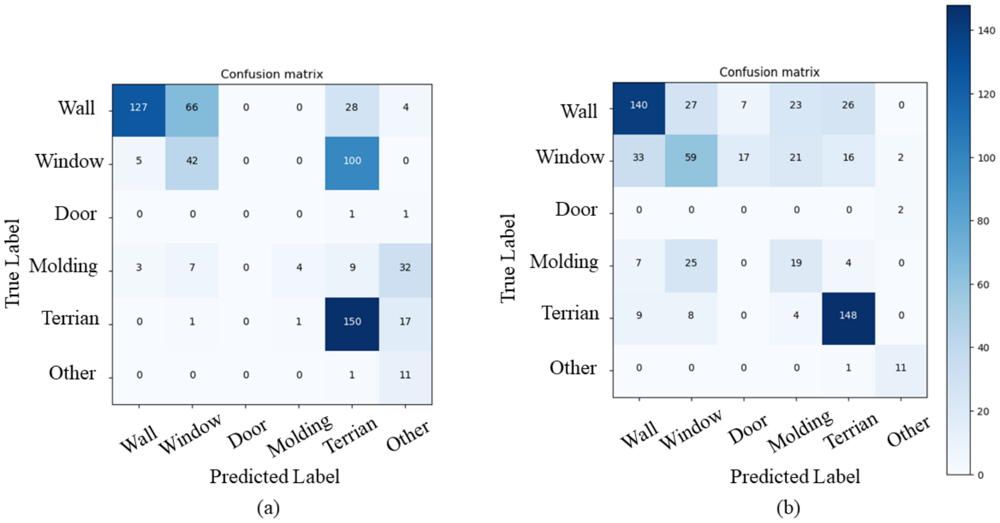


Figure 7.11 PointNet++ classification overall accuracy score on building No 59 with and without geometric features, (a) PointNet++ with xyz only 54.8%, (b) PointNet++ with xyz and 6 geometric features 62.1%.

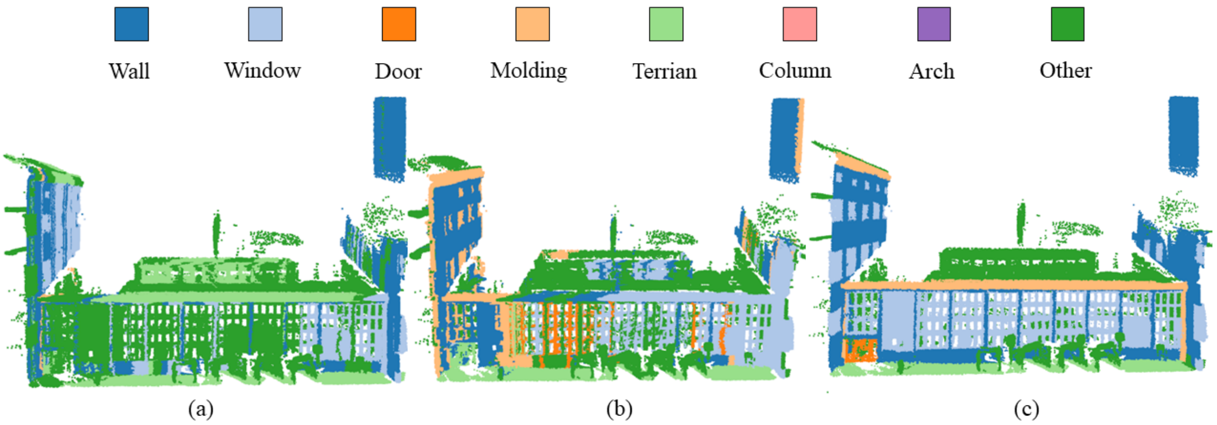


Figure 7.12 PointNet++ classification overall accuracy score on building No 59 with and without geometric features, (a) PointNet++ with xyz only 54.8%, (b) PointNet++ with xyz and 6 geometric features 62.1%, (c) ground-truth.

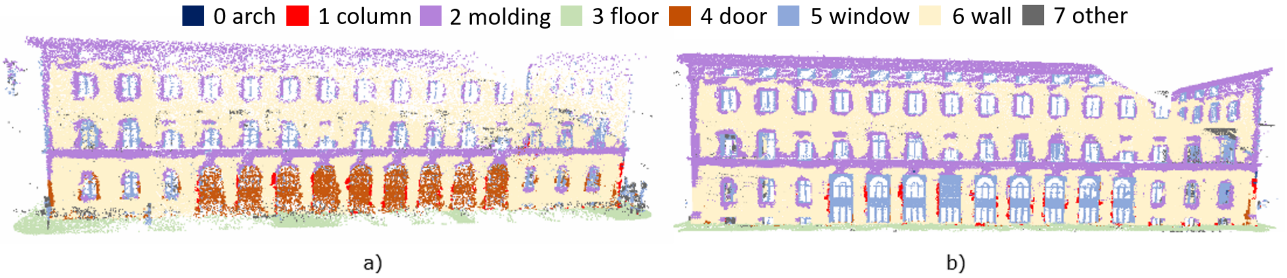


Figure 7.13 3D semantic segmentation performed using a) the baseline PT network, b) the proposed conflict-based method CC.

Table 7.4 Overall accuracy for different combinations of geometric features and algorithms.

Datasets	RF	PointNet	PointNet++
59_XYZ	33.2%	39.6%	54.8%
59_XYZ+9F	49.7%	55.8%	39.2%
59_XYZ+6F	49.1%	52.1%	**62.1%**
23_XYZ	68.4%	69.1%	83.1%
23_XYZ+9F	66.5%	78.5%	84.7%
23_XYZ+6F	64.2%	85.5%	**87.5%**

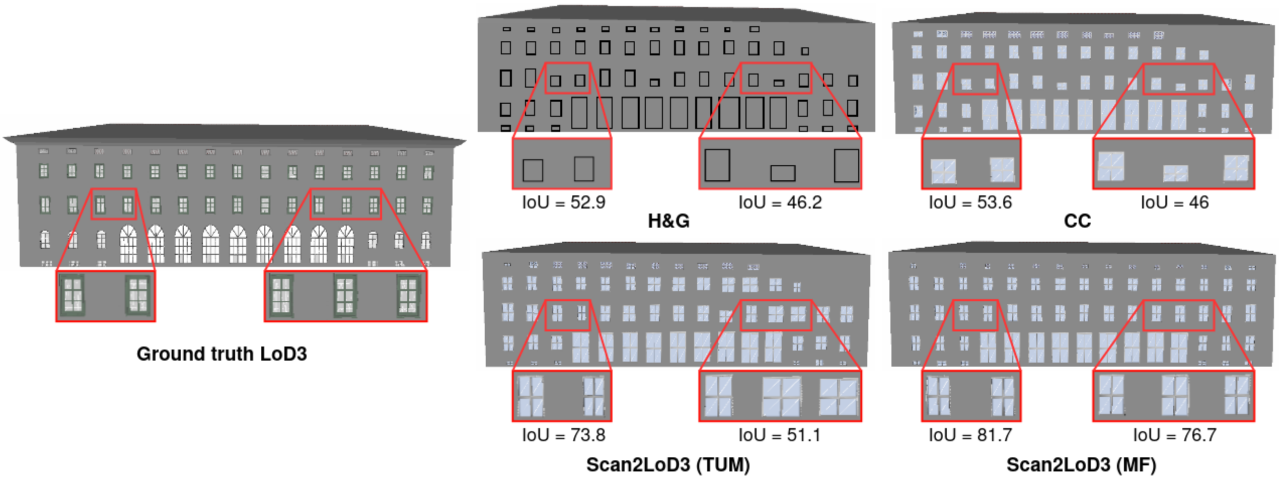


Figure 7.14 Comparison of different segmentation results on the exemplary façade A: The method identifies complete window shapes despite the presence of window blinds (red boxes).

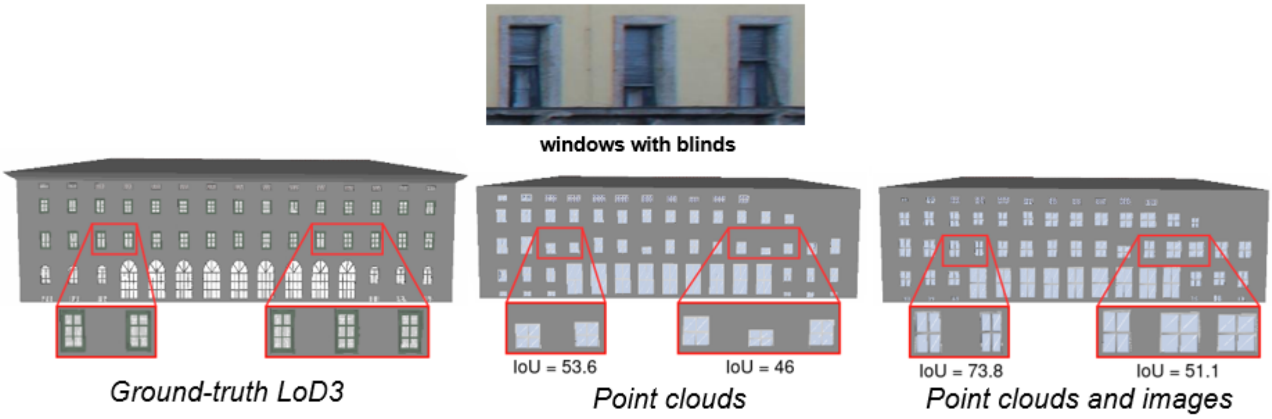


Figure 7.15 Comparison of different segmentation results on the exemplary façade A: Method without (center) and without (right) image texture map.

Method	OA	μP	μR	μF1	μIoU	molding	floor	door	window	wall	other
PT	63.4	58.5	53.2	53.8	41.4	48.2	84.8	1.5	48.7	81.5	58.4
PT+Ft.	72.6	**66.4**	66.7	63.3	52.0	68.3	92.9	5.1	54.6	86.6	72.7
CC	**75.3**	65.9	**71.9**	**65.4**	**52.9**	67.6	86.3	13.6	59.8	85.9	79.1

Table 7.5 Overall accuracy (OA), average recall (μR), average F1 (μF1), average intersection over union (μIoU), and F1 scores per class, given in percents.

	median IoU ↑			
Façade	A	B	C	Total
Openings	66	17	20	103
PT+Ft. [Zhao et al., 2021]	7.3	4.6	3.7	7.3
M-RCNN [He et al., 2017]	63.7	47.4	38.6	58.4
CC (TUM)	66.5	56.4	**53.2**	60.6
Scan2LoD3 (TUM)	63.9	52.9	38	62.1
Scan2LoD3 (MF)	**78.4**	**62.3**	40.6	**76.2**

Table 7.6 Comparison of opening segmentation using only: 3D point clouds (Pt+Ft.), images (M-RCNN), binary masks (CC), and the method with TUM and MF conflict maps.

To further measure the accuracy of the segmentation, the median per-instance intersection over union (IoU) metric was selected for all openings of building 23 (Table 7.6, Figure 7.14). This setup enabled us the comparison to the introduced modified PT network enhanced by the geometric features (PT+Ft.) working only on point clouds; Mask-RCNN (M-RCNN) using only images [He et al., 2017]; method using ray-casting and binary point cloud masks (CC) [Wysocki et al., 2022a]; the method fusing three maps (i.e., conflicts, point clouds, images), once with the TUM-MLS-2016 conflict map (TUM) and on the higher accuracy conflict map of MF (MF).

The experiments corroborate that, in contrast to the tested methods, the proposed solution identifies even closed openings, their full shapes, and reaches higher accuracy (Table 7.6 and Figure 7.14). This fact enabled the whole-shape reconstruction of, for example, windows covered by blinds, which resulted in up to 20% higher IoU on the TUM-MLS-2016 dataset (red boxes, Figure 7.14). Similarly to the detection results, the accuracy of laser measurements significantly influenced the IoU results: The method tended to overestimate opening shapes on the TUM point cloud, whereas on MF, the shapes were approximately 14% more accurate. On the other hand, Scan2LoD3 was sensitive to poor segmentation results (façade C, Table 7.6).

7.3 3D semantic reconstruction

The calculations in this section were performed on the semantic 3D building models of the TUM main campus described in the previous section 6.3; while the point cloud data comprised the TUM-FAÇADE, TUM-MLS-2016, and MoFa 3D point clouds described in the previous section 6.2.

7.3.1 Enrichment accuracy evaluation

The reconstruction performance was tested for reconstructing windows and doors. To this end, the comparison was conducted to measure the accuracy of reconstruction by comparing the method using the TUM-FAÇADE data to the well-established and mesh-oriented Poisson reconstruction [Kazhdan & Hoppe, 2013] and to the CC method (Figure 7.14, Figure 7.16, and Table 7.7). The MF point clouds' results were also added to highlight the influence of point cloud accuracy. As shown in Table 7.7 and in Figure 7.16, the 3D building priors provided more accurate reconstruction results than the standard Poisson reconstruction (i.e., root mean square (RMS) lower by 52%); the former also achieved the watertightness. Among the prior-driven methods, the improvement related to higher detection rate and IoU was notice-

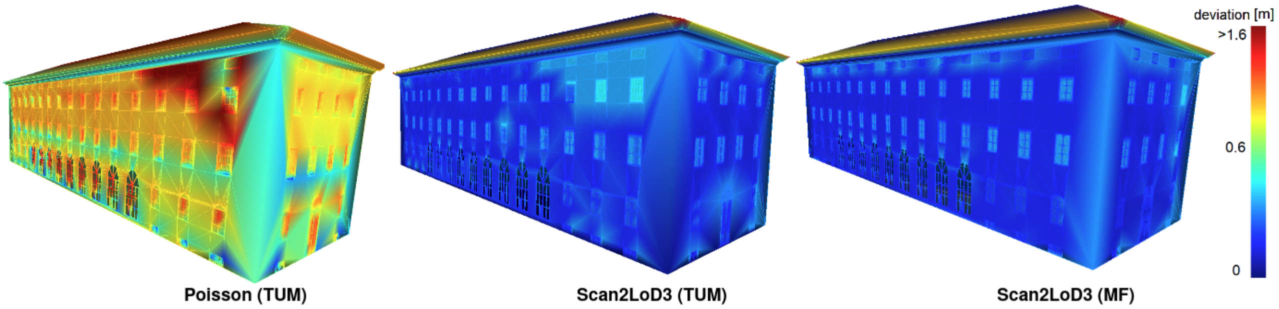


Figure 7.16 Comparison of the Poisson to the reconstruction approach: Deviations are projected onto the ground-truth LoD3 model.

Method	vs. GT LoD3 ↓		
	μ	RMS	WT
Poisson (TUM) [Kazhdan & Hoppe, 2013]	0.35	0.54	✗
CC (TUM)	0.31	0.34	✓
Scan2LoD3 (TUM)	0.23	0.26	✓
Scan2LoD3 (MF)	**0.13**	**0.25**	✓

Table 7.7 Comparison of mesh-based Poisson, building-prior-driven CC, and the proposed method using the ground-truth LoD3 model and measuring watertightness (WT).

able: Scan2LoD3 had lower mean and RMS scores by up to 26% and 24%, respectively, compared to CC (Table 7.7). It is worth noting that the eaves were incorrectly reconstructed in any of the presented methods.

The validation was conducted using the ground-truth point clouds, which represented the underpasses, while volume and surface changes were compared using refined building models (LoD3) and raw building models (LoD2). The results of the validation are shown in Table 7.8 and Figure 7.17, while Table 7.11 and Table 7.12 present surface and volume changes, respectively. The most remarkable result to emerge from the experiments was that deviations in automatically modeled underpasses ranged from 1 cm to 23 cm and had a mean mode score of 12 cm, as shown in Table 7.8.

Table 7.8 Validation of automatically modeled underpasses compared to the ground-truth point clouds.

