Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Civil and Environmental Engineering

Øyvind Håbrekke and Fredrik Dokka Nordstad

Estimating the Height of Facades with Street-level Imagery using Facade Parsing, Floor Segmentation, and Urban Rules

Master's thesis in Engineering and ICT
Supervisor: Hongchao Fan
June 2021

NTNU
Norwegian University of Science and Technology

Øyvind Håbrekke and Fredrik Dokka Nordstad

Estimating the Height of Facades with Street-level Imagery using Facade Parsing, Floor Segmentation, and Urban Rules

Master's thesis in Engineering and ICT
Supervisor: Hongchao Fan
June 2021

Norwegian University of Science and Technology
Faculty of Engineering
Department of Civil and Environmental Engineering

NTNU

Kunnskap for en bedre verden

NTNU
Kunnskap for en bedre verden

Department of Civil and Environmental Engineering

TBA4925 - Geomatics, Master's Thesis

Estimating the Height of Facades with Street-level Imagery using Facade Parsing, Floor Segmentation, and Urban Rules

Authors:

Øyvind Håbrekke

Fredrik Dokka Nordstad

Supervisor:

Prof. Dr. Hongchao Fan

June, 2021

NTNU
Faculty of Engineering Science and Technology
Department of Civil and Environmental Engineering

Master thesis

(TBA4925 - Geomatics, Master thesis)

Spring 2021

for

Øyvind Håbrekke and Fredrik Dokka Nordstad

Estimating the Height of Facades with Street-level Imagery using Facade Parsing, Floor Segmentation, and Urban Rules

BACKGROUND

Facade height is a key variable in studying the character and scale of urban environments. Accurately estimating the height requires high spatial accuracy and complete building data, both dependent on the utilization of expensive and advanced state-of-the-art methods.

TASK DESCRIPTION

The goal of the assignment is to investigate how street-level imagery can be used to estimate the facade height in a wider geographical region, by exploiting architectural principles including symmetry and repetitive patterns.

Specific tasks:

- Develop a pipeline for estimating façade heights
- Select experimental study area and test our method
- Evaluate the results

ADMINISTRATIVE/GUIDANCE

The work on the Master Thesis starts on January 15th, 2021.

The thesis report as described above shall be submitted digitally in INSPERA at the latest at June 11th, 2021.

Supervisors at NTNU and professor in charge:
Hongchao Fan

Trondheim, June, 2021

Abstract

Facade height is a key variable in studying the character and scale of urban environments. However, accurately estimating the height requires high spatial accuracy and complete building data, dependent on expensive and advanced state-of-the-art methods. This study aims to lower the threshold for large-scale height estimation by using more accessible technology. Specifically, we investigate how we can use street-level imagery to estimate the facade height in a wider geographical region by exploiting architectural principles, including symmetry and repetitive patterns. In addition, we implement a method that automatically segments the facades into separate floors and use extensive knowledge of their inherent features and attributes to estimate the facade height.

To test our pipeline, we conducted an experimental study on street view imagery in a contained geographical area of Trondheim, Norway. We automatically detected facade objects to segment the floors with a RANSAC regressor and then applied a set of defined urban rules to adjust the resulting height further. The results indicated that segmenting the floors contributed to an accurate estimation of the facade height and that the rules aided in adjusting the height estimation. We also discovered that the quality of street view imagery significantly influenced the results. Finally, to evaluate the method, we considered an optimal subset of imagery and found that the correctness of the floor segmentation was 92%. Furthermore, we achieved adequate results regarding the height estimation in the whole study area, with progressively larger errors as the building height increased.

Sammendrag

Fasadehøyde er en vesentlig faktor i studiet av bymiljøets karakter og omfang. Estimering av høyde krever nøyaktige romlige målinger og komplett bygningsdata, der begge avhenger av bruken av dyr og toppmoderne teknologi. Denne studien har som mål å senke terskelen for å gjennomføre storskala høydeestimering ved å benytte seg av lett tilgjengelig teknologi. Hovedsakelig ønsker vi å undersøke hvordan bilder tatt fra gatenivå kan benyttes for å estimere fasadehøyden i et større geografisk område. Dette gjennomføres ved å utnytte symmetri og repeterende mønstre basert på arkitektoniske prinsipper. Vi implementerer en metode som automatisk segmenterer fasadene i separate etasjer og bruker kunnskap om deres iboende egenskaper og attributter for å estimere fasadehøyden.

Et eksperimentell studie ble gjennomført i den hensikt for å teste metoden på bilder tatt fra gatenivå i Trondheim, Norge. Fasadeobjektene ble automatisk detektert for å videre segmentere etasjene med et regresjonsanalyseverktøy (RANSAC) og deretter ved å anvende våre definerte urbane regler for å ytterligere justere den resulterende høyden. Resultatene indikerte at segmentering av etasjer bidro til en nøyaktig estimering av fasadehøyden og at reglene supplerte med å styrke høydestimasjonen. Det ble videre gjort kjent at kvaliteten på bildene fra gatenivå i stor grad påvirket resultatene. Evaluering av metoden resulterte i en korrekthet på 92% for eteasjesegmentering på et utvalg bilder med god kvailtet. Videre ble det oppnådd tilfredsstillende resultater for høydestimasjonen i hele studieområdet, med gradvis større feil etter hvert som byggehøyden økte.

Preface

This paper is a master thesis written for the Department of Civil and Environmental Engineering at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. It is a part of the study program Engineering and ICT - Geomatics and was written in the spring of 2021.

We want to thank our supervisor, Hongchao Fan, for his assistance and guidance during the writing of this thesis. We would also like to thank our coworkers, fellow students, and family for their input and support.

Trondheim, June 2021

Øyvind Håbrekke
Fredrik Dokka Nordstad

Contents

List of Figures

List of Tables

1 Introduction

This chapter will present the motivation, existing state-of-the-art methods, introduce the problem and our solution. Additionally, the research objectives are defined, and an outline of the thesis is presented at the end.

1.1 Motivation

The global human population distribution has shifted from rural towards urban settlements in the last two centuries, with more than 50% of people being urban as of 2021. Moreover, it is predicted to be steadily increasing over the next decades, estimated that approximately 69% of the world's population would be living in cities by 2050 (UNDP 2016). Furthermore, the rapid growth of cities significantly impacts the socio-economic processes and has substantial environmental effects, posing a considerable challenge to sustainable urban development. In turn, this gives rise to the need of having up-to-date and consistent data on the characteristics of the urban environment and its morphology, where the building height is considered one of the key geometric parameters for understanding urban process regimes (Frantz et al. 2021).

Earth has already been comprehensively mapped in 2D, yet, the vertical dimension remains untapped of its limitless potential. Therefore, mapping inhabited areas as a 3D representation of reality require a description of the vertical dimension. This will enable the description of building height and facade extent in urban environments. Furthermore, as the parameters directly influence many quantities and relations, they enable the description of the floor space and urban morphology (Esch et al. 2020).

The importance of building height as a parameter when regarding settlement characteristics is undeniably significant. As such, the description and management of the height contribute to enabling the reconstruction of 3D city models (Biljecki et al. 2015), that in turn can be utilized to enable detailed analysis of the energy and

environmental effects, including the estimation of renewable energy potential and greenhouse gas emissions (Resch et al. 2016; Borck 2016), in addition to the management of smart cities and accurate interpretation of the population distribution (Gong et al. 2011). Moreover, extensive knowledge of heights in an urban environment can contribute significantly to the planning and expansion of infrastructures such as electricity systems, telecommunications, and water systems. In particular, this is beneficial for developing countries, as stated by Duncan 2012, suggesting that the adoption of 3D city modelling can improve the quality of life. However, the reconstruction of 3D city models is a demanding and expensive task requiring extensive knowledge and readily available geospatial data, severely limiting the feasibility of arriving at an adequate solution. As of now, few open-sourced alternatives have the spatial accuracy required for the particular task of estimating the height and producing complete 3D city models.

1.2 State-of-the-art

The height of buildings can be estimated utilizing a vast array of geo-information tools and remote sensing approaches. In the following, a selection of state-of-the-art methods will be presented, such as high-definition surveying, remote sensing, and aerial photogrammetry. We will also describe the shortcomings of existing methods, both in terms of a technical and economic aspect.

The use of LiDAR to detect the height of buildings when considering the vertical dimension can be accomplished through large-scale multi-source data analytic procedures by exploiting earth observation (OE) satellite data, such as the Sentinel-2, with the use of digital terrain models (DTM) and normalized digital surface models (nDSM) from the TanDEM-X digital elevation, as well as Open Street Map (OSM) data and the Global Urban Footprint (GUF). This enables the generation of spatially detailed maps of 3D building structures at a continental or global scale. The subsequent quality of the output of this method is sufficiently accurate for describing the urban morphology at a city level. Yet, for the precise estimation of individual building heights, the results are still, in some cases, insubstantial (Esch et al. 2020).

While some methods rely on the complex use of remote sensing (LiDAR) with DTM and DSM, various researchers have been exploring a more elementary approach using celestial geometry and remote sensing satellite images to estimate building heights. More specifically, various approaches in regards to the relationship between a building and its coincident shadows have been explored, e.g. Comber et al. 2012 and Qi et al. 2016, where the latter applied their method on images acquired from Google Earth. As such methods yields promising results, obtaining detailed features from satellite images, such as shadows cast by buildings in an urban scene, provide challenges. In particular, the presence of intensity heterogeneity and feature complexity complicates the shadow detection process, that is emphasised in the research done by Liasis and Stavrou 2016. A more compound shadow-overlapping algorithm has been developed by Kadhim and Mourshed 2018 with the incorporation of identifying building shadow regions on very high resolution (VHR) satellite images with the use of solar information gathered from image metadata, together with the application of morphological operations and the Jaccard similarity coefficient, in turn enabling a measure of similarity between the sets of data.

Photogrammetric analysis of satellite, aerial, or drone imagery and ALS-based height estimations are generally more accurate when compared to the aforementioned space-borne methods due to increasing errors as the images are captured in higher altitudes (Sirmacek et al. 2012; Baltsavias 1999). Moreover, they provide a finer spatial resolution, increasing the accuracy of height estimations on individual buildings, especially when paired with official cadasters or open-sourced building footprints (Frantz et al. 2021). On the other hand, these methods provide results that have a lesser spatial extent, as they are not covering large areas such as complete countries or regions.

The existing methods present a set of technical and economic problems, as implementing them requires extensive expert knowledge, many work hours, and accessibility to exceptional technology and other high-cost equipment. Another problem is that the datasets are still proprietary and come with considerable data purchase costs. In addition, the resulting data output from these approaches is of varying quality, as large-scale height estimation includes systematic errors and overlapping

footprints when considering the spatial resolution and distinction of individual buildings (Frantz et al. 2021). Evidently, airborne laser scanning (ALS), photogrammetry, airborne or space-borne VHR imaging all face the challenges of continuity and regional inconsistency. Furthermore, the utilization of official cadastres or volunteered geographic information (VGI) with open data alternatives to georeference the resulting height estimations to specific building footprints may cause problems concerning spatial accuracy and completeness (Brovelli and Zamboni 2018).

1.3 Objective and the Proposed Solution

This method aims to enable large-scale building height estimation in urban environments using street-level imagery to segment floors and consider urban rules to enhance the facade height estimation as a cheap alternative to the already existing state-of-the-art methods available.

We propose a solution that entails exploiting facade patterns by acquiring knowledge of the facade objects and their relative positioning, enabling an overview of the inherent structure that we can utilize when performing the floor segmentation. We seek to estimate the facade height by applying a standardized metric for individual floors and improving the estimation using urban rules that consider various architectural principles.

The work conducted in this thesis is based on the premise of the following three research questions:

Q1: Will the number of floors enable an adequate estimation of the facade height by using floor segmentation on street-level imagery?