Building	$Median$ [m]	$Mode$ [m]	RMS [m]
A	0.18	0.01	0.29
B	0.19	0.07	0.38
C	0.25	0.06	0.52
D	0.28	0.23	0.45
μ	**0.22**	**0.12**	**0.41**

7.3.2 Library object matching evaluation

A set of histogram distances is used to assess the similarity of the histograms within the presented BoW approach with the Chi-square-distance (subsection 5.2.2). Many different histogram distances are available [Cha, 2008], whereby the BoW-established histogram distances are selected for this evaluation. The Minkowski distance is given as [Cha, 2008]:

$$D = \left(\sum_{i=1}^{n} |q_i - p_i|^p \right)^{\frac{1}{p}} \tag{7.1}$$

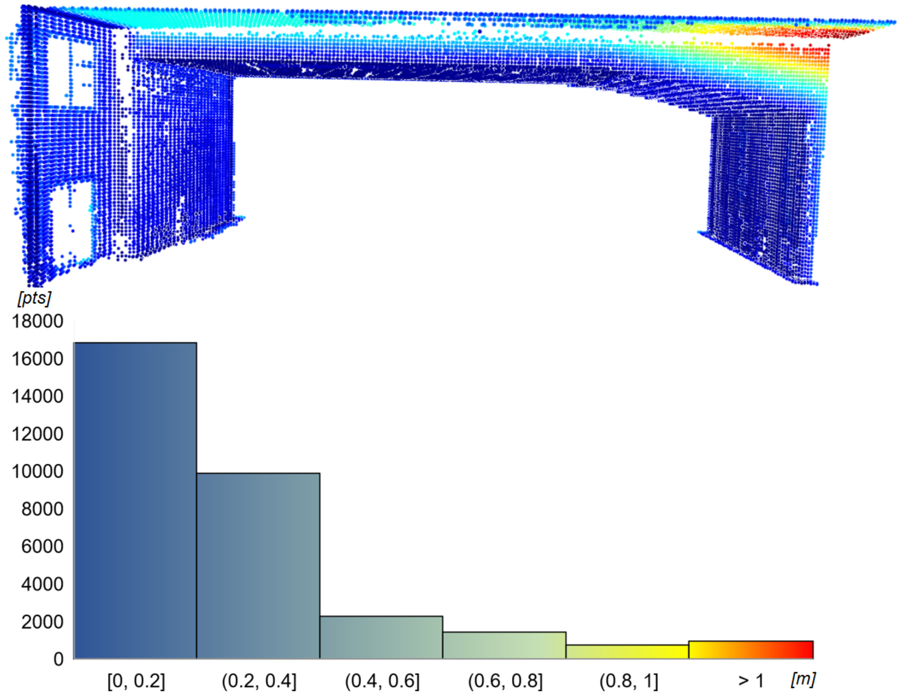


Figure 7.17 Histogram and visualization of projected Hausdorff distances to a ground-truth point cloud for underpass B, given in meters

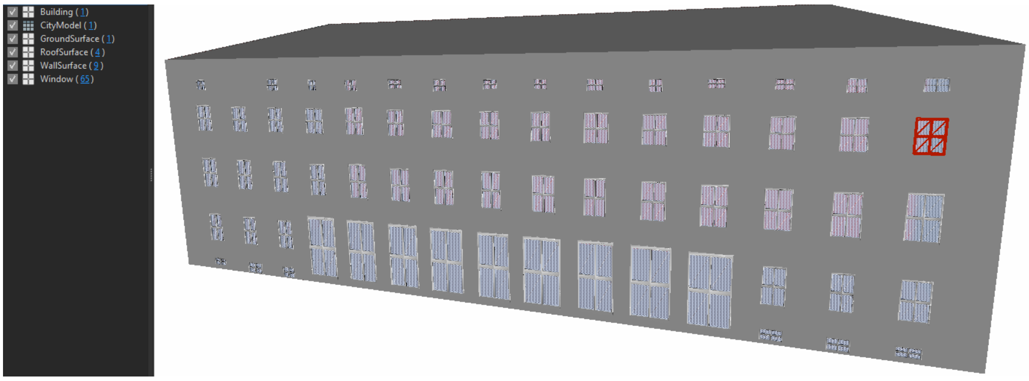


Figure 7.18 Enrichment result on an example of windows embedded into the CityGML 2.0 model (left table) with the highlighted one instance (red).

The Minkowski distance is characterized by the parameter p. It can be interpreted as a more general form of the Manhattan distance $(p = 1)$, the Euclidean distance $(p = 2)$ and the Chebyshev distance $(p = \infty)$ [Cha, 2008]. It is used to evaluate the similarity of two histograms by pairwise comparison of the individual bins and subsequent accumulation of the obtained distances. The Jensen-Shannon divergence is given as [Bamler & Shi, 2022]:

$$\text{JSD}(P, Q) = \frac{1}{2}\left(\text{KL}(P\|M) + \text{KL}(Q\|M)\right) \tag{7.2}$$

where KL$(P\|Q)$ is the Kullback-Leibler-Divergence between the probability distributions P and Q. The Kullback-Leibler-Divergence given as [Bamler & Shi, 2022]:

$$\text{KL}(P\|Q) = \sum_i P(i) \ln\left(\frac{P(i)}{Q(i)}\right) \tag{7.3}$$

Like the Kullback-Leibler divergence, the Jensen-Shannon divergence is based on the concept of entropy [Cha, 2008] and is used to quantify the similarity of two probability distributions. The Jensen-Shannon

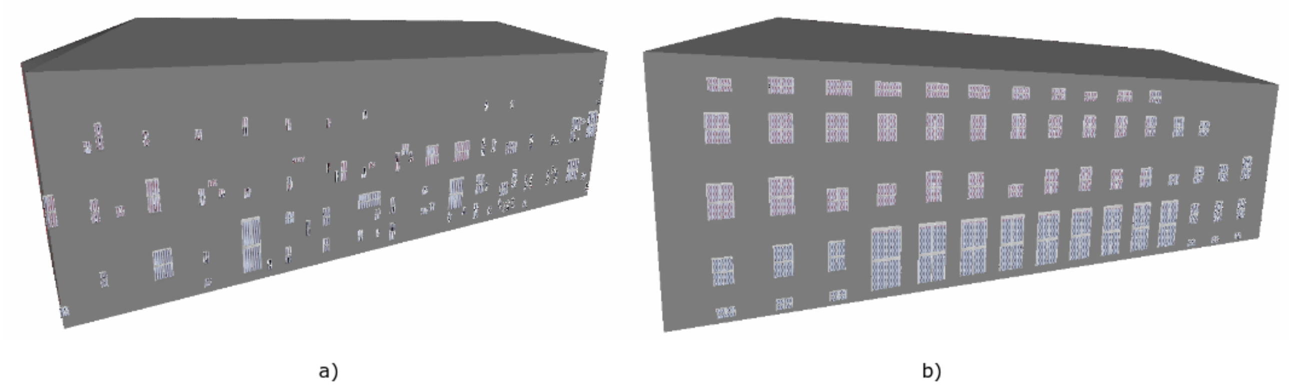


Figure 7.19 Influence of the segmentation accuracy based only on point clouds with: a) the baseline PT network, b) the proposed conflict-based reconstruction method CC.

divergence can be interpreted as a symmetric version of the Kullback-Leibler divergence [Cha, 2008]. The method was inferred on the models superimposed with random noise and on the $building\ 81$ from the labeled MLS point cloud of the TUM-FAÇADE data set [Wysocki et al., 2022c]. This object was selected owing to its window-type variability. To ensure the comparability of the experiments, the same number of feature cluster bins n was applied in all of them. The suitable n was determined in a heuristic way by performing experimental clustering and feature descriptor quantization for different values of n. Then, the occurrence frequency of the respective cluster centers was evaluated. Based on this analysis, $n = 25$ was established as an optimal number of clusters. When sampling the CAD models to a point cloud, transparency properties have to be taken into account. As shown in Figure 7.20, in the pre-processing, the glass was removed from all models to exploit the presence or absence of window bars. In Figure 7.20 b) detailed window bars become apparent due to glass components not being sampled. In contrast, Figure 7.20 a) illustrates that when the glass components are sampled using the same method as the rest of the window, the window appears as an opaque object. With MLS scans, the laser beams usually penetrate the glass and thus penetrate the interior of the building, justifying the removal of glass from the CAD models.

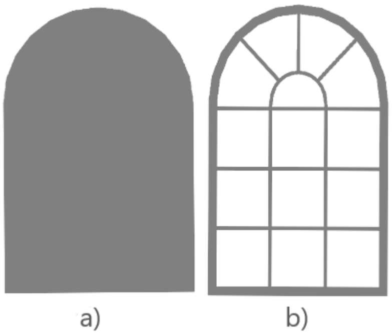


Figure 7.20 Simplified point cloud to 2D representing CAD window model: a) without removal of window glass b) with removal of window glass.

Experiments on models superimposed with random noise and real-world point clouds

For the experiments, random noise is added to the point clouds that are sampled from the pre-processed CAD models, as illustrated in Figure 7.21. The Chi-Square histogram distance is employed for these experiments. Results from these experiments are summarized in Figure 7.22, Figure 7.23 and Table 7.9. Incorporating HOG descriptors improves the matching quality. Table 7.9 shows that the overall accuracy is improved from 0.47 to 0.69, while the kappa-coefficient is improved from 0.36 to 0.65. However, the approach is sensitive to noise:

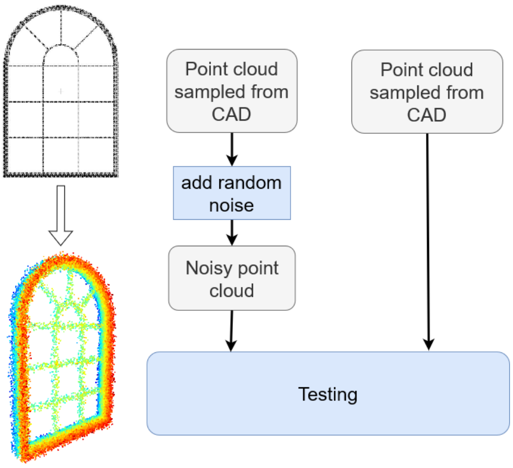


Figure 7.21 Workflow representing the addition of random noise to the point clouds sampled from CAD model.

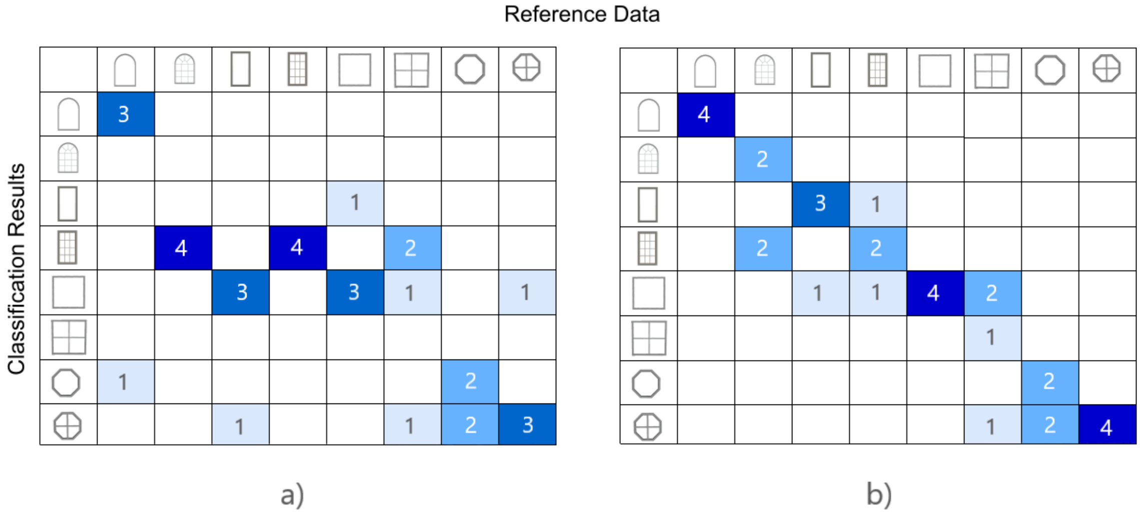


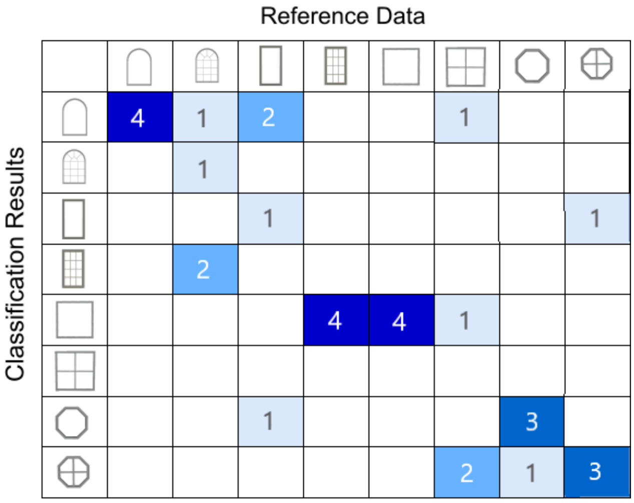
Figure 7.22 Confusion matrices: a) ORB, b) ORB and HOG.

By doubling the noise level, the kappa coefficient is diminished to 0.43, while the overall accuracy drops to 0.5. Also, there is an observable dependency of the matching quality on the type of CAD model. For example, Figure 7.22 and Figure 7.23 show that the matching is more stable for the arched and octagon-shaped-windows than for the rectangular and quadratic-windows.

Further experiments were conducted with different feature combinations and histogram distances. Figure 7.24 summarizes the standard deviations and the variances of the user's and producer's accuracy for six of these experiments with different sets of hyper-parameters. The arched window with no bars and the two octagon-shaped windows matched the most robustly compared to the other window types. For this test, the TUM-FAÇADE dataset was employed. The tested façade only comprised rectangular and arched windows with window bars (Figure 7.26). An observable improvement concerning the matching quality with the incorporation of HOG-descriptors, as the overall accuracy increases to 0.57. Also, the dependence on

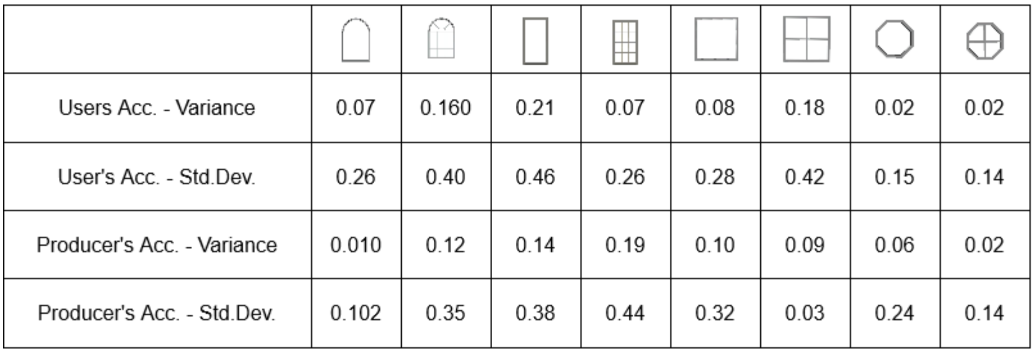
Table 7.9 Results of the experiments with the models superimposed with random noise.

Experiment	Overall Accuracy	Kappa Coefficient
ORB, Chi-Square dist.	0.47	0.36
ORB and HOG, Chi-Square dist.	0.69	0.65
ORB and HOG, Chi-Square dist., noise $*2$	0.50	0.43


Reference Data
Classification Results

	4	1	2			1		
		1						
			1					1
		2						
				4	4	1		
			1				3	
						2	1	3

Figure 7.23 Confusion matrix: ORB and HOG, noise level with factor 2.



Users Acc. - Variance	0.07	0.160	0.21	0.07	0.08	0.18	0.02	0.02
User's Acc. - Std.Dev.	0.26	0.40	0.46	0.26	0.28	0.42	0.15	0.14
Producer's Acc. - Variance	0.010	0.12	0.14	0.19	0.10	0.09	0.06	0.02
Producer's Acc. - Std.Dev.	0.102	0.35	0.38	0.44	0.32	0.03	0.24	0.14

Figure 7.24 Standard deviation and variance of user's and producer's accuracies from experiments with six different combinations of hyperparameters.

the histogram distance is shown in Table Table 7.10. A correlation of the matching quality with the building height in Figure 7.26 is observed, which presumably can be attributed to the point density of the MLS point cloud, decreasing with increasing altitude.

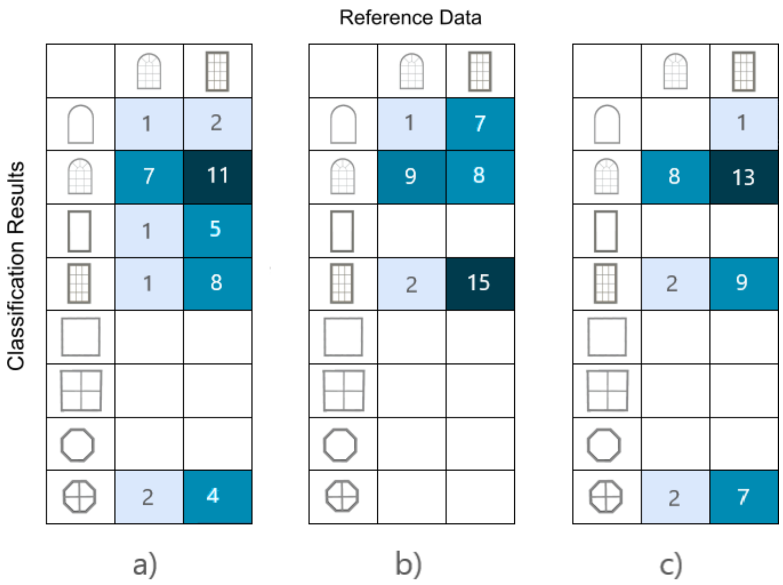
The experiment with ORB, HOG, and the Jensen-Shannon Divergence was also conducted using dense feature sampling instead of extracting key points. Figure 7.27 summarizes the results from this experiment. However, with an overall accuracy of 0.45, there is a decrease in the accuracy.

7.3.3 Enrichment impact on the building model accuracy

As hypothesized, the experiments showed that underpasses contribute significantly to a building's envelope surface; the mean score for the surface difference was 13%, reaching up to 20%. Note that the score for underpass B was 4% since the building model is vast and thus has a large surface area. All surface-related comparisons are shown in Table 7.11. The mean score for volume deduction was 11%, reaching up to 18%, as shown in Table 7.12.

Table 7.10 Results of the experiments on the TUM-FAÇADE dataset.

Experiment	Overall Accuracy
ORB, Jensen-Shannon Divergence	0.36
ORB and HOG, Jensen-Shannon Divergence	0.57
ORB and HOG, Minkowski-Distance	0.41


Reference Data
Classification Results
a)
b)
c)

	1	2
	7	11
	1	5
	1	8
	2	4

	1	7
	9	8
	2	15

		1
	8	13
	2	9
	2	7

Figure 7.25 a) ORB, Jensen-Shannon Divergence, b) ORB and HOG, Jensen-Shannon Divergence, c) ORB and HOG, Minkowski-Distance.

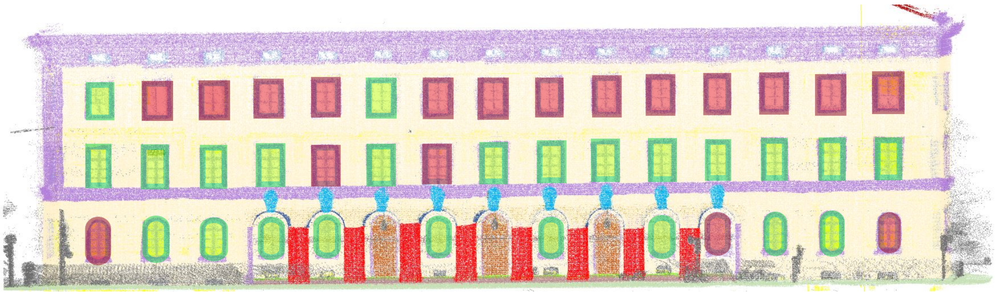
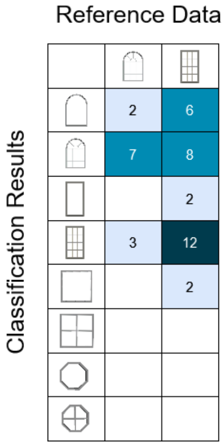


Figure 7.26 Correctly (green) and falsely (red) matched windows.


Reference Data
Classification Results

	2	6
	7	8
		2
	3	12
		2

Figure 7.27 Results of the experiments on the TUM-FAÇADE dataset with dense feature sampling.

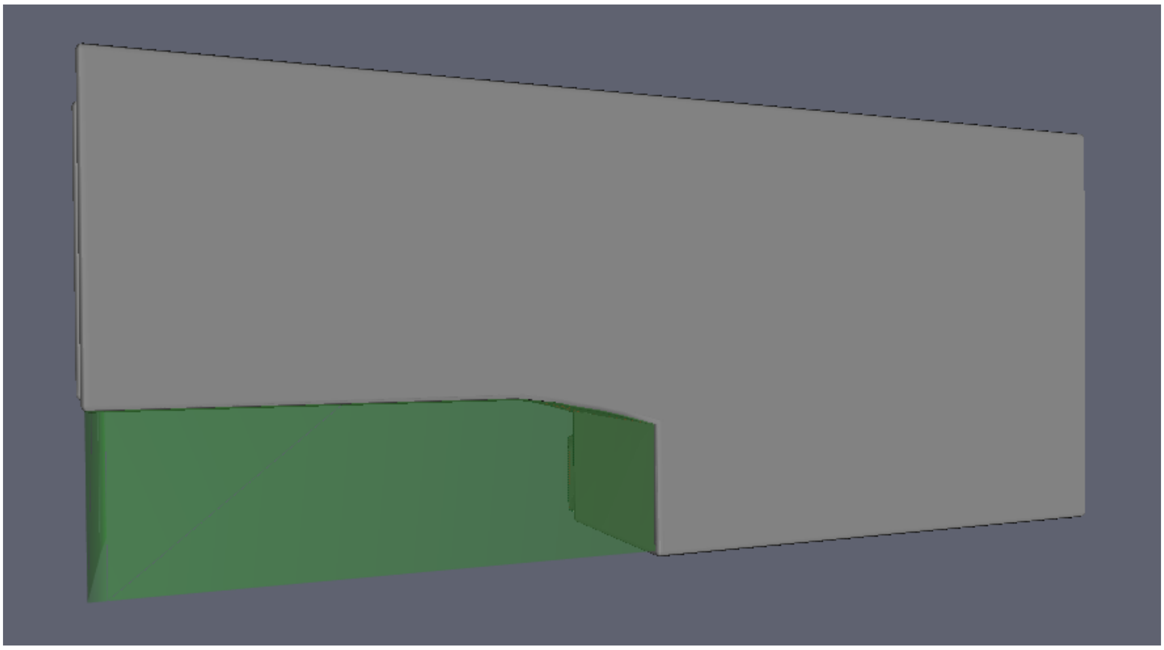


Figure 7.28 Exemplary volume gain (green) of LoD2 when compared to the LoD3 model enriched in an underpass: Difference of 1311 [m^3] was observed for LoD3 5123 [m^3] and LoD2 6434 [m^3].

Table 7.11 Surface refinements for refined models (LoD3) compared to raw (LoD2) building walls and ground surfaces.

Building	Surf. LoD2 [m^2]	Surf. LoD3 [m^2]	Diff. [m^2]	Diff. [%]
A	2,724	2,448	276	10
B	10,132	9,716	416	4
C	1,918	1,534	384	20
D	1,058	874	184	17
μ			**315**	**13**

Table 7.12 Volume refinements for refined models (LoD3) compared to raw building models (LoD2).

Building	Vol. LoD2 [m^3]	Vol. LoD3 [m^3]	Diff. [m^3]	Diff. [%]
A	11,994	11,116	878	7
B	70,110	68,962	1,148	2
C	6,434	5,403	1,031	16
D	2,898	2,380	518	18
μ			**894**	**11**

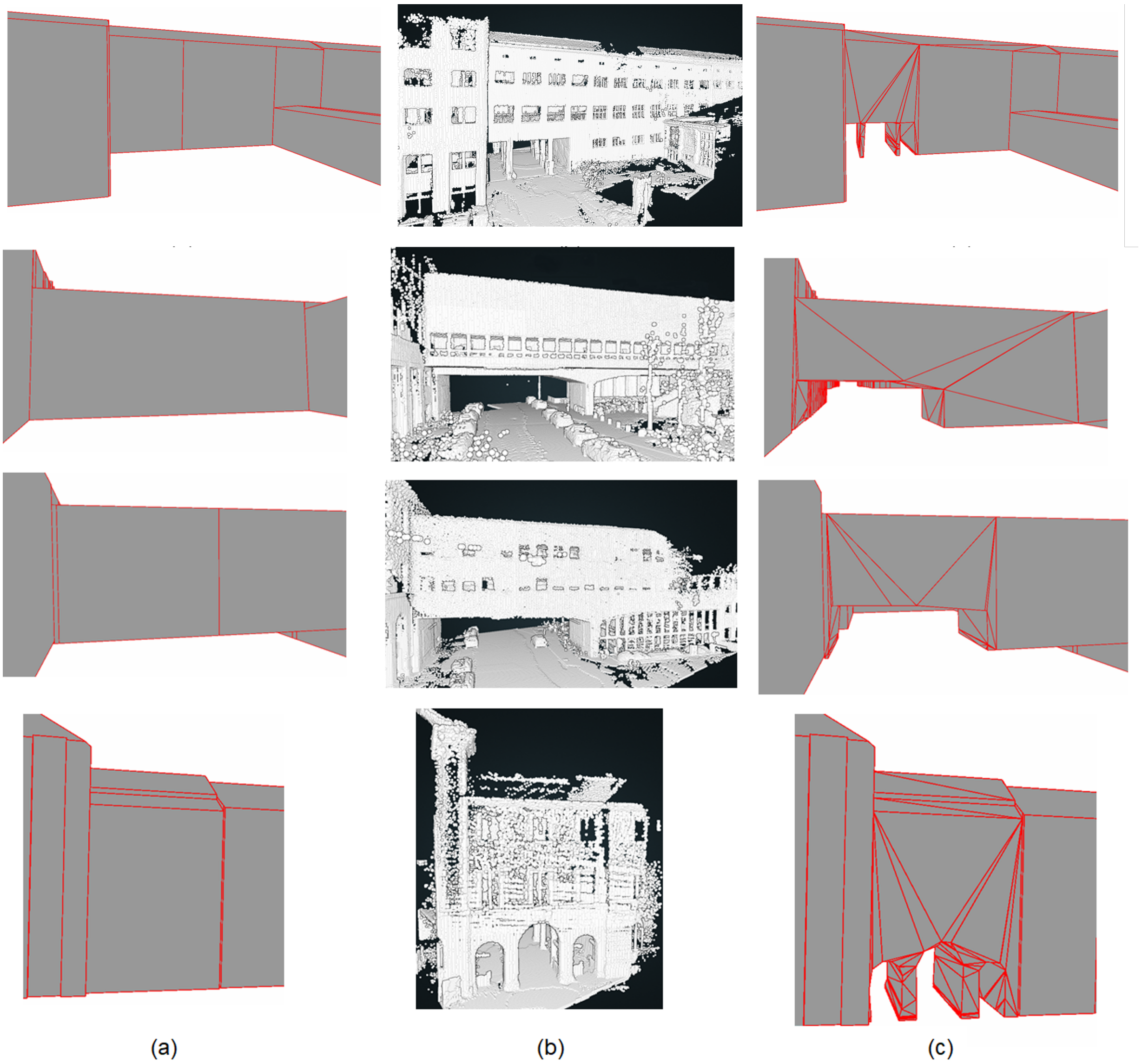
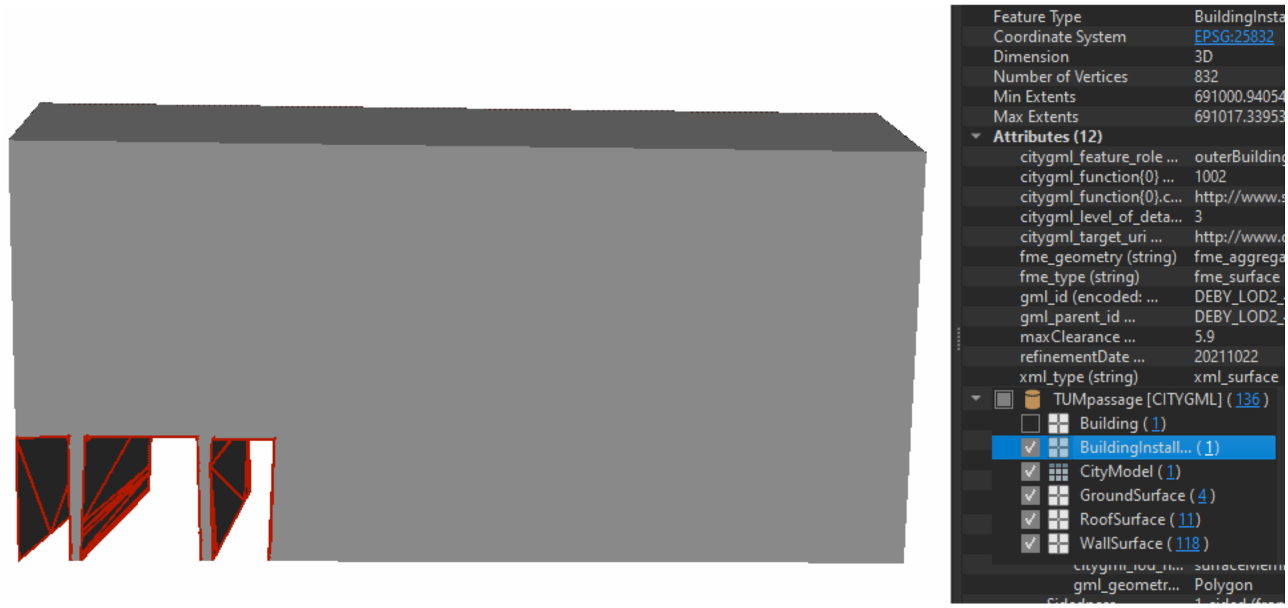


Figure 7.29 Refinement results for the buildings A, B, C, and D. a) Raw city model, b) point cloud, c) refined city model.



Feature Type	BuildingInsta
Coordinate System	EPSG:25832
Dimension	3D
Number of Vertices	832
Min Extents	691000.94054
Max Extents	691017.33953
Attributes (12)	
citygml_feature_role ...	outerBuilding
citygml_function{0} ...	1002
citygml_function{0}.c...	http://www.s
citygml_level_of_deta...	3
citygml_target_uri ...	http://www.c
fme_geometry (string)	fme_aggrega
fme_type (string)	fme_surface
gml_id (encoded: ...	DEBY_LOD2_
gml_parent_id ...	DEBY_LOD2_
maxClearance ...	5.9
refinementDate ...	20211022
xml_type (string)	xml_surface

gml_geometr...	Polygon

Figure 7.30 Refinement result on an example of an underpass embedded into the CityGML 2.0 model (right table).

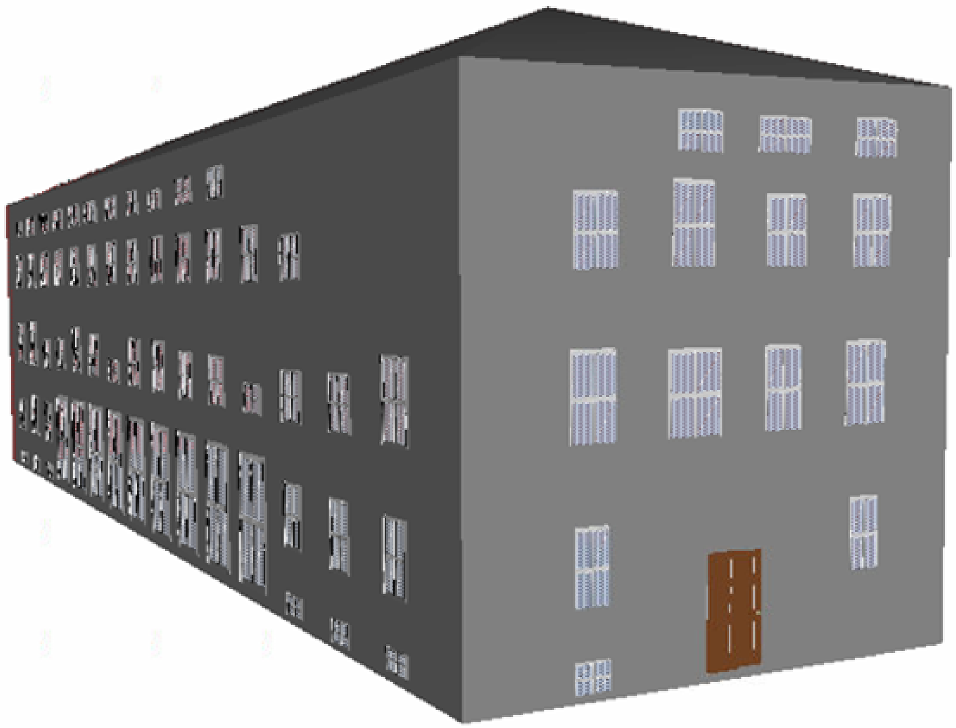


Figure 7.31 Enrichment result on an example of windows and doors in the presence of incomplete measurements: Elements are not reconstructed when the segmentation yields no results (top corner of the building model).