Q2: Will extensive knowledge about the features of a facade aid in the estimation of the facade height?

Q3: Will extensive knowledge about the attributes of a building aid in the estimation of the facade height?

1.4 Thesis Outline

In Chapter 2 we introduce the background knowledge and the fundamental concepts that guided our work. In particular, the architectural principles and the underlying symmetry inherent in building facades are unveiled and coupled to the urban landscape as a whole. Furthermore, we will present work done within facade parsing and construction of grammar rules to form a theoretical framework for the rest of the thesis.

Then, in Chapter 3 the structure of the method will be described in detail. At first, we explain the pre-processing step, including the management of building information data and street view imagery. Then, to describe the floors on facades, we explain the implementation of our floor segmentation technique. Next, we present a set of urban rules that are exerted to quantify an estimation of the real-world counterpart. In the end, we generate a 3D model to visualize the final result of our method.

Chapter 4 begins with a brief introduction of our experimental study, environment, and study area. The following experimental results are then presented, and a two-part evaluation is carried out. The final section of the chapter summarizes our key findings, entailing a discussion that interprets the results and uncovers the limitations of our method.

The thesis ends in Chapter 5 with concluding remarks answering the research objectives and summarizing the findings of our work. Finally, some suggestions for future work are presented.

2 Background Knowledge and Related work

This chapter will introduce background knowledge and related research and its relevance to our method. Hence, a description of the architectural principles applicable for laying the foundation for recognising symmetries and patterns is presented. Moreover, we will also introduce the work related to the process of logically structuring buildings, recognizing their inherent features, including objects and patterns, and subsequently modelling their facades as a whole.

2.1 Architectural Principles, Symmetry, and Patterns

The most prominent pattern within architectural principles is the use of combined symmetries. Human perception depends on combined symmetries to reduce information overload, as disorganized information is much harder to process for the human (Figure 1). Such patterns of complex symmetries and mathematical relations are often found in human creations, where architects and structural engineers cooperate to ensure that mathematical principles are followed to construct a functional building (Salingaros 2020).

Mathematics also impacts the aesthetic footprint of constructions, as proportional ratios often are used to determine relative dimensions of architectural components (Kappraff 2016). Relying on these observations, a generalization of the structure of a building, or more specifically a building facade, can be done by exploiting the mathematical relations found in the existing environment. Both temporal changes and cultural influences within the field of architecture will provide minor and major distinctions in the facade structure, particularly modern architecture known for deviating from the traditionally applied architectural principle (Nia and Rahbarianyazd 2020). However, the underlying composition of buildings usually follows the same mathematical principles and arrangements of the facade.

Such principles and patterns could also be applied at a larger scale, as the urban space is a complex collection of buildings, blocks, and neighbourhoods separated by

a structured road network (Vanegas et al. 2010). Moreover, the underlying structure of a city is defined by a large set of compound variables such as land policies and regulations set by the local government, which again affects the individual features of a building. For instance, most city areas apply certain building regulations to ensure that national or local legal framework are met, e.g. requiring at least one operable window with a clear view in each open living area as stated in Direktoratet for Byggkvalitet 2017 (§13-4) and placed with a minimum height above the floor as stated in Direktoratet for Byggkvalitet 2017 (§12-17). Urban areas elucidate the outcome of this, as limited space between buildings causes a more dense distribution of windows on the front and rear facade to fulfil the aforementioned regulations.

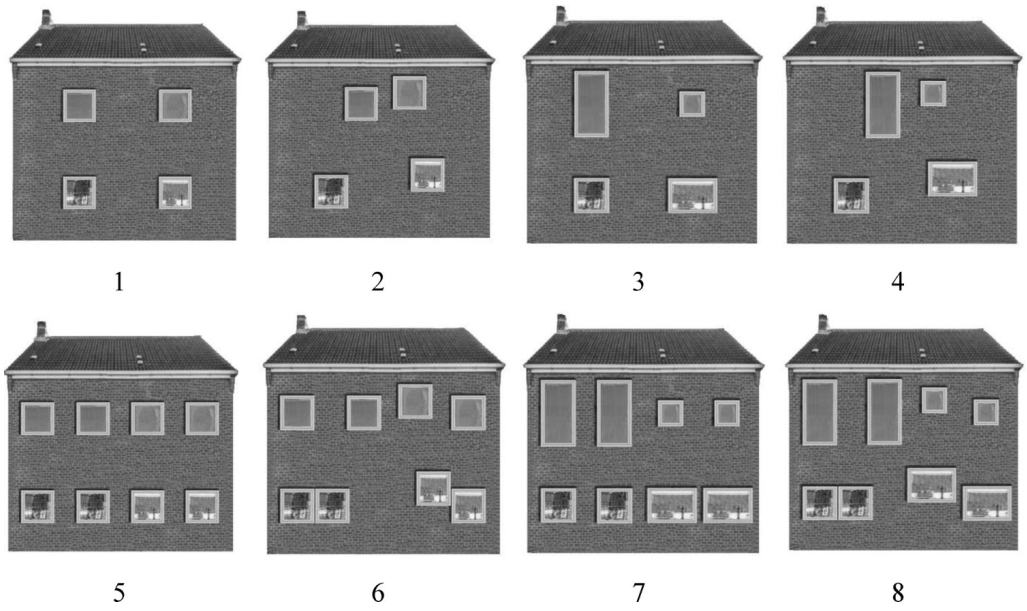


Figure 1: Research have shown that there is a strong positive correlation between symmetry and aesthetic appreciation, where symmetry along the vertical axis is usually perceived as more dominant compared to horizontal-symmetry (Aydin and Mirzaei 2021).

The urban landscape, in general, has been significantly influenced by large-scale urbanization, where limited space of land area has lead to a vertical expansion of the cities worldwide and a centralized population density (Ding 2013). Furthermore, a city is usually intentionally designed by urban planners that have applied urban morphology elements to promote sustainable urban development, dividing the city into a hierarchical network (Chen 2014; Li et al. 2004). I.e. offices, malls, and

other commercial buildings are often located in the city centre or other high-density areas, implying that each division of the city may have its own city image following certain characteristics such as a common height, building size, and facade pattern. This follows the concept of *building typology*, which states that the set of buildings serving the same function, usually share the same structural features Kelbaugh 1996.

The building typology also applies to sets of blocks and neighbourhoods, providing coherence and shared meaning in the built environment. Therefore, interpreting the area in context may give good indications on a general idea of the architectural principle applied among the vast selection of diverse facade structure and patterns current existing in the physical world today. In addition, other concepts such as wholeness and harmony, as defined by Jiang 2016, could further strengthen the idea of similarities among facade patterns and within neighbourhoods. Furthermore, as a city is a concentrated reflection of the city culture, certain local features may further shape the architectural principles and establishing a convincing theme of the urban environment (Qiao, Yiqing 2017).

2.2 Floor Segmentation and Urban Rules

In general, the procedure of modelling the facade, building, and architecture have been explored extensively throughout the last decades. However, these methods focus solely on generating and reconstructing accurate 3D models. On the contrary, our method aims to estimate the facade height by exploiting structural arrangement and repetition of facade objects. The following related work will therefore cover methods utilizing the same principles through the implementation of rules.

Becker and Haala 2009 proposed an automatic approach where grammar rules are generated from observed 3D facade geometries and further used to create synthetic facade structures for unknown building parts. Applying architectural principles such as the column-wise arrangement of facade objects and their interrelationships, grammar rules were derived and applied to verify and generate 3D models.

The parsing of building facades is crucial in semantically structuring buildings, facades, and 3D city model reconstruction and is prevalent within the domain of computer vision. In many cases, urban structures follow ordering principles and characteristics of symmetry that can be exploited to describe urban morphology and the inherent semantics accurately. Initially, the methods primarily depended on prior knowledge and grammar rules proposed by human experts, including parsing on weak architectural principles (Becker and Haala 2009).

However, newer methods have become better at reducing these challenges by basing the parsing on shape grammar rules. Yet, prior knowledge is still constraining the methods, making them harder to generalize and culminating in erroneous output, such as irregular arrangements of facade objects and perspective distortion will skew the results. With the application of deep learning, Schmitz and Mayer 2016 approached facade parsing as an image segmentation problem which yielded promising results. Still, the network suffers from the lack of implementation of man-made rules found on facades. Further on, image segmentation poses several challenges with the great variations of environments and the occurrence of occlusions, visual perspective, and changes in illumination (Liu et al. 2017).

Kong and Fan 2020 approached the problem by proposing a new pipeline based on convolutional neural networks (CNNs), connecting semantic segmentation and object detection by utilizing PSPNet and YOLO (Redmon and Farhadi 2018) for parsing facade images. Semantic segmentation is carried out for parsing walls, and object detection parses the following facade elements. The method is tested on scenes with non-optimal conditions by including images with foreground occlusion, varying illumination conditions, and complex backgrounds. The results from the pipeline are good, proving the potential for the method as a general facade parser dealing with complex scenes. The method can also be expanded upon by including more grammar and context rules to improve their facade parsing subnetwork. Furthermore, the pipeline can be applied to practical problems such as 3D model reconstruction, facade height estimation, and renewable energy potential analysis.

With most 3D buildings being reconstructed automatically or semi-automatically

from LiDAR and image data, Fan et al. 2021 proposed an interactive approach for VGI 3D buildings modelling and semantic labelling on images. The outline of the facade was marked manually through the developed software, yielding high accuracy LoD3-level 3D building models with no preliminary knowledge required by the users. Despite promising results on a smaller scale, manual intervention is a still necessary input, which will be laborious for reconstructing city models for large-scale analysis.

The architectural principles and related work presented in this chapter form the theoretical framework for the rest of the thesis. Considering the patterns inherent in facades enables logical structuring and modelling, generating accurate and semantically rich 3D models to represent the urban scene. However, several challenges arise when parsing facades as the diversity of the urban scene and complexity of facade structures complicate extracting the relevant features, requiring high-quality imagery and a comprehensive dictionary of grammar rules. Furthermore, as previous work has shown, exploiting simple patterns and repetition of objects is enough to describe the vast majority of existing facades. Yet, the height has not been properly described by the current methods, even though it is a highly prominent feature. Therefore, our work aims to describe the vertical dimension, as made possible by exploiting the same principles and ideas, to automatically generate a digital twin of the urban environment, emphasising the third dimension.

3 Floor Segmentation and Rule-based Height Estimation

In this chapter, we present our method. Initially, we lay out our approach to preprocess the necessary data to segment floors on facades from street view imagery. Then, we derive a set of urban rules from facade patterns to estimate the facade heights.

3.1 Overview

The method is based on analyzing simple architectural features and patterns where the number of detected floors clearly indicates the facade height and structure. A RANSAC regressor was utilized to achieve the aforementioned segmentation, from which the resulting output is highly dependent on the quality of the acquired street view imagery and building data. Furthermore, a set of urban rules are deployed to exploit the patterns and structure of facades and their inherent objects to provide an estimation of the facade height. Finally, a 3D model visualizes the height estimations.