8 Discussion

8.1 3D model and point cloud fusion

One of the pivotal downsides of the established top-view looking acquisition is that, e.g., it prevents capturing building façades and thus limits the achievable LoD of the reconstructed object. The recent interest in detailed road space modeling is driven by several factors, where the development of automated driving functions is a pivotal one. This trend is reflected in an increased number of mobile mapping units scanning road environments. This, however, results in an influx of geodata such as MLS point clouds and HD Maps that depict the road network and its space supporting the navigation and simulation of automated vehicles. Nevertheless, HD Maps may be valid for several test categories of automated driving functions, but as soon as more complex physical sensor effects are demanded for testing, they are not sufficient anymore [Schwab & Kolbe, 2019]. For that purpose, more detailed geometrical and semantical representations of real environments are needed. Moreover, the geodata flood is strengthened by the growth of connected devices equipped with LiDARs, cameras, and RGB-D sensors. Consequently, the question arises of how preexisting models can be geometrically matched with the unstructured influx of other geodata.

In the case of the presented methods, confidence intervals are used in conjunction with BayNet to establish a framework for fusion. The BayNet enables seamlessly rebuilding the structure, adding new components, and obtaining results at the intermediate steps as well. The threshold probabilities in Figure 3.4 are generic since they should depend on an end application. The fusion itself aims to couple confirming object parts. As such, only the output of *Raw geometry confirmed* in Figure 3.4 should be used for such purposes. Also, P_{high} and $P_{moderate}$ thresholds may be used for façade reconstruction purposes, whereas this might not be the case for change detection tasks. For example, the fusion process assumes that 3D building models exist and seeks confirmation of the structure in MLS measurements: For outdated city models, a further application-specific module has to be designed. Primarily, P_{low} should reject areas of less priority. As such, other city objects (e.g., road) and building parts unconfirmed by MLS observations are considered of low importance. The rationale for this is that occlusions are blocking spatial information acquisition. This, however, limits 3D processing capabilities within these areas and, as such, should be rejected.

Moreover, the presented dissertation introduces CIs discretized to 3D boxes. As shown in Figure 3.5, these might be used directly as soft evidence or further discretized to 2D patches, voxels, or octrees. This process, however, shall avoid element-wise discretization. Also, point clouds are often not parallel to walls (as for simplicity shown in Figure 3.5). In such cases, the additional intersection conflicts have to be solved. To avoid this, a scene-wise partition is designed. Furthermore, the discretization method should be application-specific. The same applies to a discretization unit size. For example, the detection of discrepancies due to wall openings will not yield plausible results when using 2D patches of 5.0 [m] x 5.0 [m] size, which should not be the case for 3D voxels of 0.01 [m] x 0.01 [m].

The fusion can also be performed on a point-to-point basis, as shown in the joint publication: Schwarz et al. [2023]. The assessed building range and assumptions derived based on a number of present wall surfaces support the transfer of semantic labels joint with the proposed plane-to-plane registration. The reported accuracy is up to approx. 82%.

Such fusion approaches can also be used for direct mesh-based reconstruction and utilization in creating an interactive game or in 3D GIS solutions; the visualization is shown in Figure 8.1. This confirms that semantic models can be used in the Unreal Engine software, which is used as an engine within tools such as CARLA that serves the purposes of automated driving research. Besides, the models can be utilized in 3D GIS solutions like the 3DCityDB-Web-Map-Client, as shown in Figure 8.1, and serve the purposes

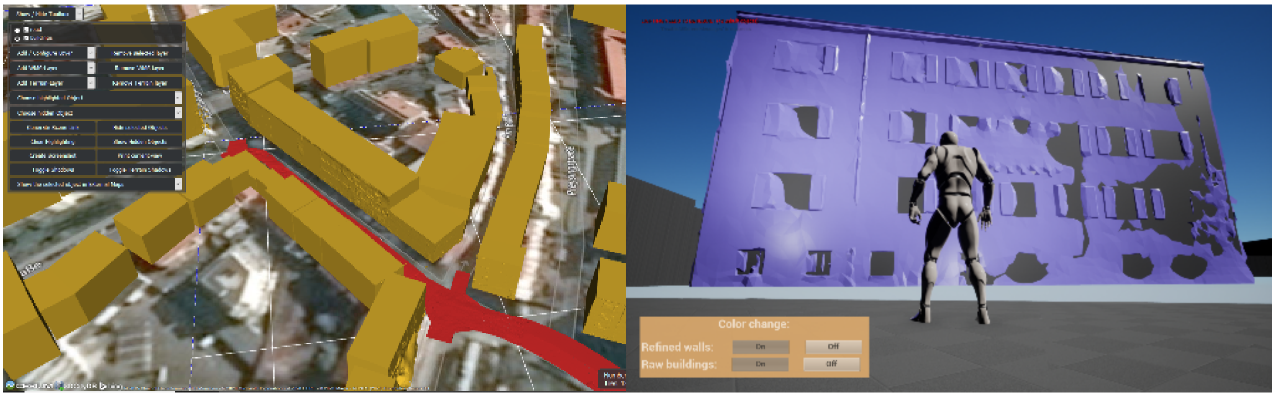


Figure 8.1 Refined models used in city models management tool (left) and automated driving simulator engine (right)

of a 3D or 4D cadastre [Zlatanova, 2000], as the concept also includes the time factor. The pipeline is expected to generate comparable results for mid-sized cities in Europe, but the transferability should be further examined for more architectural styles, such as skyscraper environments of megacities.

8.2 3D semantic segmentation of façade elements

Throughout the years, the photogrammetry community witnessed the emerging methods for LoD1 and LoD2 building model reconstruction. Nowadays, such 3D models are ubiquitous and perceived as a core element of an urban digital twin. In this work, a strategy is presented that leverages their traits to support the detection and segmentation of façade elements contributing to the LoD3. Such strategy is an intuitive and consequent next step of semantic segmentation enabling LoD3 reconstruction.

As exemplified by the experiments, such an approach offers high robustness, reaching even 90% detection rate for the method analyzing point clouds and 3D models and reaching up to 98% when combined with images for the observable openings. On the other hand, the essential requirement for conflict-supported segmentation is existing 3D building models. Such models are widely available in many cities and countries around the globe [Wysocki et al., 2022d]. In case they are absent, there are proprietary solutions and open-source solutions [Huang et al., 2022; Haala & Kada, 2010] facilitating low LoD reconstruction, which, however, require co-registered ALS point clouds and building footprints.

The experiments corroborate that DR is also dependent on the density of measurements per façade. For example, for the densely covered façade A it estimated 90% DR-AO and detected 100% of measured openings (DR-MO); for the highly occluded side-façade C it estimated 28% and 50%, respectively (see Table 7.3). The robustness of the method is underlined by the very low false alarm rate: Roughly 1% false alarm rate for both measured (FR-MO) and all openings (FR-AO) was noted.

The proposed CC method yields limited results when openings are uncovered by point clouds (*unknown* regions) or partially measured (e.g., blinds before windows), as exemplified by Figure 8.2. The issue of half-closed openings is mitigated by adding images and relying on probability range instead of binary conflict or confirmation decision, as shown in Figure 7.14. Note that it also provides a slight improvement in method robustness, yet it does not mitigate the influence of other objects obstructing the façade view. As such, the segmentation is expected to perform well on objects that are measured but not partially measured.

Analogically to the building reconstruction methods, the point cloud semantic segmentation methods are investigated. The years of machine learning methods development resulted in robust solutions, such as random forest. However, their performance strongly relies on hand-crafted features, and as such, their generalization capabilities are limited. The advent of deep learning methods promises to mitigate these issues. This strategy's premise is that there are multiple datasets available, exposing the networks to various examples, ensuring high perturbations, and, as such, learning the features from data. If such premise is considered fulfilled for image-related use cases, such premise is not granted for sparsely avail-

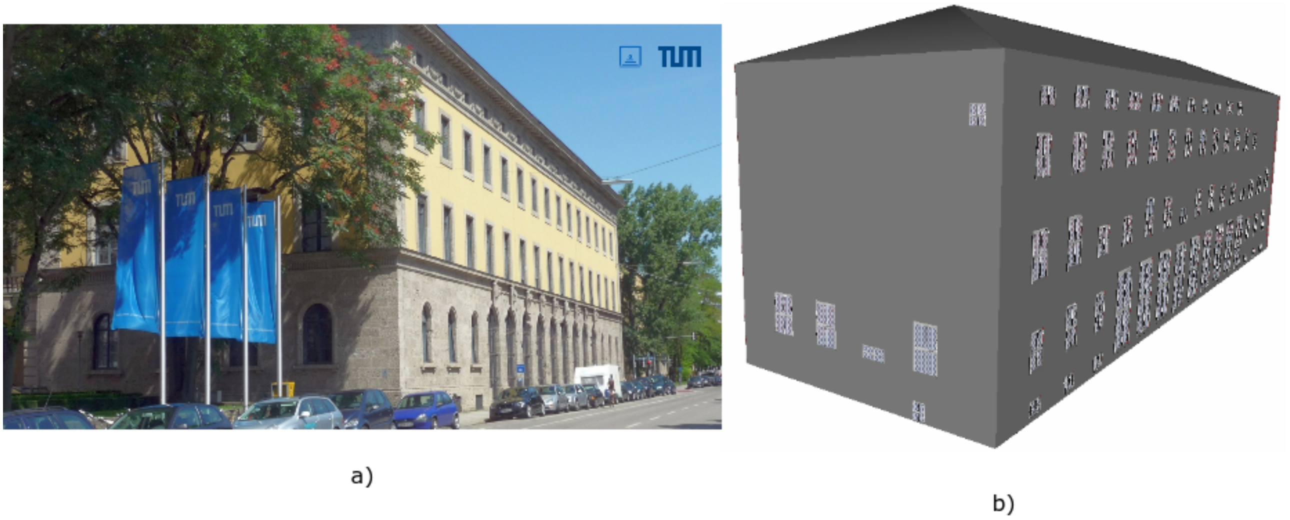


Figure 8.2 Openings reconstruction for two selected façades in the presence of occluding objects, such as vegetation (trees), street furniture (flags), and impenetrable windows (blinds): a) Photo, b) refined façade

able point cloud data, let alone façade-level point clouds. Therefore, a marriage of the learned features with the hand-crafted features is proposed, which is realized by early-fusing the hand-crafted features into the deep neural networks.

As shown in the experiments, such fusion provides a significant performance boost in case of unbalanced and sparse data. For the established point-wise networks, PointNet and PointNet++, the accuracy increase of 5% and 10% is observable, respectively. As such, it can be used for façade-related segmentation, where data is sparse. The disadvantage of the method is the reliance on the selected features, and as shown in the experiments, the usual rule of the more, the better, proves invalid in this case. The pre-processing, calculation time and selection of an appropriate radius also add to the disadvantages of the method.

When compared to the ground-truth openings, the proposed segmentation reached roughly 61% accuracy (Table 7.6); yet the method is limited when windows are partially measured (e.g., blinds before windows), as exemplified by several windows in the third-row in Figure 7.14b.

The back-projected, classified conflicts increased the accuracy of semantic point cloud segmentation by approximately 12% (Table 7.5). Note that the precision and intersection over the union score for CC remained similar to the PT+Ft. score, while F1 score for *floor* dropped by about 6%. Remarkably, the proposed CC method improves segmentation of *window*, *door*, and *other* classes by approximately 11%, 12%, and 21%, respectively.

8.3 Enrichment of building models using façade elements

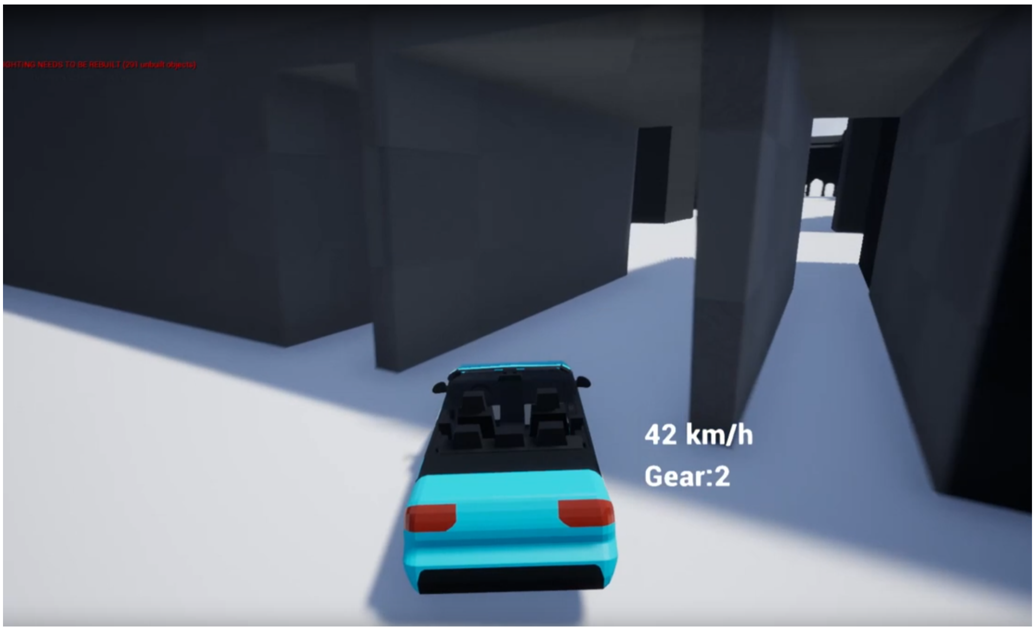
8.3.1 3D semantic reconstruction

The backbone of the presented refinement strategy is the analysis of laser rays in combination with 3D models. This is an intermediate step allowing the creation of conflict probability maps as textures on 3D model surfaces but preceding the final semantic 3D reconstruction. From the perspective of downstream applications, conflict maps already yield important information. For example, the ratio of openings to non-openings is essential to heating demand estimation, even without specifying the opening semantics [Apostolopoulou et al., 2023]. Nevertheless, the benefits of explicit geometry modeling are much greater than solely placing textures on surfaces, as highlighted by the analyzed applications and benefits of the explicit semantic LoD3 model (Figure 2.4).

The presented strategy relies on geometric priors of low LoD building models. The strategy reconstructs objects only in the identified conflict-related places: It allows assuming watertightness of given prior and close shapes in unmeasured space. Such space is frequently observable in the case of the street-level MLS, as the mobile mapping platforms rarely enter the inner-yards of buildings and partially capture roof surface, rendering in most cases incomplete coverage of building structure.

Subsequent modeling also benefits from such a setup, as it is based on conventional CSG operations, minimizing the reconstruction task complexity. However, reliance on existing models requires flawless models and assumes their adherence to the CityGML standard, or similar. On the other hand, in case the models do not exist for a particular region of interest, they can be generated using such established solutions as PolyFit [Nan & Wonka, 2017] for reconstructing LoD2 without footprints; 3Dfier for prismatic generation of LoD1 building models [Ledoux et al., 2021]; or City3D for LoD2 building models [Huang et al., 2022].

In the introduced approach to CSG, the CSG traits are utilized to add new objects and delete unnecessary objects. Note that the same object is used to remove the unnecessary object, which step ensures retaining the watertightness of models. The objects are fitted from the pre-defined library of objects per type. Nowadays, there is a myriad of open-source libraries collecting façade elements in different architectural styles. The approach caters to a highly detailed appearance and minimizes topological and geometrical issues compared to the established from scratch approach. However, in the current experiment setup, only one element per object type was used. It provides limited variation to the appearance of the final façade elements, yet, in the co-authored publication [Froech et al., 2023b], the possible enhancement in variations of object types is shown.