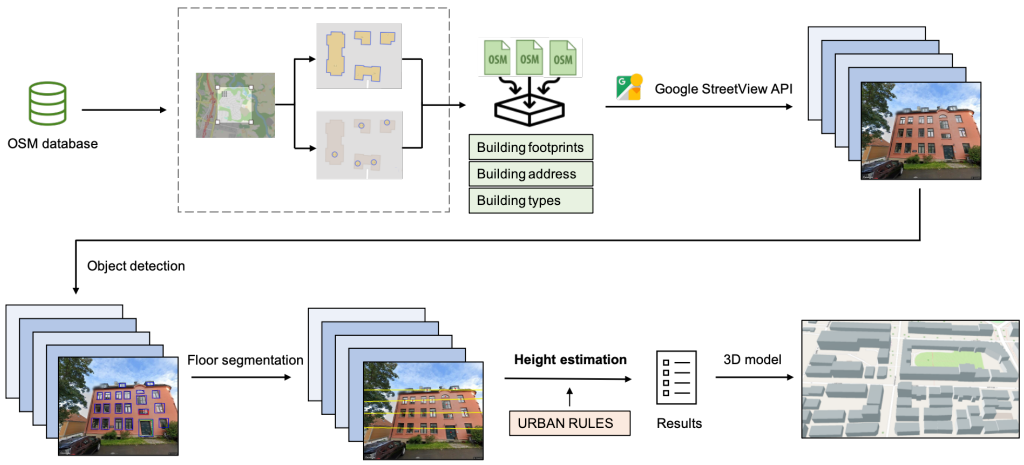


Figure 2: This flowchart visualizes the main steps of our method.

3.2 Pre-processing

The initial step of our method covers the task of collecting data and preparing it for further utilization. It entails collecting and manipulating the input data, arranging it so that the continuous integration and management become uniform. We used several data sources to retrieve the necessary data combined for the ensuing floor segmentation and height estimation. The method requires input that handles building information, particularly the footprints and building types, and the corresponding street-level imagery of the facades to connect the results with a real-world counterpart.

3.2.1 Data Acquisition

When choosing the building information source, the consistency of the data and the spatial accuracy were central properties to consider. In addition, the building information must be able to aid in georeferencing the facades and associate the digital image files with locations in physical space. In this regard, we chose a data source that could provide viable information befitting our requirements; OpenStreetMap (OSM). It is a widely acknowledged collaborative project with high data quality and appears to be consistently preferred over other open sources of retrieving spatial data (Mooney and Minghini 2017).

Initially, the bounding coordinates of the study area were given as an input in a query performed on the OSM database to gather all building footprints. In the matter of terminology evolving the elements of OSM, a *Way* is defined as a linear feature on the ground (e.g. road, wall, or river) and consists of an ordered list of nodes which normally contains at least one tag to describe its features, such as the building type as seen in Table 1 (OpenStreetMap 2021b). An additional retrieval of building footprints, including enclaves and/or multiple exclaves, was necessary, as such buildings are instead defined in the OSM data structure as multi polygon *Relations*.

As of yet, there exists no connection between the building *Node* and the *Way* describing the relation between the footprint and address node in OSM (Figure 3), and acquiring the address of the building footprint is necessary, as it facilitates for the subsequent process of connecting the fetched street-level images to its corresponding building data. Therefore, we queried the OSM data and checked whether the footprint polygon contained an address node to connect the address and the building footprint itself. Furthermore, we also handled the filtering and exclusion of undesirable building data. Consequently, whence building annexes (e.g. garages and sheds) were removed from the building data as they contained no address node. Finally, after connecting the footprints and address nodes, the data was extracted and combined into a joint JSON file.

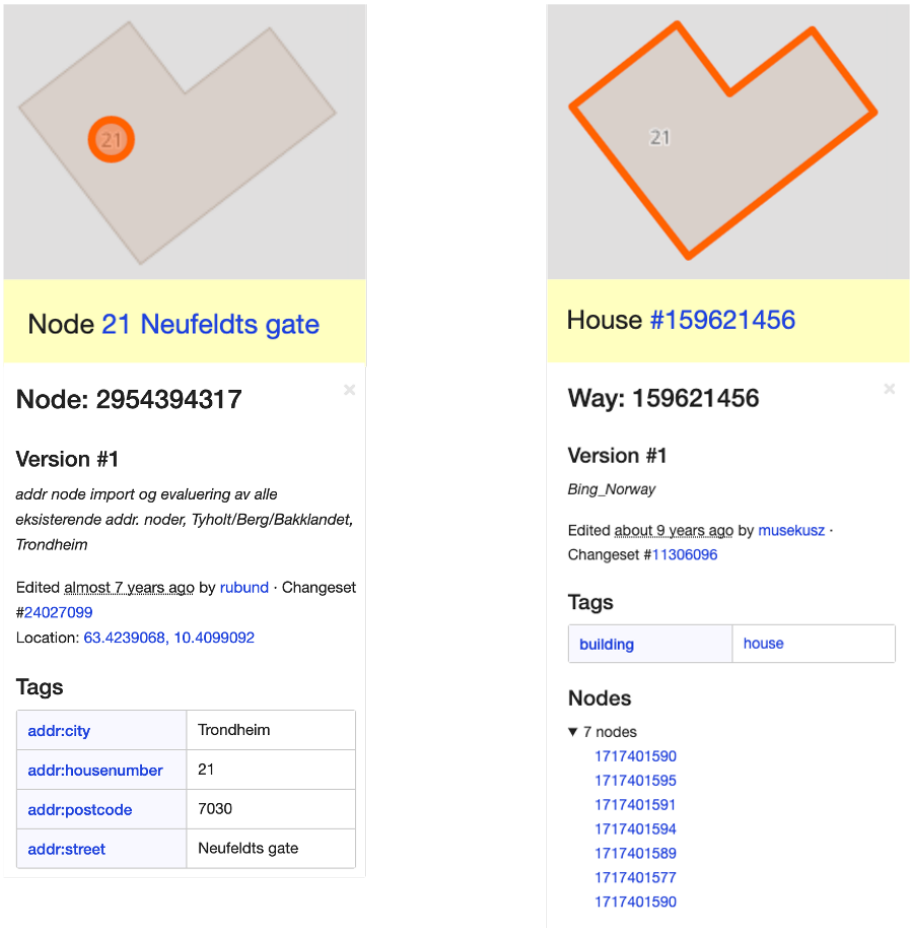


Figure 3: This figure illustrates how OpenStreetMap separates between *nodes* and *ways* (houses). Each node contains a location with both coordinates and an address (Source: OpenStreetMap 2021b).

Table 1: OpenStreetMap JSON elements.

Type	Common Attributes	Tags	Description
Node	id, lat, lon	addr:city, addr:postcode, addr:street, addr:housenumber	A node represents a specific point on the earth's surface defined by its latitude and longitude.
Way	id, bounds, geometry	building, amenity	A way is an ordered list of nodes used to represent linear features and can be a closed boundary such as a building footprint.
Relation	id, bounds, members, role, type, geometry	building, amenity	A relation is a multi-purpose data structure that describes relationships between elements (nodes, ways, or other relations).

Next, when acquiring the facade imagery, the images were desired to be aligned perpendicularly with the considered facade (Figure 4). This enables desirable behaviour for both the detection method and subsequent height estimation, as an image would be less likely to produce erroneous features. However, the existence of data that could guarantee this for a given study area is close to non-existing and requires extensive manual data acquisition. Nevertheless, manually capturing the street view imagery would indeed result in superior facade height estimates. Thus, balancing manual and automatic data collection from the street view imagery providers (s.a. Google Street View and Mapillary) essentially boils down to a trade-off between data with low availability that is more facilitated and data with high availability is less facilitated.

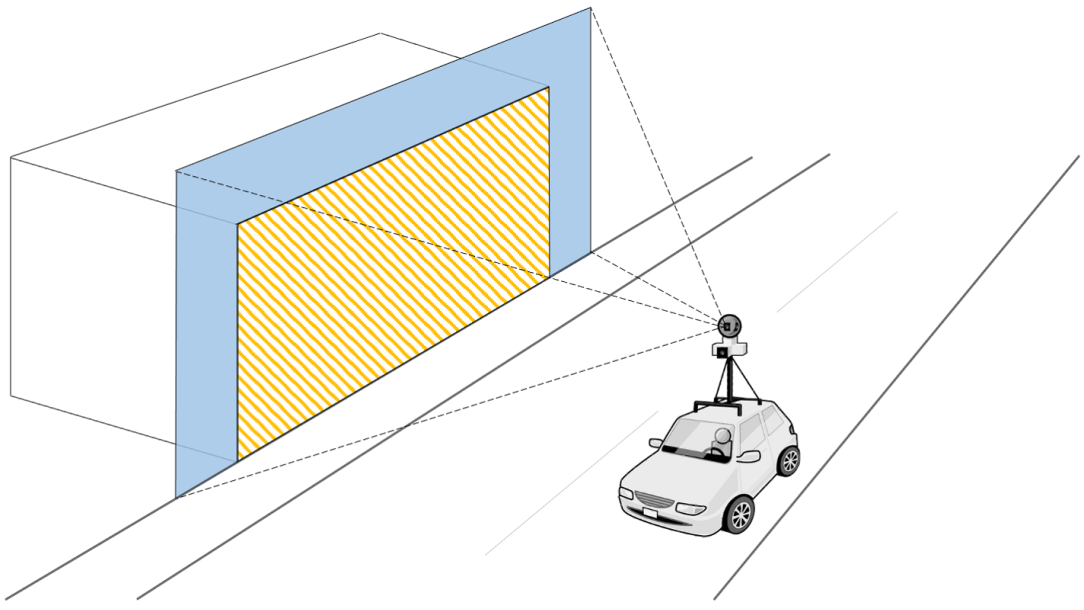


Figure 4: This figure illustrates the preferred scene of retrieving street-level facade imagery, with the camera perpendicularly aligned with the building facade.

Therefore, choosing the method for collecting street-level imagery was accomplished by considering the quality and coverage of the imagery provider. We observed that volunteered street view imagery (VSVI) platforms (e.g. Mapillary or KartaView) provided less extensive street-level imagery coverage. Thus, choosing Google Street View enabled us to automatically collect a satisfactory amount of imagery within the broader geographical area with adequate quality. Furthermore, the retrieved building addresses from OSM were used as input in the Street View Static API to request street-level images specified by a location parameter. Using the address string instead of longitude/latitude values, the API requests an image with a direct view of the specified address location. In contrast, requesting a location using longitude/latitude returns an image that is closest to the position of the given location, with a slightly higher margin for error. Moreover, handling addresses makes the results more human-readable and facilitates communication when managing data processing between OpenStreetMap and Google Street View. Using the address text string as an HTTP URL request parameter contributes to the overall higher performance of the method concerning the quality of street view imagery. Moreover, an image is mainly taken from the road adjacent to the facade at street level, where the street covers a larger portion of the lower part of the image. Therefore, tilting the camera angle slightly upwards will increase the probability of including all floors of a facade, namely on images of tall buildings where the distance between the camera position and the facade is short. We did this by adjusting the pitch parameter, representing the relative angle of the street view camera along the vertical axis.

3.2.2 Facade Object Detection

Further on, the detection of facade objects was conducted by integrating the facade parsing pipeline developed by Kong and Fan 2020. The pipeline included three subnetworks, where we only chose to consider the network handling window/door/balcony detection. Using the Google Street View images as input (640x640 pixels), the following output was four normalized coordinates portraying the bounding boxes of the detected objects, with a corresponding class label and precision threshold. We chose to consider windows and doors throughout our method, as we assumed that

these were the most influential features on a facade. The output was then saved as a list of objects. Each object contained information about the positioning and geometric extent used to segment the facade using the RANSAC floor segmentation algorithm. In addition, we extended the object geometry to include the three points on the centerline of the object, namely the left edge, right edge, and the centre of mass of the bounding box.

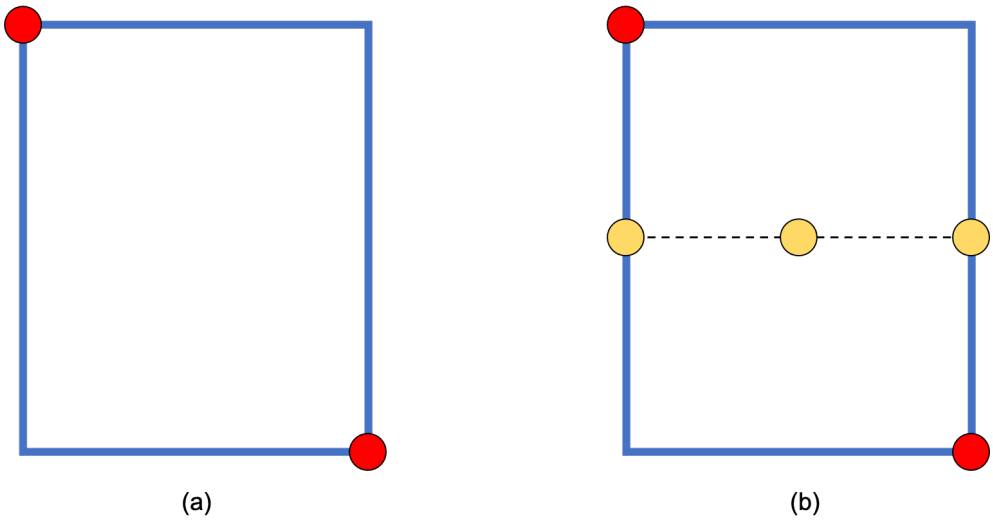


Figure 5: This figure illustrates how adding three centerline coordinates can aid in the floor segmentation process. The initially given bounding box coordinates (red) from the detection method is shown in (a), and the added centreline coordinates (yellow) is visualized in (b).

3.3 Determination of Floors on Facades

Facade heights can typically be expressed in either absolute metrics or as the number of floors. However, supplementary information about the arrangement and use could also be derived from knowing the number of floors. In general, the floor-to-floor height difference is negligible among a representative set of buildings with similar use (e.g. residential), given the conventional use of standardized regulations (Direktoratet for Byggkvalitet 2017). However, the architectural style and its integration into the surrounding terrain may cause variations in the interpretation of the total facade height, as shown in Figure 6. Moreover, between distinct building

types, the floor-to-floor height may also deviate strongly from the average practice to meet the requirements of their service, in addition to internal variations within the same structure (Council 2021).

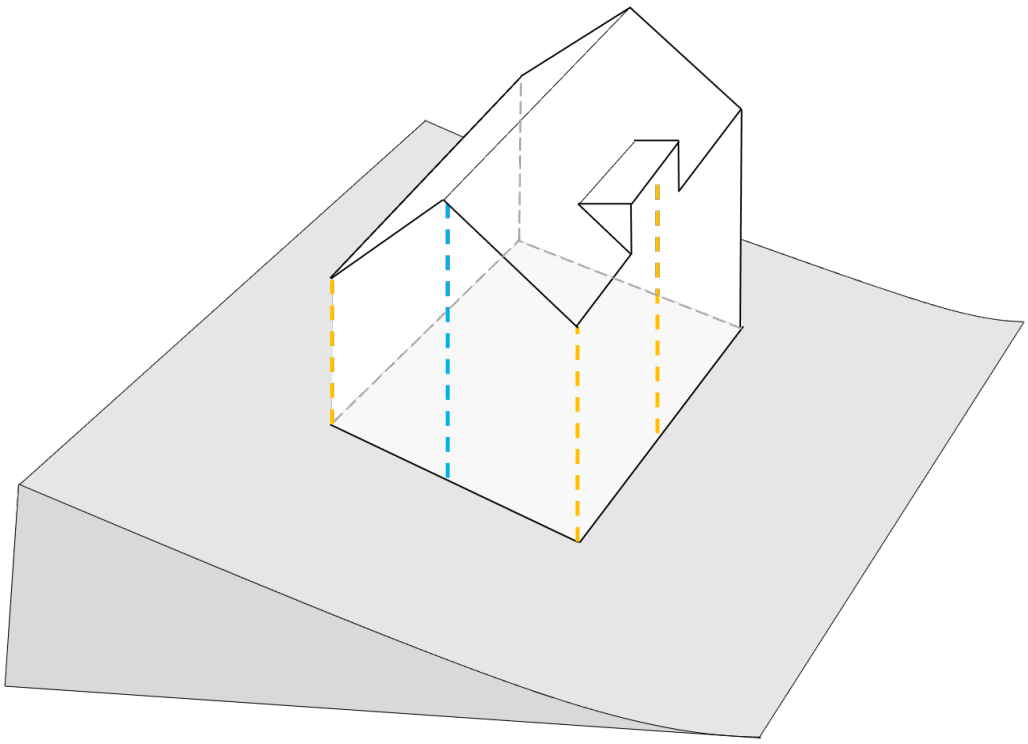


Figure 6: This figure shows the different facade height definitions and how they are affected by architectural style and surrounding terrain. The dashed lines are examples of the facade height, with cornice height (yellow) and ridge height (blue).

To accurately describe and interpret the structure and features of a building facade, identifying key characteristics is conceivably important. In particular, the relations between and among windows and doors are essential aspects to consider when attempting to describe a facade. Given these objects, one can utilize metrics such as the presence, quantity, and position of the objects to infer knowledge about the semantics of the facade. Such knowledge is based on the inherent features of already established architectural patterns that are fundamentally predictable and pragmatic.

A random sample consensus (RANSAC) algorithm from Scikit-learn was employed to enable the segmentation of floors (sklearn.linear_model.RANSACRegressor, Pedregosa et al. 2021). The algorithm was used to allow robust estimation of parameters

from subsets of inliers in the observed data, namely the centre coordinates of objects (s.a. windows and doors) found by the aforementioned detection method, to fit a model to determine separate floor lines. The additional centerline coordinates added in the pre-processing will further enhance a potential floor line, especially where the number of windows is low. As the RANSAC implementation can only estimate one such model for a particular dataset, we tweaked the method to enable the fitting of multiple lines to the observed data (Zuliani et al. 2005), enabling multi-floor segmentation of facades on the provided street view imagery.

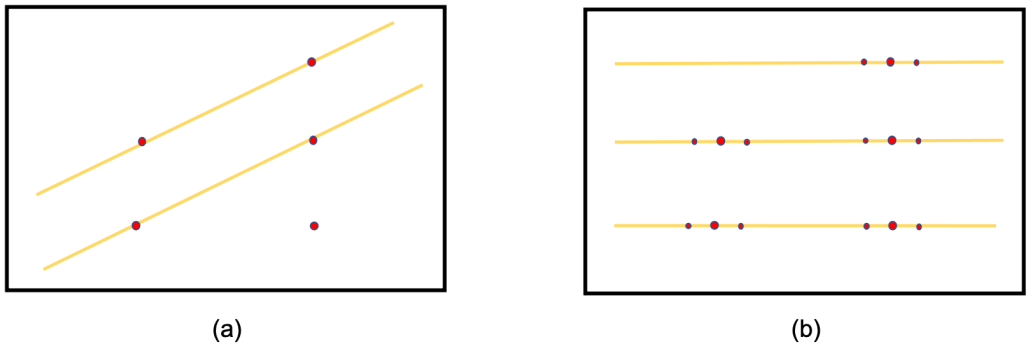


Figure 7: This figure illustrates how adding two extra centre line coordinates can aid in the floor segmentation process, with window coordinates represented as points (red) and intermediate/outlier windows (yellow).

A significant advantage of using the RANSAC algorithm in this particular case is that the observed data is symmetrically structured and logically ordered, as facade objects tend to follow standard architectural principles. Further on, the handling of potential errors and inaccuracies arising from the previous steps was deemed necessary in order to ensure greater consistency and easier manipulation of data throughout the process. Therefore, we included a step for handling error correction and adjustment of the output. We intended to manage the correctness of the representation of the separate floors, rather than being an extensive step for handling the inclusion and exclusion of specific facade objects for each floor. This is caused by the fact that the number of floors is dependent on sets of objects constructing each of the separate floors present on the facade. Therefore, when estimating the facade height, sets of objects are more important and serve as a primary feature, and individual objects only serve as secondary features.

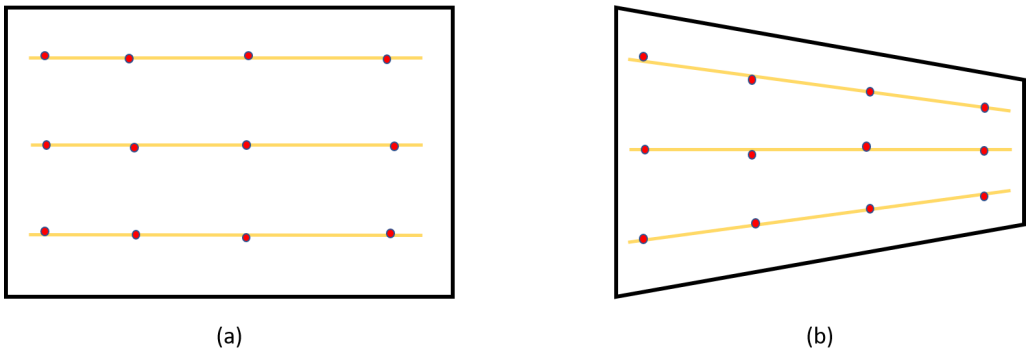


Figure 8: This figure shows a facade oriented perpendicularly (a) and a distorted facade (b), where both are valid facades. The figures contain the facade outline (black quadrilateral) with floors (yellow lines) and facade objects centres (red dots).

Managing the floors to represent a facade accurately required a reassessment of the output generated by the RANSAC regressor. Even though the floor segmentation was primarily returning valid and logically aligned floors (Figure 8), there was still a possibility that the method had recognized arbitrarily fitted lines as floors. The issue of overlapping and misaligned floor lines was one of the main problems arising from this implementation of RANSAC, as the algorithm fits any model that is not rejected due to its inherent restrictions (Figure 9). Evidently, the preceding behaviour was deemed erroneous for the facade floor segmentation, requiring corrections to achieve the desired outcome.

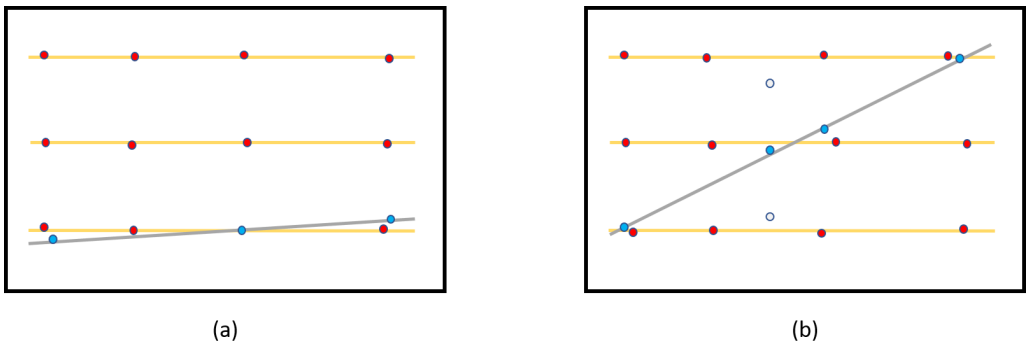


Figure 9: This figure shows two facades with invalid floor segmentation. Facade (a) has overlapping floor lines, and facade (b) has misaligned and misclassified floor lines. The figures contain the facade outline (black quadrilateral) with floors (yellow lines), facade objects centres (red dots), misclassified floors (grey lines), objects that could potentially be part of a misclassified floor (hollow blue dots) and objects part of the misclassified floors (blue dots).

We implemented the corrections by inspecting the object-to-object, object-to-floor, and floor-to-floor relations of the segmented facade. By assessing these relations, we determined if the floor lines erroneously intersected and merged the floors that were too close in relation to the average distance between all floors. We also effectively removed any floor lines that were misaligned and significantly deviated from the average absolute slope of all floor lines, considering that the facade may be prone to perspective distortion. By doing this, we were able to circumvent the issues demonstrated in Figure 9.