IGHTING NEEDS TO BE REBUILT (291 unbuilt objects)

Figure 8.3 A screenshot from an Unreal Engine application showcasing usage of enriched models as a virtual testbed for testing car navigation systems: The LoD3 model with underpass mimics the real situation and enables drive through the building, impossible in footprint-extruded LoD2 or LoD1.

The key element of the presented strategy is to rely on given measurements. As such, the reconstruction's performance firmly correlates with the preceding semantic segmentation. Its score is embedded into the final façade elements to provide a measure of confidence for further downstream tasks and comparison for the next measurement epoch comparison. Also, in contrast to various other publications [Hensel et al., 2019], the geometries are not further co-aligned, as the reliance on symmetry might corrupt the factual state of façades coming from measurements. Intuitively, not all end-users are interested in feature-based applications but rather in succinct 3D visualizations. Therefore, to deliver full-coverage LoD3, this direction is explored in the supervised master's thesis employing generative adversarial networks and diffusion models to inpaint missing conflicts on façades [Froech, 2023].

9 Conclusion

In this chapter, I present the conclusion drawn from my conducted research. It commences with an overarching conclusion and then provides a more detailed explanation in the form of answers to the research questions posed at the beginning of this dissertation (section 1.1).

According to the Cambridge Dictionary, the word refinement means "a small change that improves something" or "the process of making a substance pure" Cambridge Dictionary [2023c]. In the case of the proposed strategy, both definitions are valid. On the one hand, the presented methods perform relatively minor additions to the 3D building model, significantly improving its accuracy and semantic completeness ("a small change that improves something"). On the other hand, the methods preserve valid existing geometry and remove invalid and obsolete geometry from LoD1 and 2 while reconstructing LoD3 building models ("the process of making a substance pure").

- To what extent can model priors of building models be used for the fusion of mobile mapping point clouds with building models?

The traits of standardized semantic 3D building models allow for certain assumptions, enabling the fusion of mobile mapping point clouds. The presented methods leveraged the planarity of 3D building models and established RANSAC plane point cloud segmentation to perform co-registration. A key factor is the assigned uncertainty value both to MLS measurements and building models, which creates a framework for uncertainty range. The subsequent analysis of model-to-point cloud fusion can be performed on a voxel level. Not intersecting, and distant façade elements are considered as possible matching. Therefore, it is concluded that MLS point clouds and 3D models can be matched within the confidence range and have plane structures for co-registration. However, façade-distant objects, such as flag poles, need more complex system knowledge relying not only on model priors. As shown in the proposed semantic segmentation method, such a concept can be extended by understanding the semantics of point clouds, which will support the subsequent model to point cloud matching. Here, of great importance are geometric priors estimated and added as features into the deep learning models, allowing for even up to 10% higher segmentation accuracy, totaling up to 87.5% accuracy.

- What level of geometric and semantic refinement of building models can be achieved using the conflict analysis approach?

As mentioned at the beginning of this chapter, the conducted refinements might be perceived as small. Yet, they contribute greatly to single models' overall geometric and semantic information gain. The geometric improvements, including underpasses, may reach up to 20% in terms of surface area and up to 18% when it comes to volume. The modeled underpasses also significantly contribute to navigation and flow-oriented algorithms, enabling passing through the building model, previously impossible in low LoD models.

When analyzing other façade elements, such as windows or doors, the solely geometric improvement compared to low LoD is relatively small. Yet, the prior-driven method exhibits traits over the mesh-based approaches. Namely, the strategy can reach up to 54% better reconstruction accuracy and, in contrast to mesh-based methods, ensures watertightness (Table 7.1). The watertightness is key to applications relying on volume calculation, such as energy demand estimation. At the same time, the accuracy of the model impacts every listed application of LoD3 models (Figure 2.4).

The impact on semantics gain is significant, as most of the laser-observable features are detected and reconstructed; for example, the refinement method can detect up to 91% of all openings. It provides a high completeness reconstruction of the present features. It is worth noting, however, that this assumption is valid for the facades that can be observed from the street level: The inner-yard facades or obscured facades cannot be reconstructed with this method. The excerpt about complementing missing façade features is presented in [Froech, 2023]. Nevertheless, measured-based street-level completeness greatly benefits systems operating on street-observable features, such as autonomous driving simulation and navigation. The exploration of such models application is undergone in the supervised thesis [Bieringer, 2023].

- What level of improvement of accuracy and semantic completeness of building models can be achieved by:
  - Conflict analysis with model and library knowledge?
  - Conflict analysis with the model, library knowledge, and images?

The experiments corroborate that complementing the conflict analysis with model-derived knowledge enhances the framework generalization. In this case, the author refers to the model as a subset of deep learning models. The experiments corroborate that conflict analysis improves 3D semantic segmentation of point clouds. For instance, the robustness (recall score) is improved by approximately 19% and the overall accuracy of segmentation by roughly 12%. The improvement is also noticeable for the single classes, especially for door and window classes, with an F1 score improving 12% and 11%, respectively. Yet, even though the overall accuracy reaches around 75%, the per-class scores for door of 13% or window of 60% require further development of the methods.

As shown by the experiments, the final reconstruction strongly relies on the façade semantic segmentation accuracy. Yet, with the library approach, the effect of accuracy on borders of segments is mitigated by fitting full shapes. This trait also prevents corrupted shape generation. The quantitative accuracy implies that the geometric accuracy of 0.34 RMS can be achieved solely by analyzing model-to-laser conflicts and using libraries. The completeness score reaches 78% of all detected openings and 90% for all laser-measured openings. The median intersection over union (IoU) of roughly 61% reflects the underestimation of openings due to occlusions and border effects.

This effect can be mitigated by supporting the Bayesian analysis with image-derived semantics. Even though on the global scale the improvement seems minor (up to 2%), such an addition can result in an improvement of roughly 20% for partially covered openings, for example, by blinds, as indicated in Figure 7.15. The detection robustness when adding image information also increases: 6% for all openings and 8% for laser-observed openings. The improvement of 24% is noticeable for the geometric-wise evaluation, too.

10 Implications and outlook

Reconstructing high-fidelity semantic 3D building models is considered a key element of creating urban digital twins. Thus, through developing novel 3D reconstruction methods, this thesis has contributed not only to the 3D building reconstruction field but also to generating urban digital twins. Even though contributions are significant, given the broad spectrum of challenges, further aspects still require consideration in the future. Although the methods are the key scientific contributions in the thesis, the work also elaborates on released open datasets and software in this section, as they aim to aid the further development of methods.

This thesis introduces two strategy-pertinent terms, i.e., refinement strategy and enrichment strategy. There is a subtle difference between these definitions. While the word *refinement* emphasizes improving the existing semantic and geometric accuracy of existing models, the *enrichment* implies the addition of new elements to the existing objects. This last chapter also highlights yet another term coined within this thesis that provides an even more abstract view of the proposed strategy: plastic surgery for 3D city models. As illustrated in Figure 10.1, the input to the system shall comprise unstructured data depicting the 3D reality in any form, be it 2D images or 3D scanner point clouds. The surgeon develops a set of methods that directly operate on the existing semantic 3D city model; crucially, the methods repair the model solely in conflict-relevant places. The final output shall represent the existing model with refined geometry and semantics while enriching its representation with new features. The author believes that the presented strategy and subsequent methods are of great relevance to both photogrammetry and computer vision as well as to the geoinformatics scientific community. The specific contributions are described in the following sections.

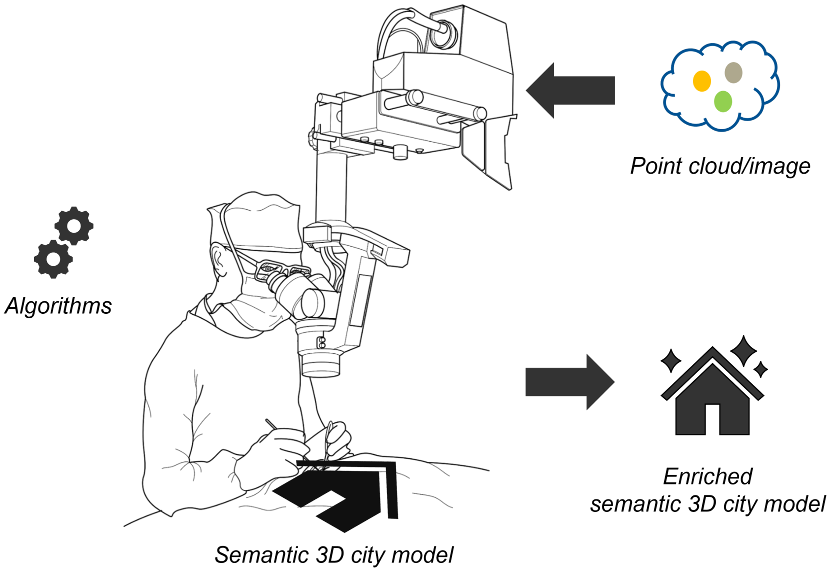


Figure 10.1 Within the thesis, terms such as refinement strategy and enrichment are used. Yet, the coined term plastic surgery for 3D city models provides a more abstract and generic view of the strategy, too.

10.1 Photogrammetry and computer vision community

10.1.1 Methods

Conflict analysis The pivotal part of the introduced methods is the ray-to-model analysis, or the coined term conflict analysis. As shown in the thesis, it provides a physical measure of the analysis process, assuming that a ray penetrates a wall when there is an opening. Such strong deterministic information enables further investigations for semantic segmentation and reconstruction beyond the realm of 3D building models. It underlines the importance of 3D scene perception, which is not limited to merely the 3D point cloud as an end-product but also to the whole measurement campaign with its positioning and laser ray physics uncertainties. Not only are conflicts important in the analysis but so is the ability to determine unknown 3D space, which is limited, if not impossible, relying only on 3D point clouds. Such determination allows discarding the vast majority of space, allowing for the conduct, for example, of more robust change detection. Moreover, the presented conflict analysis framework allows for identifying facade-extruded (e.g., balconies) and not only facade-intruded (e.g., windows) objects. Since conflict map results are prone to occlusions, inpainting techniques can be adapted to mitigate shadow-like features on the conflict areas [Froech et al., 2025]. Such a feature opens new possibilities for facade reconstruction approaches, which shall be investigated in future work. The author strongly believes that the 3D change detection, segmentation, and reconstruction community should investigate this approach further.

Bayesian networks Fusion approaches, as in the case of the proposed early-fusion of geometric features or late-fusion of analyzed conflicts, serve as a sound solution for circumventing data imbalance for specific tasks. The impact of different fusion stages on the performance of deep learning networks is worth investigating. Of special interest is conflict analysis, as it provides deterministic, physical conflict information into the stochastic reasoning of the network. In the case of this work, the Bayesian network is explored for the late fusion. The Bayesian network has a very important trait of including data but also expert evidence, which is essential in data-sparse issues. It is worth exploring alternative multi-layer perceptron (MLP) late-fusion approaches, as they derive weights based on provided data (evidence) and not expert-assigned weights. However, as it is a supervised learning problem, this approach will necessitate a great deal of annotated datasets. Future work shall be dedicated to exploring this research direction.

Prior models Yet another research direction frequently addressed proposes capitalizing on the advent of image-based deep learning networks by projecting the labels onto 3D point clouds. This research direction shows promising results; however, it is limited to the camera field of view and the image's two-dimensional nature, with its errors and flaws inevitably propagated onto 3D laser point clouds. Moreover, the façade reconstruction proves difficult if the orthorectification is not applied, which requires scale information extracted from control points. The presented thesis shows that street-level images can enhance the 3D reconstruction of façade elements when target planes of geometric priors are used. Essentially, it skips the traditional orthorectification approach. The ubiquity of such building models and the CityGML standard enforcing planar wall surfaces provides a promising research direction.

Multimodal uncertainty Currently, the presented methods assume a level of uncertainty of the given models and measurements. Such an assumption addresses the issue, however, not completely. The uncertainty of a 3D model itself can stem from various sources, which can result, for example, in footprint accuracy different from eve accuracy, as the footprints can be measured by the surveyor and eves extracted from a certain assumed height of ALS measurements. The laser scanner measurements have varying accuracy along the ray as well and resemble a conic shape in reality, extruding along the measurement distance. In the case of the shown uncertainty model, it is assumed that a ray follows the defined vector and does not deviate. The influence of more detailed uncertainty modeling should be addressed in future work. Nevertheless, the uncertainty in the presented approach is propagated throughout the process, starting with conflict analysis, through semantic segmentation, and ending with 3D reconstruction. Such an approach is of great importance since if the uncertainty is skipped at one stage, it is only implicitly propagated, leading to false assumptions and uncertainty for further processing stages. The presented framework also allows incorporating different modalities than optical images or geometric point clouds. It is foreseeable that thermal point cloud reconstruction and thermal image understanding may profit from the proposed workflow. Also, the uncertainty becomes increasingly important for reconstruction using optical images and Gaussian Splatting [Zhang et al., 2024].

Façade elements reconstruction As shown in this thesis, the methods' evaluation focus lies on reconstructing openings, such as doors, windows, or underpasses. Although these are essential parts of buildings that are key to LoD 3 reconstruction and downstream tasks, the test on façade-extruded elements shall be conducted in the future. In fact, the conflict analysis framework is expected to not only distinguish wall-intruded elements but also wall-extruded elements within the defined uncertainty range. Such deterministic, physical information has proved critical in this study, and it is foreseen to be critical also for wall-extruded elements. As such, it is believed that the strategy applicability to other façade elements should be investigated in future work.

10.1.2 Open datasets and software

TUM-FAÇADE Within the course of this thesis the façade-level point cloud benchmark dataset – TUM-FAÇADE – has been released [Wysocki et al., 2022c]. At the time of writing, it comprises 333 million manually annotated points, facilitating training and validation of both stochastic and deterministic semantic segmentation methods. The dataset is in active development, and newly acquired point cloud subsets can enrich the testing capabilities. Currently, the dataset is being reviewed as a standard semantic segmentation dataset benchmark of the widely-used Open3D ML framework [Zhou et al., 2018].

Cities around the world possess distinct architectural styles, composed of modern football stadiums but also of medieval churches, simple concrete dwellings, and sophisticated marble single houses. This variety itself poses numerous generalization challenges. Although the selected testing sample had comprised various architectural building types, it showcased the methods' applicability and assumptions on a typical central European architecture. As such, the ultimate generalization capabilities of the presented methods require worldwide testing, with presumably several samples of cities around the world to showcase the methods' generalization. The increasing interest in the openly released TUM-FAÇADE dataset shows that this study holds great potential to inspire other researchers to pursue this path and release new façade-level datasets unlocking such tests. It is anticipated that the increasing availability of mobile mapping systems and their data will contribute to this trend.

Furthermore, the developments of new open façade-level benchmark datasets will facilitate solving the imbalanced data problem. Observing fields such as image recognition or natural language processing, one can see that methods and data availability solved many challenges. If this trend is projected onto the point cloud segmentation, it will aid the robustness of the façade-level semantic segmentation methods, ultimately enhancing detailed 3D reconstruction. Yet, labeling 3D data is a much more complex than images or text, whereas its automatic label extraction is also limited. To overcome this issue, a great deal of resources need to be invested.

TUM2TWIN The TUM2TWIN dataset [Barbosa et al., 2023] is a project initiated within the scope of this thesis, which comprises 14 open data LoD3 building models and its respective LoD2 textured building models; more are expected as the dataset is actively developed. The building models overlap with the point clouds in the TUM-FAÇADE dataset, providing a holistic 3D representation of the environment.

One of the possible approaches to minimize the labeling effort is using highly-detailed LoD3 models for automatic labeling of geo-referenced data, as shown in the supervised project investigating the applicability of synthetic point cloud data created based on LoD3 models [Schwab et al., 2023b]. Furthermore, such detailed models can serve for 2D data annotation: texture masks for openings can be easily generated. This trait yields another possibility for testing detection and segmentation algorithms on a model surface.