3.4 Height Estimation by using Urban Rules

In some cases, it might be relevant to consider the spatial extent of the objects contained on each floor, as an object can provide valuable information about the mathematical relations of a facade. However, it is important to note that a single object will not provide any significant information regarding the height of a building, but a single object measured up against the sets of objects contained on the whole facade may provide valuable insight.

With the refined images from the previous step, we improved the estimation of the facade height by investigating the topography of the arrangement of objects on the facade. Then, based upon commonly known architectural patterns observed on existing facades, we exploited mathematical metrics and relations among the detected objects and floor lines to adjust the relative physical height between the floors. A set of rules were then applied to potentially adjust the estimated floor height if the given rule returned a true statement.

1. **Rule: Basements**

   The facade consists of a basement if the windows located on the lowest detected floor are significantly smaller than the rest.

   Consequence: Reduce the height of the lowest floor.

2. **Rule: Shop Floor**

   The facade is a commercial service building if the windows located on the lowest detected floor are significantly larger than the rest.

   Consequence: Increase height of the lowest floor.

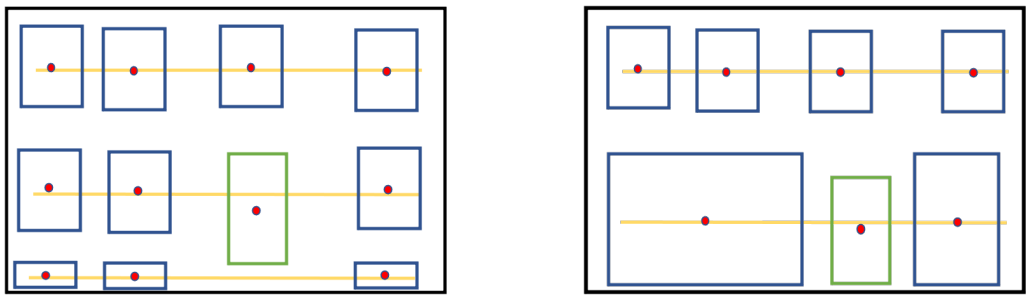


Figure 10: This figure shows two examples where the rules are relevant, i.e. the presence of basement floor (a) and store floor (b). The figures contain the facade outline (black quadrilateral) with floors (yellow lines), doors (green quadrilateral), windows (blue quadrilateral) and facade objects centres (red dots).

We used general knowledge and observations to decide if a relation between objects or sets of objects is decisive, in addition to manually tweaking variables to see which values yield the best results. If neither of the aforementioned rules has been enforced, we performed an additional search for any relevant building data. Particularly, using the building type of a facade enabled further possible adjustment of the estimated height, as it may contain relevant characteristics that influence the structure of a facade (Figure 11).

1. **Rule: Commercial**

   The building type is Commercial

   Consequence: Increase height of all floors.

2. **Rule: Service**

   The building type is Service

   Consequence: Increase height of all floors.

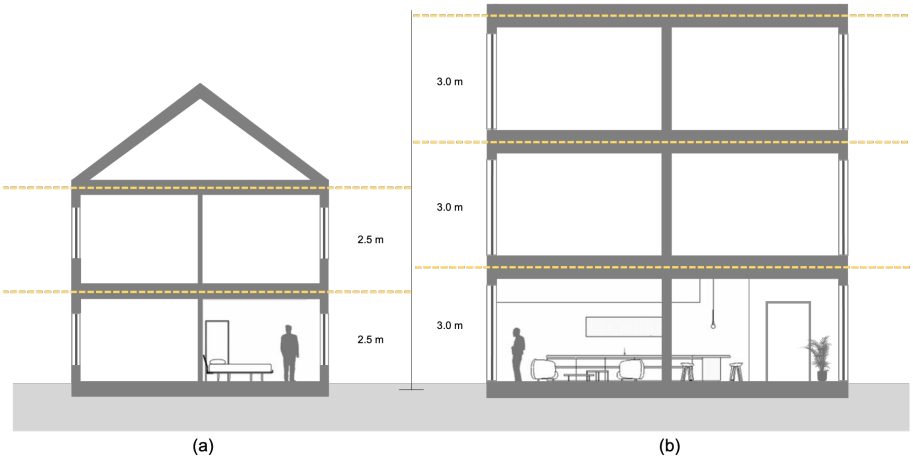


Figure 11: This figure shows an example of the floor-to-floor height: (a) residential building; (b) commercial building.

In essence, the collection of the defined rules are based on a standardized floor height and slight variations given certain features or attributes that are either decided by mathematical relations found on the facade or from data extracted from the OSM database. The standardized floor height is dependent on building regulations and may therefore vary in different areas and countries. However, we assumed that a general floor-to-floor height was around 250 centimetres for residential buildings and 300 centimetres for commercial and service buildings (§ 12-7, § 12-8)(Chun and Guldmann 2012).

Note that the minor height differences among neighbouring buildings caused by variations in parapets, foundation heights, and floor thickness do not adversely affect the scale, consistency, and character of a street scene (Council 2021). Therefore, applying a standardized metric will give an accurate estimation of the urban scene as a whole, generating a digital twin facilitating large-scale surveying and analysis of the urban environment.

With the implementation of urban rules and floor segmentation, we performed a large-scale estimation of the facade height for our given study area. Finally, we combined mapping data from OSM and the Norwegian Mapping Authority (NMA). Finally, we used software provided by our geomatics research group at NTNU to generate a 3D model to visualize the results.

3.5 Evaluation

In order to evaluate our method, we chose to decompose the problem into two parts. One part checks the viability of our method when deployed on a large-scale urban environment for height estimation. The other assesses the accuracy of the floor segmentation and application of urban rules for an optimal subset of street-level imagery through manual inspection.

To evaluate the resulting height estimations, we approached the problem by considering a LiDAR point cloud covering the study area to compare our results with ground truth height data, extracted from LiDAR data provided by The Norwegian Mapping Authority. The ground truth height values for each building in our dataset were calculated by segmenting the LiDAR point cloud in the XY-plane using the building footprint coordinates acquired from OSM and obtaining the average z-value of the selected subset.

Additionally, the manual inspection of the selected subset of optimal facade images was performed to assess the correctness of the implemented floor segmentation, thereby eliminating the influence of erroneous features caused by irrelevant factors regarding segmentation quality. Initially, the correctness was measured by simply comparing the detected and actual number of floors observed per the facade. Furthermore, we measured the degree of error by counting the number of detected floors that deviated from the true number. Finally, the impact of the defined rules was checked by comparing whether there was an improvement of the height estimation if a relevant rule was applied or not.

4 Experimental Study

In this chapter, we will present the objective of the experiment, how the experiment was conducted, and the subsequent results. First, we estimate the facade height of buildings from street view imagery in the chosen study area by utilising our floor segmentation and rule-based method. Then, we continue with presenting the results, culminating in a discussion of the following output.

4.1 Experiment environment

We developed the height estimation pipeline in a Python environment, utilizing Overpass API (Olbricht 2021) to access OSM data and Google Street View Static API (Google 2021) to access street view imagery in Google Street View (GSV) to collect all required data for pre-processing. We employed the convolutional neural network pipeline of Kong and Fan 2020 to enable facade parsing and object detection of the input data. Then we implemented the floor segmentation algorithm using the RANSAC regressor from Pedregosa et al. 2021. The 3D model representation was done using the 3D model generation software from our Geomatics research group, currently available in internal testing.

GPU: NVIDIA GTX 980 Ti
Processor: Intel Core i5-8600K CPU
Memory: 16 GB Memory

4.2 Study Area

The experiment was conducted as a two-fold case study within the region of Trondheim, Norway. The chosen area is enclosed with a rectangle (Figure 12), with minimum and maximum longitude and latitude of respectively (63.40795, 10.33578) and (63.44873, 10.46711). We did the study to carry out a large-scale survey of the

area and ensure a comprehensive examination of the performance of our method. The intention was to investigate the viability of the method as a large-scale analysis tool and to check if it could be deployed in a complex city environment realistically. To further examine the quality of the implemented floor segmentation and the applied rules, a manual selection of facade imagery was made in order to simulate an optimal testing environment.

Figure 12: This figure shows the chosen study area of Trondheim, Norway (red quadrilateral) (OpenStreetMap 2021a).

We intentionally chose the study area to cover multiple city district types to test the flexibility of our method. Furthermore, Trondheim consist of several urban districts with different characteristics, including industrial areas, city centres and suburban areas. It is worth mentioning that the distinct features may influence our results within the chosen study area, such as cultural and environmental factors. However, the fundamental principles of a facade structure are shared between cities in general, and the impact could be assumed to have minor significance. Additionally, given our familiarity with the study area, our underlying knowledge was exploited to evaluate further and interpret the results.

Table 2: Building types in OSM.

	Residential	Commercial	Service	Industry
Include	Apartments, Terrace	Retail, Commercial, Office	School, University, Hospital, Kindergarten	Warehouse, Industry
Exclude	Garage, Shed			

4.3 Experimental Results

In this subsection, we provide the results of the main parts of our method: the floor segmentation and the application of urban rules.

The collected data from OSM and Google was retrieved throughout March, April, and May 2021. A total of 16288 unique buildings were extracted from the OSM dataset within the study area. After filtering the building data through the pre-processing step and disregarding any non-unique buildings, i.e. buildings not containing an address node and areas lacking or erroneous street view imagery, 6233 buildings were included. Finally, the method disregarded 1283 images, and we ended up with 4950 images that yielded adequate results considering the floor segmentation and rule-based height estimation. In addition, we selected 50 images distributed uniformly throughout the study area to analyse the results further manually, all of which concurred with the requirement of proper alignment to their respective facades and were representative of the different district types (Figure 13).

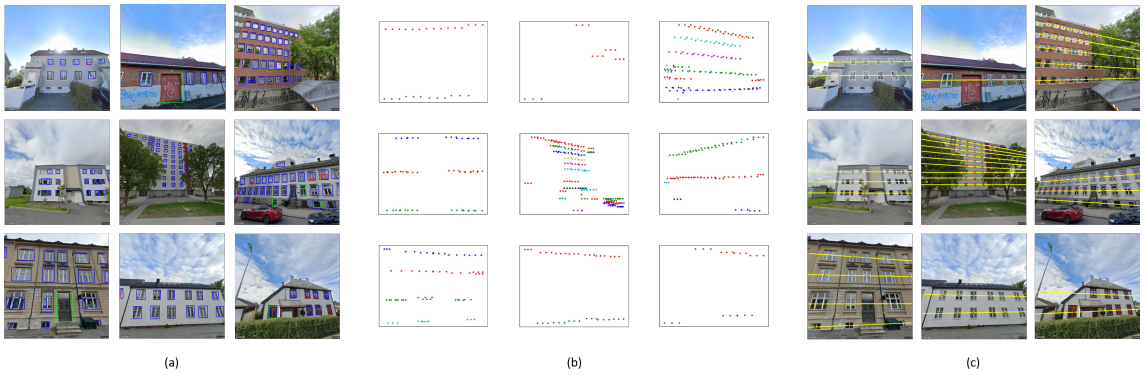


Figure 13: Visualization of the results from an example subset of street-level imagery. (a) The completed input from the pre-processing step with detected objects on the facade imagery. (b) The corresponding plots from the RANSAC algorithm with object coordinates (inlier points) where the different colours represent floor affiliation. (c) The segmentation of each unique floor on the different facades.

In Figure 13, we see the input data from the pre-processing step (Figure 13a), the plots from the RANSAC object fitting (Figure 13b), and a visualization of the final

results of the floor segmentation of RANSAC displayed on corresponding the street-view imagery (Figure 13c). We can observe that the detected floor lines are generally fitted to the corresponding floors on the street-view images with satisfactory accuracy (Figure 14). Furthermore, as observed throughout the results, when considering optimal street view imagery and a clear and distinct facade comprising an overall symmetrical facade structure, the floor segmentation was frequently correct.