Another advantage of having ground-truth LoD3 building models is opening new 3D semantic reconstruction evaluation possibilities, especially when 3D semantic point clouds depict the same objects. To the best of my knowledge, it is the first dataset of its kind, enabling testing the whole pipeline from the point cloud acquisition, façade-level semantic segmentation, to 3D semantic reconstruction. It is believed that such datasets are necessary for the community tackling each sub-task of such pipeline, but also for holistic workflow testing.

CityGML2OBJ 2.0 The converters holding a semantic of 3D building models are rare and mostly proprietary: To foster the research about semantic 3D reconstruction, this thesis contributed to a release of an updated open converter of CityGML encoded in GML to OBJ file format, while retaining the input semantics [Froech et al., 2023a].

Such a converter is needed in the community, as most photogrammetry and computer vision libraries support the OBJ data format, but not the more complex CityGML. One of the reasons is a different motivation for using 3D geometries: The simplicity of such formats as OBJ enables relatively easy validation of 3D reconstruction methods. Also, the geoinformatics software that handles CityGML reading and writing is typically not well-known in the photogrammetry and computer vision communities. Therefore, it is believed that a lightweight converter can bridge the gap between communities and foster the development of 3D reconstruction methods.

10.2 Geoinformatics community

10.2.1 Methods

Conflict analysis The ray-to-model analysis shows great potential in detecting missing objects in existing models. As such, the conflict probability maps projected as textures to models enrich the building model representation. It may prove essential to the development of energy simulation methods, owing to the fact that conflict maps depict mainly openings in the model. This trait is of great importance for the analysis of openings and non-openings area of models, which trait is often used to assess potential heat loss in energy conversion efficiency approaches. Further foreseeable methods may include analyzing the at-scale quality of reconstructed LoD2 and LoD1 building models. It is expected that the presented strategy can be adapted to analyze other urban elements' quality, such as road surfaces or city furniture. Moreover, the conflict analysis is based on the voxel structure, which is expected to generalize towards issues such as analyzing clearance of road spaces, often represented as oriented bounding boxes.

Bayesian networks The pivotal characteristic of Bayesian networks is their ability to address input data uncertainties throughout the process. It is of special importance to the final LoD3 models, as further analysis relies not only on accuracy but also on data fidelity. However, in current reconstruction approaches, this information is frequently neglected. Such information opens new research directions for 3D building models in various downstream tasks. For example, the analysis of 3D building models in supporting the positioning of autonomous cars, where the provided object-assigned probability score can play a vital role. Furthermore, the presented approaches show how Bayesian networks may operate on 3D building models, showing great potential in examining evidence and expert domain knowledge: Similarly to inspecting avalanche risk [Stritih et al., 2020], expert knowledge may prove indispensable in tackling data-sparse scenarios.

Prior models The overarching goal of the thesis is to leverage the existing LoD1 and LoD2 building models and upgrade them into LoD3 models. As mentioned before, this approach has several advantages to the processing itself, but, it might prove even more essential for the cadastre-oriented development. The low LoD building models are frequently created based on the cadastre-derived footprints and complemented with aerial observations for roof type and height modeling. Crucially, such semantic 3D building models and their entities posses hyperlinks to external data sources, such as solar potential analysis, internet of things sensors, and energy demand estimation. The 3D models also have topological relations to other models in the set and hierarchical relations within each model. Owing to that, as in my approach, the hyperlinks should remain intact to retain the links; in contrast, the loss of such links is imposed in the case of the from-scratch approach. Furthermore, parts of backyard-facing walls and roof surfaces will be concealed to the street-level measurements, disabling at-once full LoD3 reconstruction. Therefore, it is believed that incremental refinement is a promising research direction, and even incomplete, can serve multiple purposes and be once again updated using next measurement epoch. Interestingly, reconstructed LoD3 models can also serve for labeling other modalities with semantic information, such as thermal images and point clouds [Zhu et al., 2024].

Multimodal uncertainty In this work, multimodal uncertainties stemming from unstructured and structured datasets are addressed. Presumably, such challenge is and will be pertinent to any spatial-data-related field of science, with geoinformatics being at the forefront of these challenges. Due to the current influx of various datasets depicting the same phenomenon with different fidelity, the traditional assessment solely on relative or global accuracy becomes insufficient; additionally assessing data fidelity (uncertainty) increases. As such, the prevailing question of choosing data for a particular downstream task depends on the highest available accuracy and reliability. It is believed that this research avenue is worthwhile, especially given the interest of the geoinformatics community in data structure development.

Façade elements reconstruction As the refinement strategy opens a new chapter in the approach of LoD3 reconstructions, a set of guidelines and new functions enabling seamless mapping of newly created features onto LoD3 building data model are proposed. Such findings may prove indispensable in further research, providing production-ready LoD3 building models into the geoinformatics community; without a hurdle of developing additional mapping processes. Moreover, as indicated by the recently introduced CityGML 3.0 Point Cloud module [Kutzner et al., 2020], 3D point clouds have become a key part of city models. This thesis also investigates the translation of point cloud class-wise semantics into the standardized CityGML classes. This translation shall allow bidirectional mapping of such classes and ease the fusion of point clouds and surface models. Presumably, hybrid solutions will prevail, offering explicit surface modeling and, in case of small and unique façade objects (e.g., a gargoyle of less than 2m^3) storing it as a set of 3D points.

10.2.2 Open datasets and software

Awesome CityGML is a project aiming to list all the openly available semantic 3D city models worldwide [Wysocki et al., 2022d]; it boasts 57 regions and cities in 18 countries. Olaf Wysocki commenced this project as a main contributor, but the goal is to encourage the community to add their known datasets. At the time of writing, the repository has been *starred* by more than 260 GitHub users. There are four explicit contributors, but there have also been several direct emails with a request to add a dataset. The list is officially recognized by the Open Geospatial Consortium (OGC) website and listed on their website as well [Open Geospatial Consortium, 2023]. Currently, there are more than 215 million buildings freely available across the world Wysocki et al. [2024a].

Such a repository is needed in the community, as the openly available city models are scattered, and finding the required 3D model is often impeded by language barriers or requires insider knowledge. Without knowing which models are available, researchers and practitioners are limited in testing their novel solutions. On the other hand, the open city models are meant to be used by the community: the stronger the exposure, the higher the usage of models. With the community's help, the repository is poised to grow and gather all available datasets worldwide.

TUM2TWIN [Barbosa et al., 2023] and **LoD3 road space models** [Schwab et al., 2021] are two datasets comprising LoD3 building models, totaling 69 building models, representing 14 buildings in Munich, Germany and 55 in Ingolstadt, Germany. Such manually created LoD3 models are the first examples of real-world city districts at this level of detail, which provides a platform for developing new use cases for such detailed building models. They can also serve as practical examples and guidelines for creating similar benchmarks; additionally, a modeling manual for manual modeling was created.

The datasets have been used in-house already, but other institutes are exploring their possibilities. For example, the LoD3 road space models were used for simulating synthetic semantic point clouds and training of deep neural network for façade-level segmentation [Schwab et al., 2023b]. Essentially, it underlines that once created, virtual 3D models can be used for annotating data, providing even more data for training neural networks that will facilitate the creation of yet other 3D models; eventually resulting in a self-perpetuating mechanism. This direction of research is expected to be investigated in the coming years.

Bibliography

3D Mapping Solutions (2023) MoSES mobile mapping platform - technical details. https://www.3d-mapping.de/ueber-uns/unternehmensbereiche/data-acquisition/unser-vermessungssystem/. Accessed: 2023-01-30.

Aijazi AK (2014) 3D urban cartography incorporating recognition and temporal integration. PhD thesis, Université Blaise Pascal, Clermont-Ferrand, France.

Aleksandrov M, Zlatanova S, Kimmel L, Barton J, Gorte B (2019) Voxel-based visibility analysis for safety assessment of urban environments. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-4/W8: 11–17.

Amirebrahimi S, Rajabifard A, Mendis P, Ngo T (2016) A BIM-GIS integration method in support of the assessment and 3D visualisation of flood damage to a building. Journal of Spatial Science, 61 (2): 317–350.

Anderson L, Cheek D, Aragao L, Andere L, Duarte B, Salazar N, Lima A, Duarte V, Arai E (2017) Development of a point-based method for map validation and confidence interval estimation: A case study of burned areas in Amazonia. Journal of Remote Sensing and GIS, 6 (1): 193.

Apostolopoulou A, Zhu M, Jin J (2023) Parametric assessment of building heating demand for different levels of details and user comfort levels: A case study in London, UK. Sustainability, 15 (10).

Bamler R, Shi Y (2022) Estimation Theory: 2 - Recap of basic propability theory - part 4. 13-5-2022, Chair of Remote Sensing Technology Technical University of Munich. Lecture of Estimation Theory.

Barbosa J, Schwab B, Wysocki O, Heeramaglore M, Huang X (2023) tum2twin: Repository of CityGML LOD3 models of the Technical University of Munich. https://github.com/tum-gis/tum2twin/tree/main. Accessed: 2023-09-22.

Basaraner M, Cetinkaya S (2017) Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS. International Journal of Geographical Information Science, 31 (10): 1952–1977.

Batty M (2018) Digital twins. Environment and Planning B: Urban Analytics and City Science, 45 (5): 817–820.

Becker S (2011) Automatische Ableitung und Anwendung von Regeln für die Rekonstruktion von Fassaden aus heterogenen Sensordaten. PhD thesis, Universität Stuttgart, Stuttgart, Germany.

Becker S, Haala N (2007) Refinement of building façades by integrated processing of LiDAR and image data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3/W49A: 7–12.

Beil C, Ruhdorfer R, Coduro T, Kolbe TH (2020) Detailed streetspace modelling for multiple applications: Discussions on the proposed CityGML 3.0 Transportation Model. ISPRS International Journal of Geo-Information, 9 (10): 603.

Berzins V (1984) Accuracy of Laplacian edge detectors. Computer Vision, Graphics, and Image Processing, 27 (2): 195–210.

Besl P, McKay ND (1992) A method for registration of 3D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 (2): 239–256.

Bieringer A (2023) Evaluation of the effect of enriched facade models on image-based localization of vehicles. Master's thesis, Chair of Photogrammetry and Remote Sensing, Technical University of Munich, Munich, Germany.

Bieringer A, Wysocki O, Tuttas S, Hoegner L, Holst C (2024) Analyzing the impact of semantic LoD3 building models on image-based vehicle localization. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 10: 55–62.

Biljecki F (2020) Exploration of open data in Southeast Asia to generate 3D building models. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, VI-4/W1-2020: 37–44.

Biljecki F, Ito K (2021) Street view imagery in urban analytics and GIS: A review. Landscape and Urban Planning, 215: 104217.

Biljecki F, Ledoux H, Stoter J (2016) An improved LOD specification for 3D building models. Computers, Environment and Urban Systems, 59: 25–37.

Biljecki F, Ledoux H, Stoter J, Zhao J (2014) Formalisation of the level of detail in 3D city modelling. Computers, Environment and Urban Systems, 48: 1–15.

Biljecki F, Stoter J, Ledoux H, Zlatanova S, Çöltekin A (2015) Applications of 3D city models: State of the art review. ISPRS International Journal of Geo-Information, 4 (4): 2842–2889.

Blaser S, Meyer J, Nebiker S (2021) Open urban and forest datasets from a high-performance mobile mapping backpack – a contribution for advancing the creation of digital city twins. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B1-2021: 125–131.

Brenner C, Ripperda N (2006) Extraction of facades using RJMCMC and constraint equations. Photogrammetric Computer Vision, 36: 155–160.

Bronstein AM, Bronstein MM, Guibas LJ, Ovsjanikov M (2011) Shape google: Geometric words and expressions for invariant shape retrieva. ACM Transactions on Graphics, 30 (1): 1–20.

Calonder M, Lepetit V, Ozuysal M, Trzcinski T, Strecha C, Fua P (2012) BRIEF: Computing a local binary descriptor very fast. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (7): 1281–1298.

Cambridge Dictionary (2023a) Facade definition, Cambridge University Press & Assessment 2023. https://dictionary.cambridge.org/dictionary/english/facade. Accessed: 2023-01-30.

Cambridge Dictionary (2023b) Front definition, Cambridge University Press & Assessment 2023. https://dictionary.cambridge.org/dictionary/english/front. Accessed: 2023-01-30.

Cambridge Dictionary (2023c) Refinement definition, Cambridge University Press & Assessment 2023. https://dictionary.cambridge.org/dictionary/english/refinement. Accessed: 2023-01-30.

Carlone L, Tron R, Daniilidis K, Dellaert F (2015) Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization. IEEE international conference on robotics and automation (ICRA), : 4597–4604.

Cha SH (2008) Taxonomy of nominal type histogram distance measures. In: Long C, Sohrab SH, Bognar G, Perlovsky L (eds) MATH'08: Proceedings of the American Conference on Applied Mathematics: 325–330.

Chacon M (2020) Sign guidelines and applications manual. Texas Department of Transportation. http://onlinemanuals.txdot.gov/txdotmanuals/smk/index.htm. Accessed: 2021-12-01.

Chaidas K, Tataris G, Soulakellis N (2021) Seismic damage semantics on post-earthquake LOD3 building models generated by UAS. ISPRS International Journal of Geo-Information, 10 (5).

Chen SH, Pollino CA (2012) Good practice in Bayesian Network modelling. Environmental Modelling & Software, 37: 134–145.

Chuprikova E (2019) Visualizing uncertainty in reasoning. Phd thesis, Chair of Cartography, Technical University of Munich, Munich, Germany.

Cignoni P, Rocchini C, Scopigno R (1998) Metro: Measuring error on simplified surfaces. Computer Graphics Forum, 17 (2): 167–174.

Csurka G, Dance C, Fan L, Willamowski J, Bray C (2004) Visual categorization with bags of keypoints. European Conference on Computer Vision Workshop (ECCVW), 1 (1-22): 1–2.

Dai A, Nießner M (2018) 3DMV: Joint 3D-multi-view prediction for 3D semantic scene segmentation. Proceedings of the European Conference on Computer Vision (ECCV), : 452–468.

Dalal N, Triggs B (2005) Histograms of Oriented Gradients for human detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1: 886–893.

De Deuge M, Quadros A, Hung C, Douillard B (2013) Unsupervised feature learning for classification of outdoor 3D scans. Australasian conference on robitics and automation, 2 (1).

Deschaud JE, Duque D, Richa JP, Velasco-Forero S, Marcotegui B, Goulette F (2021) Paris-CARLA-3D: A real and synthetic outdoor point cloud dataset for challenging tasks in 3D Mapping. Remote Sensing, 13 (22): 4713.

Dong Z, Liang F, Yang B, Xu Y, Zang Y, Li J, Wang Y, Dai W, Fan H, Hyyppä J et al. (2020) Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 163: 327–342.

Douglas DH, Peucker TK (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica: The International Journal for Geographic Information and Geovisualization, 10 (2): 112–122.

Dukai B, Peters R, Wu T, Commandeur T, Ledoux H, Baving T, Post M, van Altena V, van Hinsbergh W, Stoter J (2020) Generating, storing, updating and disseminating a countrywide 3D model. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIV-4/W1-2020: 27–32.

Fan H, Wang Y, Gong J (2021) Layout graph model for semantic façade reconstruction using laser point clouds. Geo-spatial Information Science, 24 (3): 403–421.

FGSV (1996) Richtlinien für die Anlage von Straßen, Teil: Querschnitte, Ausgabe 1996; RAS-Q 96, Forschungsgesellschaft für Straßen- und Verkehrswesen e.V. https://beck-online.beck.de/. Accessed: 2021-11-16.

Fiutak G, Marx C, Willkomm P, Donaubauer A (2018) Projekt 3D Digitales Landschaftsmodell (3D-DLM) Am Runden Tisch GIS e.V., Abschlussbericht (Demonstrationsphase): Datenvorverarbeitung, Anwendung Des 3Dfiers, Abbildung Auf CityGML-Datenmodell, Bereitstellung Der Ergebnisdaten & Qualitätsbewertung. https://rundertischgis.de/projektarbeit/3d-digitales-landschaftsmodell.html. Accessed: 2021-11-11.

Froech T (2023) Inpainting of unseen façade objects using deep learning methods. Master's thesis, Chair of Geoinformatics, Technical University of Munich, Munich, Germany.