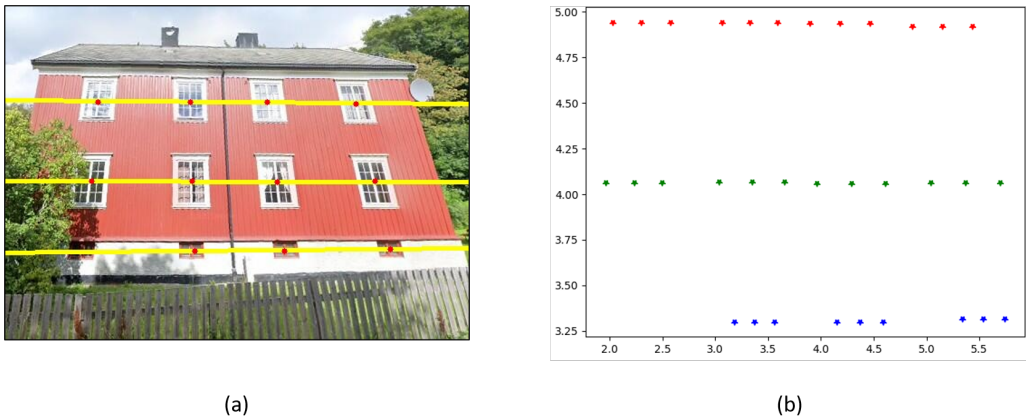


Figure 14: Visualization of the results of the multi-RANSAC regressor algorithm; in (a) the object centres (red points) and floor lines (yellow lines) on the facade imagery, and in (b) a plot of the corresponding segmented floors with object coordinates (red, green, and blue points).

In contrast, we also observed that the floor segmentation on street-view images that were sub-optimal in terms of positioning, alignment, occlusion and background noise performed significantly worse (Figure 15). Specifically, we observed that the number of segmented floors varied depending on the framing of the street-level image. Furthermore, the misalignment was observed as the image was slanted along the road axis due to the image capture interval. In addition, the occlusion of facades mainly came from trees, cars, and other city furniture. Finally, non-relevant facades were sometimes included in the background or adjacent to the facade in question, resulting in faulty floor segmentation, as seen in Figure 15. Nevertheless, the resulting number of floors estimated was adequate in many cases, or the images had no detected facade objects or segmented floors. They were consequently disregarded as part of the 1283 removed images.

Figure 15: Example visualization of the floor segmentation results on street-level imagery that was either incorrectly positioned, occluded, misaligned, or included background noise.

Furthermore, Figure 16 shows a set of examples visualizing how the method performed on buildings with complex facades, usually consisting of buildings with fewer detected objects on each facade. We observed that the recurring issue of the complex facades was mainly due to asymmetrical facade patterns and irregular distribution of facade objects. As a consequence, no logical ordering of floors could be described by our method.

Figure 16: Visualization examples of the floor segmentation results on street-level imagery with complex facades.

Moreover, Figure 17 represents a selection of buildings including a basement floor, with the top row representing facades with correct basement detection and the bottom row representing undetected basement floors. We see that the basements are not considered mostly due to the detection method's inability to discover the smaller windows. In addition, we also observed that basement windows were in some cases occluded by various objects (e.g. fences, cars, and vegetation). However, in most cases, we could clearly distinguish between the size of the basement windows and the windows above. The heights of the foundation on which the basement windows are placed were repeatedly similar, facilitating for an accurate height estimation result in total.

Figure 17: A set of facades where the basement rule has been applied on the detected basements (top row), and a set of facades where the basement rule has not been applied (bottom row).

An example of facades with shops is shown in Figure 18, and we can observe that the shop floors remain undetected. We observed this recurring situation throughout the results, as the detection method could not locate shop floor windows on numerous facades.

Figure 18: Visualization examples of the floor segmentation results on street-level imagery with undetected shop floors.

Table 2 represents the building types found in our study area, yet most observed buildings had no information about their type. Moreover, even the few building types found, such as offices and universities, were mostly complex with abnormal facade object size and distribution. Consequently, the impact of the rules was minimal.

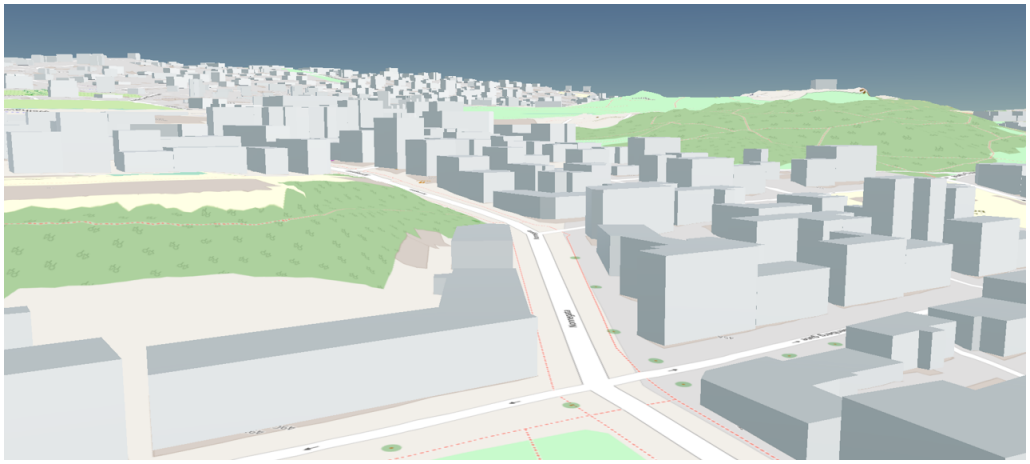


Figure 19: A view of the resulting 3D model representing a neighbourhood in the study area of Trondheim (3D model from Geomatics group, NTNU, Trondheim).

The final height estimation result on a large scale, as is visualized in Figure 19 and Figure 20, gives an overall sufficient impression of the scale and character of the urban environment. In particular, when observing areas with sporadically missing or erroneous data, the neighbouring 3D buildings provide a decent representation of the general area as a whole, based on the average value to adjacent 3D models and the character of the urban block.

(a) (b) (c)

Figure 20: 3D visualization of the results in the study area of Trondheim. (a) An urban residential area, (b) a city centre, and (c) a suburban residential area.

Finally, comparing the generated 3D model to the real street view imagery, as presented in Figure 21, we can see a clear correlation between the two representations, proving adequate height estimations for the particular area. In general, we observed the best representations of the city environment in urban and suburban residential neighbourhoods.

(a) (b)

Figure 21: This figure shows a comparison between a street view image with a shop floor (a) and its corresponding, one-story, 3D model representation (b). ((a): Google, "Streetview," digital images, Google Maps (GoogleMaps 2021), photograph of Båhus gate, 7030, Trondheim, captured: Aug 2020)

4.4 Evaluation

We chose to evaluate our method with a decomposed two-part approach. In the first part, we intended to evaluate the correctness of the floor segmentation rigorously. Whereas, in the second part, we evaluated the method based on how well it performed as a large-scale height estimator and its accuracy of the portrayal of

the real-world counterpart. In this subsection, we thereby present an evaluation of the results, both the performance of floor segmentation on optimal images and the height estimation in various scales.

4.4.1 Evaluation of Floor Segmentation

The manual evaluation of the floor segmentation on the subset of 50 manually selected facades is presented in Table 3 and Table 4. Table 4 presents the correctness of the floor segmentation, applied rules and their associated errors. Overall, the correctness of the floor segmentation yields good results with an accuracy of 0.92. Furthermore, a total of four facades had segmentation errors with $Mean_{ErrorDegree} = 1.250$ and $StandardDeviation = 0.274$, $Min_{ErrorDegree} = 1$, and $Max_{ErrorDegree} = 2$. Three facades were segmented with $ErrorDegree = \pm 1$ and one facade was segmented with $ErrorDegree = \pm 2$.

Table 3: Evaluation results of the floor segmentation.

Buildings	**Segmentation Errors**	**Error Degree**	**Basement Detection**	**Shop Detection**
1-10	0	-	5/6	0/2
11-20	2	1, 1	3/3	1/2
21-30	2	1, 2	4/5	-
31-40	0	-	2/3	1/3
41-50	0	-	5/5	2/2
	46/50	-	19/22	4/9

Table 4: Statistical floor segmentation results.

Mean	**Standard Deviation**	**Variance**	**Max**	**Min**
1.250	0.274	0.075	2	1

We observed that the applied basement rule yielded significantly better results when compared to the applied shop rule. In total, the basement rule was applied on 22 facades with a correctness of 0.867, compared to the nine facades containing a shop with correctness of 0.444. It is worth mentioning that none of the rules was incorrectly applied in situations where they were not relevant.

4.4.2 Evaluation of Height Estimation

The height comparison between the estimated heights and the ground truth data in our study area is visualized in Figure 22. The apparent missing data can be explained by the inability to estimate any or too few facade heights, in addition to having detected no buildings contained within a particular unit of the hexbin. As a result, we can see that the bulk of our estimated heights had a value in the bottom part of the spectrum with height difference values ranging from 0 to 2 floors (0 to 5 meters). Hereof, we can presume that the height estimations were, for the most part deviating from the ground truth with one or fewer floors when disregarding the roof height. This can be further visualized in Figure 23. An interesting observation from Figure 22 is that hospital complex, university campuses, and industry zones located in our study area were the main benefactors of the poor height estimations.

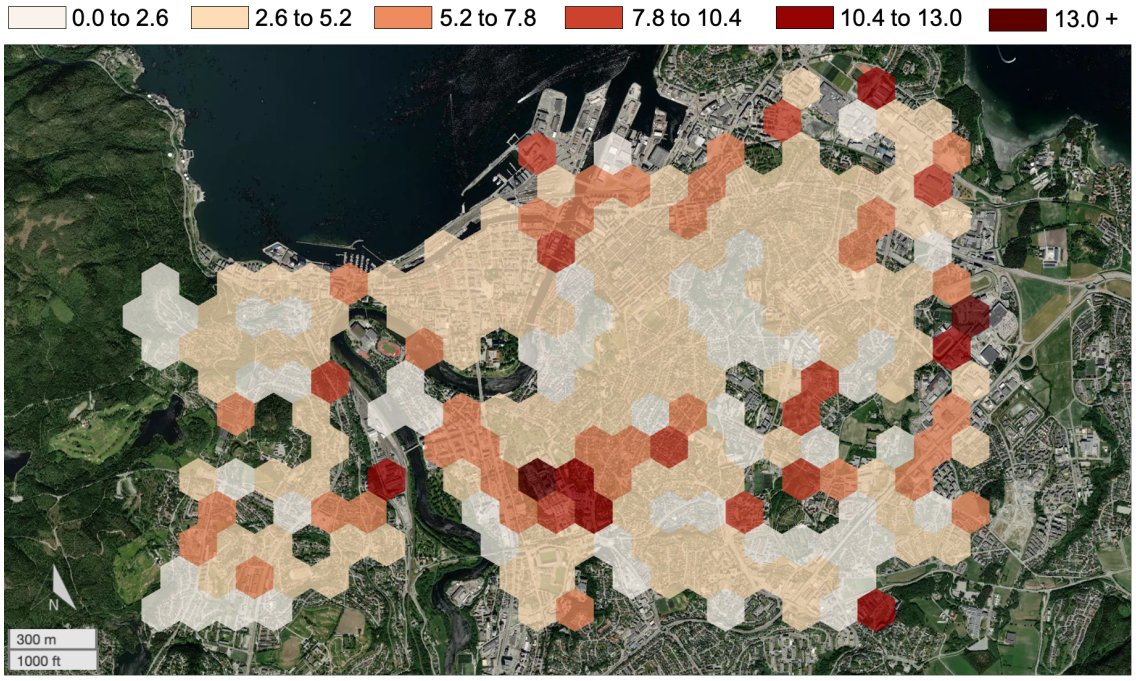


Figure 22: The hexbin plot represents the difference between the ground truth heights and estimated heights. The legend shows the height difference in meters in six distinct categories.

Further elucidation from Figure 23 emphasises the distribution's discrete nature, as buildings are individually considered, with a one-to-one relationship between

the estimated heights and the ground truth. It is worth noting that the height estimation is segmented into explicit intervals as the rules assign a limited set of values. Meanwhile, the ground truth is collected from a wide range of values in the vertical dimension of the LiDAR output.

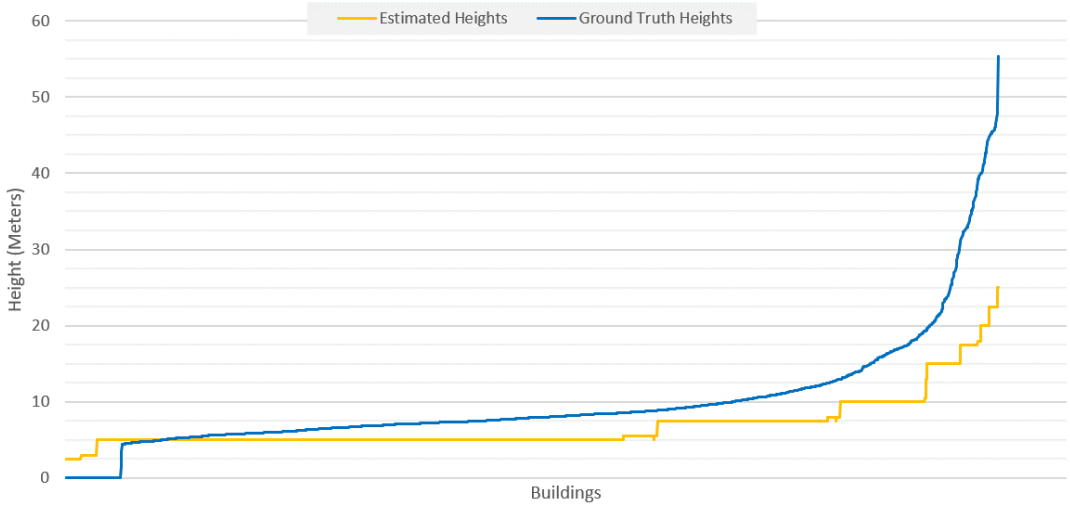


Figure 23: Visualization of the height results compared to the corresponding ground truth heights from the distribution of buildings in the whole study area of Trondheim, sorted from smallest to largest height values.

Evident from the distribution curves in Figure 23, the height difference can be observed to vary between 0 to 5 meters, with a $Mean_{0 \leq y \leq 5} = 2.537m$, for buildings ranging from 5 to 10 meters. In contrast, the difference between building heights beyond 10 meters diverges with a $Mean_{y \geq 10} = 14.372m$. Thus, the aforementioned observations seem to be correlating with the results found in Figure 22.

Generally, the larger values in Figure 22 can be explained by the presence of taller buildings or complex facades within the hexbin unit, as they often are erroneously estimated. For example, the university campus area of NTNU is a compound scene of several large buildings with many facade objects and a varying degree of complexity, causing significant deviation in the difference between ground truth heights and estimated heights (Figure 22). This is an obvious limitation of our method since the height estimation became increasingly erroneous when the buildings went past 20 meters high (Figure 23).

Table 5: Statistical height estimation results.

	Height Estimation	**Ground Truth Heights**
Mean	6.850	9.922
Standard Deviation	3.554	7.733
Variance	12.631	59.803
Max	25.000	55.340
Min	2.500	1.490

Moreover, the statistics in Table 5 explain the relation of the height estimation and the ground truth heights. As a result, we can observe that there is a clear distinction in the statistical data between the two, highlighting the tendency that the ground truth heights were on average greater than the estimated heights, as can be seen with the logical expressions $Mean_{GT} > Mean_{Estimate}$ and $Max_{GT} \gg Max_{Estimate}$.

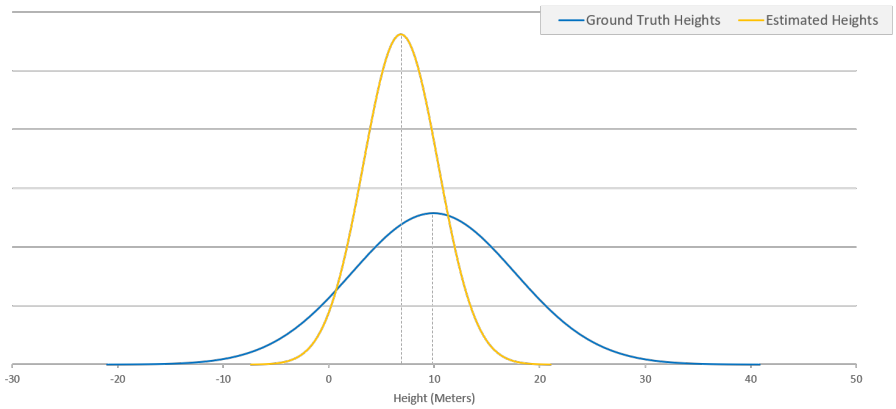
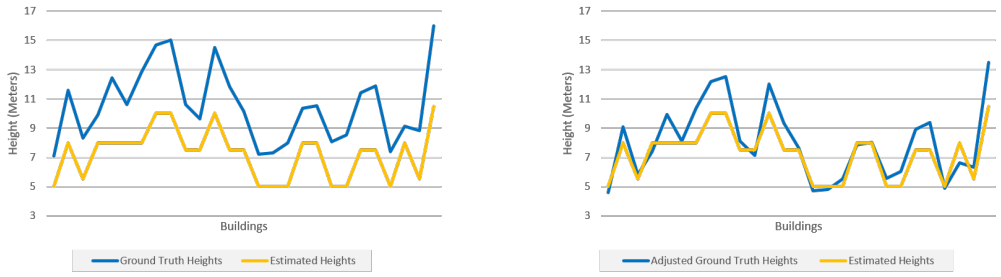


Figure 24: Visualization of the normal distribution of height results compared to the normal distribution of ground truth heights from the distribution of buildings in the whole study area of Trondheim.

Figure 24 illustrates the central properties of the two datasets by highlighting the distribution of the height values. As expected, we observed that the height estimation results were distributed with a relatively small $Variance_{Estimate} = 12.631m$ about the $Mean_{Estimate} = 6.850m$, as a consequence of the limited set of values generated and the absence of higher buildings compared to the ground truth data. Furthermore, we observe that about 68% of the estimated buildings were distributed in the height range $x\,\epsilon\,(3.297, 10.403)$ within one standard deviation σ, as calculated from $Pr(\mu - 1\sigma \leq X \leq \mu + 1\sigma) \approx 68.27\%$. Meanwhile, the ground truth is distrib-

uted within the range $x \epsilon (2.189, 17.655)$ about $Mean_{GT} = 9.922m$. Moreover, there was an uneven distribution of heights when comparing the estimations and ground truth values, further supporting the observations made earlier in Figure 23.



(a) (b)

Figure 25: Visualization of the height estimation results compared to the ground truth heights from the distribution of buildings in the selected street Båhus gate. (a) Shows the distribution of heights for the estimated heights and the corresponding ground truth heights. (b) Shows the distribution of heights for the estimated heights and the ground truth heights adjusted with the mean difference.

An evaluation of a selected residential area was done to more accurately evaluate how the method performs on a smaller scale with recurrently similar architectural patterns. The result of the height estimation can be seen in Figure 25, with seemingly better results. By adjusting the ground truth height with the mean difference, the data was closer fitted to the height estimations, resulting in a matching height profile.

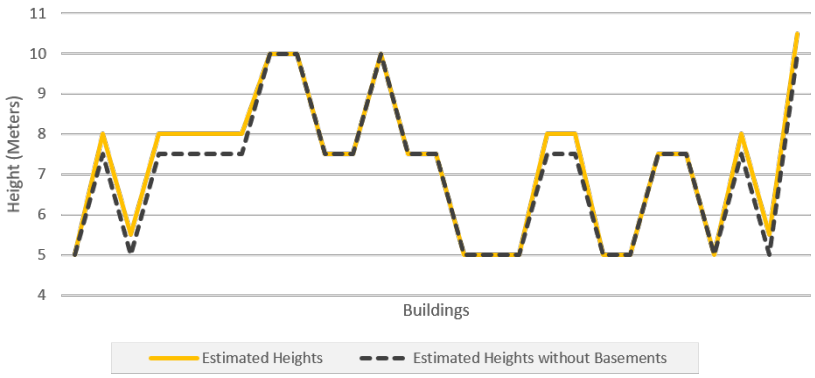


Figure 26: Visualization of the height estimation, both with the inclusion and exclusion of detected basement floors on the buildings in Båhus gate.

By disregarding detected basements added from the applied basement rule, we can see how the resulting height profile would have been estimated only regarding regular floors (Figure 26). Therefore, we deduce that the implemented rules result in a more accurate height estimation, as a consequence that it gives a better approximation to the ground truth heights in this particular area (Figure 25).

Finally, we can see a trend that the various errors found in our results were mostly found within clearly separated areas, typically service or industrial areas. Furthermore, we observed that a reason for this might be their inherent facade complexity and tall stature. In contrast, we observed that the floor segmentation correctly estimated the number of floors, especially in urban/suburban residential areas. Furthermore, the applied rules contributed to an increase in height estimation accuracy.

4.5 Discussion

The following subsection will discuss how the data acquisition step, with the management of building information and street view imagery, impacts the floor segmentation and rule-based height estimation and the final results' quality.

4.5.1 Building Information Acquisition

The pre-processing of the data can be regarded as having a significant impact on the overall quality of the results. Evidently, the importance of data acquisition and management evolve around facilitating an optimal floor segmentation and height estimation. We witnessed from the results that high-quality input with clear and distinguishable features allowed for more effective estimation and less erroneous output.

The data integrity of OSM is dependent on the regularity of community updates that implicates variations in data quality provided by the OSM communities in various areas. Nevertheless, the data accuracy is not significantly impaired with the use of OSM as the primary data source, as the spatial accuracy and completeness of

the data were adequate (Brovelli and Zamboni 2018). Moreover, building type and amenity were observed to be an infrequently distributed tag and rarely found in our study area.

Furthermore, gathering the contained address nodes within the building footprints from OSM was a resource-heavy and mathematically complex process resulting in systematic errors. It was not guaranteed that a node was located within its corresponding footprint, and as a result, some building footprints that would have been viable for further processing were excluded. Upon inspecting the 6233 returned viable buildings of the total 16288 buildings in the study area, we observed that most of the excluded footprints were building annexes (e.g. garage, shed), yet several buildings were still erroneously disregarded. This was particularly true for buildings described as multi-polygon relations in the city centre or townhouses in residential areas, as the address nodes were sometimes outside the limits of the footprint.

In some areas, the complete absence of building footprint information may result in an inability to georeference the output of our method, hence requiring a way of managing this limitation. Here, creating a generic building extent mimicking the footprint at the location the photo was captured, based on the number of windows detected on the horizontal line, aligning it to the terrain and surrounding buildings would be necessary.

To conclude how the management of building information and its acquisition impacted the results of our method, moreover by assessing the consequences that followed from the exclusion of building footprints and consequent building information, we found that even though some data was lost in the process and by utilizing a VGI open data platform, gathering footprints to georeference the results proved to be adequate, as seen in Figure 20. Consequently, the most important aspect to consider when facilitating the correct georeferencing of footprints is that their coordinates should be properly aligned to represent the real world counterpart as accurately as possible. Furthermore, by examining the 3D visualisation of the results, the footprints were aligned and positioned relative to the base map satisfactorily. Finally, the building information, especially the footprint, serves to enable the connection

of data and reconstruction of 3D models to visualize the estimated heights of buildings. The selection of provider, albeit authority cadastres or volunteered open data, is important to consider as completeness and spatial accuracy are critical metrics.