Froech T, Schwab B, Wysocki O (2023a) CityGML2OBJ 2.0: Command line converter of CityGML (.gml) to OBJ (.obj) files, while maintaining the semantics. https://github.com/tum-gis/CityGML2OBJv2. Accessed: 2023-08-23.

Froech T, Wysocki O, Hoegner L, Stilla U (2023b) Reconstructing façade details using MLS point clouds and Bag-of-Words approach. Recent Advances in 3D Geoinformation Science (proceedings of 3D GeoInfo 2023), : 337–355. Cham: Springer.

Froech T, Wysocki O, Xia Y, Junyu X, Schwab B, Cremers D, Kolbe TH (2025) FacaDiffy: Inpainting unseen facade parts using diffusion models. Accepted to the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Geospatial Week 2025).

Gadde R, Marlet R, Paragios N (2016) Learning grammars for architecture-specific facade parsing. International Journal of Computer Vision, 117 (3): 290–316.

Gargoum SA, Karsten L, El-Basyouny K, Koch JC (2018) Automated assessment of vertical clearance on highways scanned using mobile LiDAR technology. Automation in Construction, 95: 260–274.

Gehrung J, Hebel M, Arens M, Stilla U (2017) An approach to extract moving objects from MLS data using a volumetric background representation. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-1/W1: 107–114.

Gehrung JA (2022) Change detection in point clouds of urban street spaces using fuzzy spatial reasoning. PhD thesis, Chair of Photogrammetry and Remote Sensing, Technical University of Munich, Munich, Germany.

Geyer J, Kassahun Y, Mahmudi M, Ricou X, Durgesh R, Chung AS, Hauswald L, Pham VH, Mühlegg M, Dorn S et al. (2020) A2D2: Audi autonomous driving dataset. arXiv preprint arXiv:2004.06320.

Goebbels S, Pohle-Fröhlich R (2018) Line-based registration of photogrammetric point clouds with 3D city models by means of mixed integer linear programming. International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), : 299–306.

Goebbels S, Pohle-Fröhlich R, Kant P (2018) A linear program for matching photogrammetric point clouds with CityGML building models. In: Operations Research Proceedings 2017: 129–134.

Goebbels S, Pohle-Fröhlich R, Pricken P (2019) Iterative closest point algorithm for accurate registration of coarsely registered point clouds with CityGML models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-2/W5: 201–208.

Grêt-Regamey A, Straub D (2006) Spatially explicit avalanche risk assessment linking Bayesian Networks to a GIS. Natural Hazards and Earth System Sciences, 6 (6): 911–926.

Griffiths D, Boehm J (2019a) A review on deep learning techniques for 3D sensed data classification. Remote Sensing, 11 (12): 1499.

Griffiths D, Boehm J (2019b) SynthCity: A large scale synthetic point cloud. arXiv preprint arXiv:1907.04758.

Grilli E, Farella EM, Torresani A, Remondino F (2019) Geometric features analysis for the classification of cultural heritage point clouds. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W15: 541–548.

Grilli E, Remondino F (2020) Machine learning generalisation across different 3D architectural heritage. ISPRS International Journal of Geo-Information, 9 (6): 379.

Gröger G, Kolbe TH, Nagel C, Häfele KH (2012) OGC City Geography Markup Language CityGML Encoding Standard version 2.0.0. Open Geospatial Consortium: Wayland, MA, USA, 2012, OGC Doc No 12-019.

Haala N, Kada M (2010) An update on automatic 3D building reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing, 65 (6): 570 – 580.

Hackel T, Savinov N, Ladicky L, Wegner JD, Schindler K, Pollefeys M (2017) Semantic3d.net: A new large-scale point cloud classification benchmark. arXiv preprint arXiv:1704.03847.

Haklay M (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and planning B: Planning and design, 37 (4): 682–703.

Hartigan JA, Wong MA (1979) Algorithm AS 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics), 28 (1): 100–108.

Hazra A (2017) Using the confidence interval confidently. Journal of Thoracic Disease, 9 (10): 4125.

He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. IEEE/CVF International Conference on Computer Vision (ICCV), : 2961–2969.

Hebel M (2012) Änderungsdetektion in urbanen Gebieten durch objektbasierte Analyse und schritthaltenden Vergleich von Multi-Aspekt ALS-Daten. PhD thesis, Technical University of Munich, Munich, Germany.

Hebel M, Arens M, Stilla U (2013) Change detection in urban areas by object-based analysis and on-the-fly comparison of multi-view ALS data. ISPRS Journal of Photogrammetry and Remote Sensing, 86: 52–64.

Heeramaglore M, Kolbe TH (2022) Semantically enriched voxels as a common representation for comparison and evaluation of 3D building models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-4/W2-2022: 89–96.

Hensel S, Goebbels S, Kada M (2019) Facade reconstruction for textured LOD2 CityGML models based on deep learning and mixed integer linear programming. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-2/W5: 37–44.

Hijazi I, Donaubauer A, Kolbe TH (2018) BIM-GIS integration as dedicated and independent course for geoinformatics students: Merits, challenges, and ways forward. ISPRS International Journal of Geo-Information, 7 (8): 319.

Hirt PR, Xu Y, Hoegner L, Stilla U (2021) Change detection of urban trees in MLS point clouds using occupancy grids. PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 89 (4): 301–318.

Hoegner L, Gleixner G (2022) Automatic extraction of facades and windows from MLS Point clouds using voxelspace and visibility analysis. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2022: 387–394.

Holst KH, Holst R (2004) Brücken aus Stahlbeton und Spannbeton: Entwurf, Konstruktion und Berechnung. Berlin, Germany: Ernst & Sohn.

Hornung A, Wurm KM, Bennewitz M, Stachniss C, Burgard W (2013) OctoMap: An efficient probabilistic 3D mapping framework based on octrees. Autonomous Robots, 34 (3): 189–206.

Hou Y, Biljecki F (2022) A comprehensive framework for evaluating the quality of street view imagery. International Journal of Applied Earth Observation and Geoinformation, 115: 103094.

Huang H, Michelini M, Schmitz M, Roth L, Mayer H (2020) LOD3 building reconstruction from multi-source images. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2020: 427–434.

Huang J, Stoter J, Peters R, Nan L (2022) City3D: Large-scale building reconstruction from airborne LiDAR point clouds. Remote Sensing, 14 (9).

Huang R (2021) Change detection of construction sites based on 3D point clouds. PhD thesis, Chair of Photogrammetry and Remote Sensing, Technical University of Munich, Munich, Germany.

Iwaszczuk D, Hoegner L, Stilla U (2011) Detection of windows in IR building textures using masked correlation. In: Stilla U, Rottensteiner F, Mayer H, Jutzi B, Butenuth M (eds) Photogrammetric Image Analysis, ISPRS Conference - Proceedings. Lecture Notes in Computer Science: 133–146.

Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y (2017) The one hundred layers tiramisu: Fully convolutional dense nets for semantic segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), : 11–19.

Kanezaki A, Matsushita Y, Nishida Y (2018) RotationNet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), : 5010–5019.

Kang Z, Yang J (2018) A probabilistic graphical model for the classification of mobile LiDAR point clouds. ISPRS Journal of Photogrammetry and Remote Sensing, 143: 108–123.

Kazhdan M, Hoppe H (2013) Screened Poisson surface reconstruction. ACM Transactions on Graphics (ToG), 32 (3): 1–13.

Kjaerulff UB, Madsen AL (2008) Bayesian Networks and influence diagrams. Springer Science+ Business Media, 200: 114.

Kolbe TH, Donaubauer A (2021) Semantic 3D City Modeling and BIM. In: Shi W, Goodchild MF, Batty M, Kwan MP, Zhang A (eds) Urban Informatics: 609–636. Springer Singapore.

Kolbe TH, Gröger G, Plümer L (2005) CityGML: Interoperable access to 3D city models. In: Geo-information for disaster management (pp. 883–899). Springer.

Kolbe TH, Gröger G, Plümer L (2008) CityGML–3D city models and their potential for emergency response. Geospatial Information Technology for Emergency Response, 257: 273–290.

Korc F, Förstner W (2009) eTRIMS Image Database for interpreting images of man-made scenes. Deptartment of Photogrammetry, University of Bonn, Technical Report TR-IGG-P-2009-01, : 1–12.

Kutzner T, Chaturvedi K, Kolbe TH (2020) CityGML 3.0: New functions open up new applications. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 88 (1): 1–19.

Lande MB (2012) Automatic registration of partially overlapping terrestrial laser scanner point clouds. https://prs.igp.ethz.ch/research/completed_projects/automatic_registration_of_point_clouds.html. Accessed: 2020-10-30.

Ledoux H, Arroyo Ohori K, Kumar K, Dukai B, Labetski A, Vitalis S (2019) CityJSON: A compact and easy-to-use encoding of the CityGML data model. Open Geospatial Data, Software and Standards, 4 (1): 1–12.

Ledoux H, Biljecki F, Dukai B, Kumar K, Peters R, Stoter J, Commandeur T (2021) 3dfier: Automatic reconstruction of 3D city models. Journal of Open Source Software, 6 (57): 2866.

Lee D, Pietrzyk P, Donkers S, Liem V, van Oostveen J, Montazeri S, Boeters R, Colin J, Kastendeuch P, Nerry F et al. (2013) Modelling and observation of heat losses from buildings: The impact of geometric detail on 3D heat flux modelling. 33rd AERSel Symposium Towards Horizon 2020: Earth Observation and Social Perspectives, : 3–6.

Li Y, Bu R, Sun M, Wu W, Di X, Chen B (2018) PointCNN: Convolution on x-transformed points. Advances in Neural Information Processing System (NeurIPS), 31.

Li Y, Ma L, Zhong Z, Liu F, Chapman MA, Cao D, Li J (2020) Deep learning for LiDAR point clouds in autonomous driving: A review. IEEE Transactions on Neural Networks and Learning Systems, 32 (8): 3412–3432.

Liao Y, Xie J, Geiger A (2021) KITTI-360: A novel dataset and benchmarks for urban scene understanding in 2D and 3D. arXiv preprint arXiv:2109.13410.

Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick LC (2014) Microsoft COCO: Common objects in context. arXiv preprint arXiv:1405.0312.

Liu H, Xu Y, Zhang J, Zhu J, Li Y, Hoi SC (2020) DeepFacade: A deep learning approach to facade parsing with symmetric loss. IEEE Transactions on Multimedia, 22 (12): 3153–3165.

Liu S (2022) Vehicle detection in aerial images using neural networks with synthetic training data. Master's thesis, Chair of Photogrammetry and Remote Sensing, Technical University of Munich, Munich, Germany.

Lowe D (1999) Object recognition from local scale-invariant features. IEEE/CVF International Conference on Computer Vision (ICCV), 2: 1150–1157.

Lucks L, Klingbeil L, Plümer L, Dehbi Y (2021) Improving trajectory estimation using 3D city models and kinematic point clouds. Transactions in GIS, 25 (1): 238–260.

Masante D (2020) bnspatial: Package for the spatial implementation of Bayesian Networks and mapping in geographical space. Version 1.1.1. https://github.com/dariomasante/bnspatial. Accessed: 2023-06-11.

Matrone F, Grilli E, Martini M, Paolanti M, Pierdicca R, Remondino F (2020) Comparing machine and deep learning methods for large 3D heritage semantic segmentation. ISPRS International Journal of Geo-Information, 9 (9): 535.

Maturana D, Scherer S (2015) VoxNet: A 3D convolutional neural network for real-time object recognition. International conference on intelligent robots and systems (IROS), : 922–928.

Mayer H (2008) Object extraction in photogrammetric computer vision. ISPRS Journal of Photogrammetry and Remote Sensing, 63 (2): 213–222.

Mayer H, Reznik S (2007) Building facade interpretation from uncalibrated wide-baseline image sequences. ISPRS Journal of Photogrammetry and Remote Sensing, 61 (6): 371–380.

Meyer T, Brunn A, Stilla U (2021) Accuracy investigation on image-based change detection for BIM compliant indoor models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4: 105–112.

Meyer T, Brunn A, Stilla U (2022) Change detection for indoor construction progress monitoring based on BIM, point clouds and uncertainties. Automation in Construction, 141: 104442.

Minghini M, Brovelli MA, Frassinelli F (2018) An open source approach for the intrinsic assessment of the temporal accuracy, up-to-dateness and lineage of OpenStreetMap. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W8: 147–154.

Moravec H, Elfes A (1985) High resolution maps from wide angle sonar. In: Proceedings of the 1985 IEEE International Conference on Robotics and Automation, 2: 116–121.

Muñumer Herrero E, Ellul C, Morley J (2018) Testing the impact of 2D generalisation on 3D models – exploring analysis options with an off-the-shelf software package. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W10: 119–126.

Munoz D, Bagnell JAD, Vandapel N, Hebert M (2009) Contextual classification with functional Max-Margin Markov networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), : 975 – 982.

Musialski P, Wonka P, Aliaga DG, Wimmer M, Van Gool L, Purgathofer W (2013) A survey of urban reconstruction. Computer graphics forum, 32 (6): 146–177.

Nan L, Wonka P (2017) PolyFit: Polygonal surface reconstruction from point clouds. IEEE/CVF International Conference on Computer Vision (ICCV), : 2353–2361.

Neis P, Zielstra D, Zipf A (2012) The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. Future Internet, 4 (1): 1–21.

Nguyen SH, Kolbe TH (2021) Modelling changes, stakeholders and their relations in semantic 3D city models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 8: 137–144.

Nguyen SH, Kolbe TH (2022) Path-tracing semantic networks to interpret changes in semantic 3D city models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-4/W2-2022: 217–224.

Nouvel R, Schulte C, Eicker U, Pietruschka D, Coors V (2013) CityGML-based 3D city model for energy diagnostics and urban energy policy support. IBPSA World, 2013: 1–7.

Nowak E, Jurie F, Triggs B (2008) Sampling strategies for bag-of-features image classification. European Conference on Computer Vision (ECCV), 3954: 490–503.

Open Geospatial Consortium (2023) CityGML, Related Links. `https://www.ogc.org/standard/citygml/`. Accessed: 2023-10-01.

Palliwal A, Song S, Tan HTW, Biljecki F (2021) 3D city models for urban farming site identification in buildings. Computers, Environment and Urban Systems, 86: 101584.

Pang HE, Biljecki F (2022) 3D building reconstruction from single street view images using deep learning. International Journal of Applied Earth Observation and Geoinformation, 112: 102859.

Pantoja-Rosero BG, Achanta R, Kozinski M, Fua P, Perez-Cruz F, Beyer K (2022) Generating LoD3 building models from structure-from-motion and semantic segmentation. Automation in Construction, 141: 104430.

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) PyTorch: An Imperative Style, High-Performance Deep Learning Library. arXiv preprint arXiv:1912.01703.

Persad RA, Armenakis C (2017) Automatic co-registration of 3D multi-sensor point clouds. ISPRS Journal of Photogrammetry and Remote Sensing, 130: 162–186.

Qi CR, Su H, Mo K, Guibas LJ (2017a) PointNet: Deep learning on point sets for 3D classification and segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), : 652–660.

Qi CR, Su H, Nießner M, Dai A, Yan M, Guibas LJ (2016) Volumetric and multi-view CNNs for object classification on 3D data. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), : 5648–5656.

Qi CR, Yi L, Su H, Guibas LJ (2017b) PointNet++: Deep hierarchical feature learning on point sets in a metric space. Advances in Neural Information Processing System (NeurIPS), 30.

Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing System (NeurIPS), 28.

Riegler G, Osman Ulusoy A, Geiger A (2017) OctNet: Learning deep 3D representations at high resolutions. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 3577–3586.

Riemenschneider H, Bódis-Szomorú A, Weissenberg J, Van Gool L (2014) Learning where to classify in multi-view semantic segmentation. European Conference on Computer Vision (ECCV), : 516–532.

Riemenschneider H, Krispel U, Thaller W, Donoser M, Havemann S, Fellner D, Bischof H (2012) Irregular lattices for complex shape grammar facade parsing. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), : 1640–1647.

Ripperda N (2010) Rekonstruktion von Fassadenstrukturen mittels formaler Grammatiken und Reversible Jump Markov Chain Monte Carlo Sampling. PhD thesis, Institut für Photogrammetrie und GeoInformation, Leibniz University Hannover, Hannover, Germany.

Roschlaub R, Batscheider J (2016) An INSPIRE-conform 3D building model of Bavaria using cadastre information, LiDAR and image matching. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B4: 747–754.

Roynard X, Deschaud JE, Goulette F (2018) Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. The International Journal of Robotics Research, 37 (6): 545–557.

Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: An efficient alternative to SIFT or SURF. International Conference on Computer Vision (ICCV).

Rusinkiewicz S, Levoy M (2001) Efficient variants of the ICP algorithm. In: Proceedings third international conference on 3-D digital imaging and modeling: 145–152.

Rusu RB, Cousins S (2011) 3D is here: Point Cloud Library (PCL). IEEE International Conference on Robotics and Automation (ICRA), : 1–4.

Schnabel R, Wahl R, Klein R (2007) Efficient RANSAC for point-cloud shape detection. Computer graphics forum, 26 (2): 214–226.

Schwab B, Beil C, Kolbe TH (2023a) r:trån: A road space model transformer library for OpenDRIVE, CityGML and beyond, version v1.3.0. https://doi.org/10.5281/zenodo.7702313. Accessed: 2023-08-20.

Schwab B, Haas Goschenhofer S, Wysocki O (2021) LoD3 Road Space Models, release v0.8.1. https://github.com/savenow/lod3-road-space-models. Accessed: 2023-01-30.

Schwab B, Kolbe TH (2019) Requirement analysis of 3D road space models for automated driving. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-4/W8: 99–106.

Schwab B, Wysocki O, Cabra RA, Kolbe TH (2023b) Generating synthetic point clouds of real cities for semantic road space segmentation. https://www.mdsi.tum.de/en/di-lab/vergangene-projekte/ss2023-tum-chair-of-geoinformatics-generating-synthetic-point-clouds-of-real-cities/. Accessed: 2023-06-11.

Schwarz S, Pilz T, Wysocki O, Hoegner L, Stilla U (2023) Transferring facade labels between point clouds with semantic octrees while considering change detection. Recent Advances in 3D Geoinformation Science (proceedings of 3D GeoInfo 2023), : 287–298. Cham: Springer.

Serna A, Marcotegui B, Goulette F, Deschaud JE (2014) Paris-rue-Madame database: As 3D mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. In: Proceedings of the International Conference on Pattern Recognition Applications and Methods. ACM, Angers, France, 6–8 March: 819–824.

Shafer G (1992) Dempster-Shafer theory. Encyclopedia of artificial intelligence, John Wiley, 1: 330–331.

Simonovsky M, Komodakis N (2017) Dynamic edge-conditioned filters in convolutional neural networks on graphs. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 3693–3702.

Special Interest Group 3D (2020) Modeling Guide for 3D Objects - Part 2: Modeling of Buildings (LoD1, LoD2, LoD3) - SIG3D Quality Wiki EN. https://en.wiki.quality.sig3d.org/. Accessed: 2023-01-30.

Stilla U, Xu Y (2023) Change detection of urban objects using 3D point clouds: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 197: 228–255.

Stritih A, Rabe SE, Robaina O, Grêt-Regamey A, Celio E (2020) An online platform for spatial and iterative modelling with Bayesian Networks. Environmental Modelling & Software, 127: 104658.

Su H, Maji S, Kalogerakis E, Learned-Miller E (2015) Multi-view convolutional neural networks for 3D shape recognition. IEEE/CVF International Conference on Computer Vision (ICCV), : 945–953.

Suveg I, Vosselman G (2000) 3D reconstruction of building models. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXIII: 538–545.

Szeliski R (2010) Computer vision: Algorithms and applications. Springer Science & Business Media.

Tan W, Qin N, Ma L, Li Y, Du J, Cai G, Yang K, Li J (2020) Toronto-3D: A large-scale mobile LiDAR dataset for semantic segmentation of urban roadways. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW): 202–203.

Tan Y, Wysocki O, Hoegner L, Stilla U (2023) Classifying point clouds at the facade-level using geometric features and deep learning networks. Recent Advances in 3D Geoinformation Science (proceedings of 3D GeoInfo 2023), : 391–404. Cham: Springer.

Thomas H, Qi CR, Deschaud JE, Marcotegui B, Goulette F, Guibas LJ (2019) KPConv: Flexible and deformable convolution for point clouds. IEEE/CVF International Conference on Computer Vision (ICCV), : 6411–6420.

Trimble Inc. (2021) SketchUp 3D Warehouse. https://3dwarehouse.sketchup.com. Accessed: 2021-10-30.

Tuttas S, Stilla U (2012) Reconstruction of rectangular windows in multi-looking oblique view ALS data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, I-3: 317–322.

Tuttas S, Stilla U (2013) Reconstruction of façades in point clouds from multi aspect oblique ALS. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W3: 91–96.

Tuttas S, Stilla U, Braun A, Borrmann A (2015) Validation of BIM components by photogrammetric point clouds for construction site monitoring. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W4: 231–237.

Tyleček R, Šára R (2013) Spatial pattern templates for recognition of objects with regular structure. In: German Conference on Pattern Recognition, Saarbrücken, Germany, September 3-6, 2013: 364–374.

Uggla M, Olsson P, Abdi B, Axelsson B, Calvert M, Christensen U, Gardevärn D, Hirsch G, Jeansson E, Kadric Z, Lord J, Loreman A, Persson A, Setterby O, Sjöberger M, Stewart P, Rudenå A, Ahlström A, Bauner M, Hartman K, Pantazatou K, Liu W, Fan H, Kong G, Li H, Harrie L (2023) Future Swedish 3D city models - specifications, test data, and evaluation. ISPRS International Journal of Geo-Information, 12 (2): 47.

U.S. Department of Transportation (2014) Mitigation Strategies For Design Exceptions: Vertical Clearance. US Department of Transportation Federal Highway Administration. https://safety.fhwa.dot.gov/geometric/pubs/mitigationstrategies/chapter3/3_verticalclearance.cfm. Accessed: 2021-03-26.

Vallet B, Brédif M, Serna A, Marcotegui B, Paparoditis N (2015) TerraMobilita/iQmulus urban point cloud analysis benchmark. Computers & Graphics, 49: 126–133.

Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2017) Graph attention networks. arXiv preprint arXiv:1710.10903.

Vermessungsverwaltung B (2023) 3D-Gebäudemodelle (LoD2). https://geodaten.bayern.de/opengeodata/OpenDataDetail.html?pn=lod2. Accessed: 2023-01-30.

Wang L, Hu H, Shang Q, Xu B, Zhu Q (2023) StructuredMesh: 3D structured optimization of facade components on photogrammetric mesh models using binary integer programming. arXiv preprint arXiv:2306.04184.

Wang Y, Sun Y, Liu Z, Sarma SE, Bronstein MM, Solomon JM (2019) Dynamic graph CNN for learning on point clouds. ACM Transactions on Graphics (tog), 38 (5): 1–12.

Weinmann M, Jutzi B, Mallet C (2013) Feature relevance assessment for the semantic interpretation of 3D point cloud data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-5/W2: 313–318.

Wilbers D, Merfels C, Stachniss C (2019) Localization with sliding window factor graphs on third-party maps for automated driving. International Conference on Robotics and Automation (ICRA), : 5951–5957.

Willenborg B, Pültz M, Kolbe TH (2018a) Integration of semantic 3D city models and 3D mesh models for accuracy improvements of solar potential analyses. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W10: 223–230.

Willenborg B, Sindram M, Kolbe TH (2016) Semantic 3D city models serving as information hub for 3D field based simulations. Lösungen für eine Welt im Wandel, DGPF Jahrestagung, : 54–65.

Willenborg B, Sindram M, Kolbe TH (2018b) Applications of 3D city models for a better understanding of the built environment. Trends in Spatial Analysis and Modelling: Decision-Support and Planning Strategies, : 167–191. Springer Verlag.

Wu J, Zhang C, Xue T, Freeman B, Tenenbaum J (2016) Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. Advances in Neural Information Processing System (NeurIPS), 29.

Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, Xiao J (2015) 3D ShapeNets: A deep representation for volumetric shapes. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), : 1912–1920.

Wysocki O, Grilli E, Hoegner L, Stilla U (2022a) Combining visibility analysis and deep learning for refinement of semantic 3D building models by conflict classification. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-4/W2-2022: 289–296.

Wysocki O, Hoegner L, Stilla U (2022b) Refinement of semantic 3D building models by reconstructing underpasses from MLS point clouds. International Journal of Applied Earth Observation and Geoinformation, 111: 102841.

Wysocki O, Hoegner L, Stilla U (2022c) TUM-FAÇADE: Reviewing and enriching point cloud benchmarks for façade segmentation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-2/W1-2022: 529–536.

Wysocki O, Hoegner L, Stilla U (2023a) MLS2LoD3: Refining low LoDs building models with MLS point clouds to reconstruct semantic LoD3 building models. Recent Advances in 3D Geoinformation Science (proceedings of 3D GeoInfo 2023), : 367–380. Cham: Springer.

Wysocki O, Schwab B, Beil C, Holst C, Kolbe TH (2024a) Reviewing Open Data Semantic 3D City Models to Develop Novel 3D Reconstruction Methods. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48: 493–500.

Wysocki O, Schwab B, Hoegner L, Kolbe T, Stilla U (2021a) Plastic surgery for 3D city models: A pipeline for automatic geometry refinement and semantic enrichment. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, V-4: 17–24.

Wysocki O, Schwab B, Willenborg B (2022d) GitHub repository: `https://github.com/OloOcki/awesome-citygml`. GitHub. Accessed: 2024-01-30.

Wysocki O, Tan Y, Froech T, Xia Y, Wysocki M, Hoegner L, Cremers D, Holst C (2024b) ZAHA: Introducing the Level of Facade Generalization and the Large-Scale Point Cloud Facade Semantic Segmentation Benchmark Dataset. Accepted to IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2025).

Wysocki O, Xia Y, Wysocki M, Grilli E, Hoegner L, Cremers D, Stilla U (2023b) Scan2LoD3: Reconstructing semantic 3D building models at LoD3 using ray casting and Bayesian networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), : 6547–6557.

Wysocki O, Xu Y, Stilla U (2021b) Unlocking point cloud potential: Fusing MLS point clouds with semantic 3D building models while considering uncertainty. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, VIII-4/W2: 45–52.

Wysocki O, Zhang J, Stilla U (2021c) TUM-FAÇADE: A database of annotated façade point clouds. https://mediatum.ub.tum.de/1636761?v=2. Accessed: 2023-01-08.

Wyvill G, Kunii TL (1985) A functional model for constructive solid geometry. The Visual Computer, 1 (1): 3–14.

Xiao W, Vallet B, Brédif M, Paparoditis N (2015) Street environment change detection from mobile laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote Sensing, 107: 38–49.

Xu Y, Stilla U (2021) Towards building and civil infrastructure reconstruction from point clouds: A review on data and key techniques. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14: 2857–2885.

Xu Y, Tong X, Stilla U (2021) Voxel-based representation of 3D point clouds: Methods, applications, and its potential use in the construction industry. Automation in Construction, 126: 103675.

Yi L, Su H, Guo X, Guibas LJ (2017) SyncSpecCNN: Synchronized spectral CNN for 3D shape segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), : 2282–2290.

Zeibak R, Filin S (2008) Change detection via terrestrial laser scanning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVI-3/W52: 430–435.

Zhang Q, Wysocki O, Urban S, Jutzi B (2024) CDGS: Confidence-Aware Depth Regularization for 3D Gaussian Splatting. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48: 189–196.

Zhang S, Lo S, Chen YH, Walter T, Enge P (2018) GNSS multipath detection in urban environment using 3D building model. IEEE/ION Position, Location and Navigation Symposium (PLANS), : 1053–1058.

Zhang X, Lippoldt F, Chen K, Johan H, Erdt M, Zhang X, Lippoldt F, Chen K, Johan H, Erdt M (2019) A data-driven approach for adding facade details to textured LOD2 CityGML models. International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), : 294–301.

Zhao H, Jiang L, Jia J, Torr PH, Koltun V (2021) Point transformer. IEEE/CVF International Conference on Computer Vision (ICCV), : 16259–16268.

Zhou QY, Park J, Koltun V (2018) Open3D: A modern library for 3D data processing. arXiv preprint arXiv:1801.09847.

Zhu J, Gehrung J, Huang R, Borgmann B, Sun Z, Hoegner L, Hebel M, Xu Y, Stilla U (2020) TUM-MLS-2016: An annotated mobile LiDAR dataset of the TUM City Campus for semantic point cloud interpretation in urban areas. Remote Sensing, 12 (11): 1875.

Zhu J, Wysocki O, Holst C, Kolbe TH (2024) Enriching Thermal Point Clouds of Buildings using Semantic 3D building Models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 10: 341–348.

Zlatanova S (2000) 3D GIS for urban development. PhD thesis, Institute for Computer Graphics and Vision, Technical University of Graz, Graz, Austria.

Acknowledgment

I would like to express my sincere gratitude to the following individuals and organizations; this three-year doctoral journey would not have been possible without their support, guidance, and encouragement.

I would start with my primary supervisor, mentor, and a friend who always supported me in the dissertation and beyond: The late Prof. Dr.-Ing. Uwe Stilla. One of the primary reasons I embarked on the doctoral journey was Uwe's Photogrammetry Selected Topics seminar, where, still as a master's student, I was fascinated by his passion for science and photogrammetry. Afterward, pursuing a PhD under his guidance was an obvious choice. His expert guidance, thought-provoking discussions, and frequent long-hours-meetings have immensely impacted my personal and professional development. As you once sent me, citing The Little Prince of Antoine de Saint-Exupéry: "When saying goodbye, the fox shares his secret with the Little Prince: It is only with the heart that one can see rightly; what is essential is invisible to the eye."

I extend my appreciation to my committee members and supervisors, Univ.-Prof. Dr. rer. nat. Thomas H. Kolbe, Prof. Dr.-Ing. habil. Ludwig Hoegner, and Prof. Dr. Sisi Zlatanova for their constructive feedback and thoughtful suggestions that significantly enhanced the quality of this dissertation. I would also like to thank Univ.-Prof. Dr.-Ing. Christoph Holst for his support and voluntary chairmanship of the committee. Also, I am indebted to Univ.-Prof. Dr. rer. nat. Thomas H. Kolbe, thank you for agreeing to become my primary supervisor in the last months of my PhD journey after Prof. Dr. Ing. Uwe Stilla passed away.

I am grateful to my colleagues and peers for their intellectual contributions, stimulating discussions, and the sense of community that made this academic pursuit a collaborative and enriching experience. I believe that a PhD journey is nothing without a team and constructive peer-level feedback. It is impossible to mention all the peers who have influenced my PhD work. To all of you who crossed paths with me and were keen to share your views, big thank you! Nevertheless, I would like to express special gratitude to three individuals, M.Sc. M.Sc. Benedikt Schwab, Dr. Yan Xia, and Dr. Eleonora Grilli, who have become my friends, great inspiration, and support, always pushing me forward.

However, the scientific team comprises not only direct peers but also students. I have enjoyed supervising you, and I have learned a great deal from you, my Mentees. Here, again, it is impossible to name all the students that I crossed paths with. Yet, I would like to thank Yue Tan, Sophia Schwarz, Tanja Pilz, and Thomas Fröch, whose implementations and scientific contributions enriched this dissertation! I am sure that our paths will cross again soon.

I acknowledge the Technical University of Munich and 3D Mapping Solutions Gmbh for providing the necessary resources, facilities, and a conducive academic environment to successfully complete this dissertation.

To the most important people in my life who constantly support me: To my parents, Jolanta and Jarosław, and my sister Julia, your unwavering belief in me and your constant encouragement sustained me during the challenging moments of this journey. Your love and understanding have been my pillars of strength. A special place in my heart is reserved for my wife, Magdalena, who always selflessly supports me both mentally and scientifically. Your contributions to this work are implicitly and even explicitly, by joint publication, present.

A heartfelt thank you to my friends who provided support, encouragement, and occasional distractions during the ups and downs of this academic endeavor. Special thanks go to Bela and Remik; your years-long camaraderie made the journey more enjoyable.

Finally, to everyone who played a role, no matter how small, in this academic journey, I offer my heartfelt thanks. Your contributions have left an indelible mark on this dissertation; I am truly grateful for that.