4.5.2 Street View Imagery and Floor Segmentation

The main problems arising from using street view imagery were perspective distortion, occlusion and background noise (Figure 15). The problems mainly emerged as the fetched images were framing the facade at an angle or that arbitrary objects occluded a facade. The way this method manages misaligned floors largely contributed to adequate results. However, in some cases, distortion of the image was large enough that the floors were mistakenly regarded as errors and consequently removed. This resulted in faulty segmentation of facades by not including actual floors whilst simultaneously including erroneously detected floors.

Additionally, occlusion reduced the number of potential objects detected, as they were either completely or partially hidden by objects such as trees, city furniture, or cars. Moreover, as the ratio of occluded facade objects and the total number of facade objects increase, the segmentation would become progressively more unreliable. In turn, if the detection step were unable to find an adequate amount of significant features, the semantics and symmetries inherent in the facade would become lacking.

Moreover, the aforementioned errors impact the overall performance of the floor segmentation in the case of automatic data acquisition, as a fair amount of images fetched from GSV were either misaligned with the address coordinate of the building or considerably occluded (Figure 15). Even though some images suffered from this, they were removed as they included too few facade objects, and consequently, no floor lines were detected.

The method showed promising results regarding floor segmentation whenever these problems did not considerably influence the images. Moreover, as observed from the results of the manual selection, control of environment variables significantly

increase the results as these errors were reduced to a minimum. Therefore, by manually acquiring and selecting images, occlusion and perspective distortion could be avoided altogether. Ultimately, the method was able to adequately segment floors given this particular problem, especially when street-level images were aligned optimally to the given facade (Figure 13).

(a) (b)

Figure 27: This figure shows the perspective distortion that may occur when fetching an image for a particular address (red quadrilateral) in (a), and the occlusion of a facade due to vegetation (b). (Google, "Streetview," digital images, Google Maps (GoogleMaps 2021), (a): Photograph of Kjøpmannsgata 42, 7030, Trondheim, captured: Aug 2020; (b): Photograph of P. A. Munchs gate 6, 7030, Trondheim, captured: Aug 2020)

Another error that may arise from street view imagery is the problem of accurately estimating the number of floors of individual buildings in densely built-up areas, as images are fetched at different positions along the street, and may either incorrectly display the facade at an angle or regard neighbouring facades as well when estimating segmenting the facade in question (Figure 28). However, we observed that this was not a substantial problem when evaluating buildings in homogeneous areas as buildings were often aligned in height, implying that their structural scale was recurrently similar. Even buildings with clear distinctions did not divide largely from the area in regards to the facade height. This can be explained by regulations of building density that local authorities may enforce, which was relevant for our case study as stated in Direktoratet for Byggkvalitet 2017 (§5).

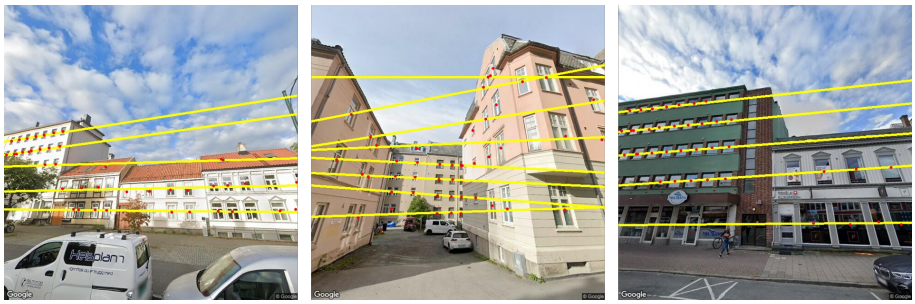


Figure 28: Visualization examples of the floor segmentation results on street-level imagery that had neighbouring facades or facades located in the rear, where the detected facade objects acted as background noise and contributed to faulty floor segmentation.

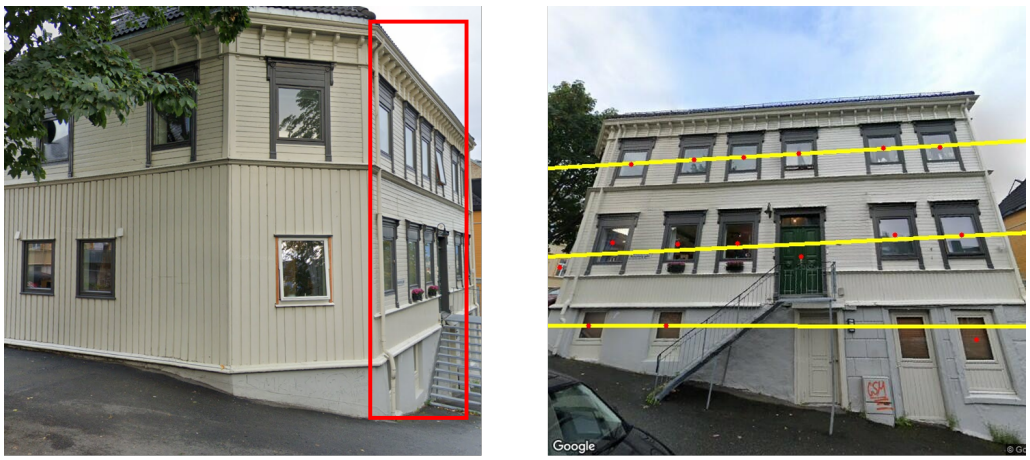
Nevertheless, the detection of nearby facades and their objects, particularly in situations where other facades were located in the background of an image (i.e. other facades partially obscured by the facade in question), impacted the resulting floor segmentation. Considering the background noise, the method erroneously recognised patterns from the detected objects in some cases and consequently segmented floors in a faulty way, as we see in the results (Figure 28). This could have been avoided by implementing a way of detecting and isolating individual facade outlines in order to manage each of them separately, as Kong and Fan 2020 mentioned could be a future improvement to their pipeline. However, even though this occurred, we observed satisfactory results in most areas where this was relevant (s.a. city centres), as seen in Figure 20. As for the manually selected imagery, this was largely avoided as the facades were aligned optimally.

Furthermore, since only street-facing facades can be observed, a given facade pictured on a street view image is not necessarily the most significant one, as buildings may have variable facade heights or include too few significant features overall. The former is especially true for buildings with gabled or saltbox roofs as one or more of the facades are dissimilar, increasing the difficulty of properly selecting the facade to estimate or simply making it impossible as all facades may need to be considered to give an adequate result of the building as a whole.

Using only a single image per building footprint had its limitations regarding quality assurance, which also applied when dealing with images taken in areas with steep

terrain. In such areas, the facade height of buildings will vary with one or multiple floors depending on which angle the building was observed from. We observed, as in Figure 29b, that the resulting floor segmentation was three floors. Meanwhile, in Figure 29a, the left facade would likely be segmented into two floors. Thus, providing a clear indication that the method was highly dependent on the angle at which the image was captured.

As such, collecting images of multiple facades for each building may provide a complete view of the building. However, problems arise as this would significantly increase the complexity of the floor segmentation process, where more dimensions must be considered, and the architectural patterns become three-dimensional. Furthermore, this would be redundant for the cases where buildings are flat or hipped, meaning that it would only be conciliatory to implement if all roof types were previously known, implying that the generality of the method outweighs the specific cases where this is relevant, at least when not considering the aforementioned roof types.



(a) (b)

Figure 29: This figure shows an example of a building in a steep terrain where the estimated height will vary depending on the position of the taken street view image. Image (a) shows both facades of the building with the segmented facade in (b) enclosed in the red quadrilateral. Image (b) shows the floor segmentation of the given street view image provided. (Google, "Streetview," digital images, Google Maps (GoogleMaps 2021), photograph of Rosenborg gate 28a/b, 7043, Trondheim, captured: Aug 2020)

Moreover, the applied facade object detection pipeline (Kong and Fan 2020) is a robust facade parser yielding sufficient results and works on complex scenes quickly and accurately. In addition, various window sizes and shapes could be detected without much trouble. However, the parser struggles with modern architecture and reflection on windows, as the method remains untrained to detect objects in these specific environments. Our method performed some error correction to prevent gross errors, where the detection method was flawed, thereby providing substantially better results.

We observed that storefronts often remained undetected in city centre areas due to not implementing the shop detection network and can therefore be regarded as a pervasive error. Furthermore, errors arise from the inability of the used detection subnetwork to handle reflectivity of windows, significantly restraining the current implementation. Respectively, the detection method we used could be supplemented by adding a sub-network specifically handling the detection of shop windows, given the recurrent nature of shops in the urban environment. This is particularly relevant in city centres, where the exclusion of shop floors may constitute a large percentage of the total facade height, significantly skewing the results and altering the surrounding environment's scale, consistency, and characteristics (Figure 30; 31).

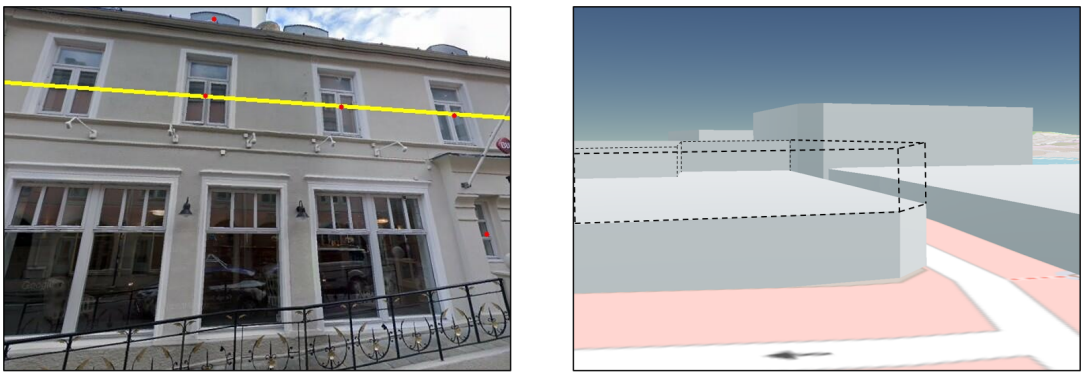


Figure 30: This figure shows a single detected floor (yellow line) of a storefront facade in (a), and the corresponding 3D building model at the location of the facade with a dashed outline representing the undetected floor in (b).

(a) (b)

Figure 31: Comparison of the floor segmentation results on street-level imagery with detected (a) and undetected (b) shop floors.

To summarize our findings regarding the floor segmentation, it was evident that the method could determine the number of floors on facades at an adequate accuracy. Finally, the quality of the images significantly influenced the output of the floor segmentation, meaning that a properly aforethought selection of imagery would be of great importance regarding the ability of our method to detect the floors adequately. Based on how the number of floors segmented on a facade, the knowledge of facade features, and building attributes influences the results, we can conclude that the method sufficiently estimates the number of floors, as evident in Figure 13.

4.5.3 Height Estimation and Urban Rules

With the research objectives in mind, we showed that syntactic knowledge of the architectural patterns and structure of facades could improve the estimation of the facade height based on street view imagery. Several rules could be implemented into the method to cover more patterns. However, only the most influential features impacting the height were utilized to arrive at the most beneficial result without over-complicating the process. It could be argued that the most evident factor for parsing a facade is the generally applied floor height standards, which we have exploited to enable a logical arrangement of the facade.

Moreover, the additional rules were implemented to enhance the results by exploiting variables, primarily to strengthen the height estimation from the floor segmentation and thereby account for less predictable features on a facade. However, as observed in the results, the impact of the rules was negligible, given that the coverage of building types in our study area was nominal.

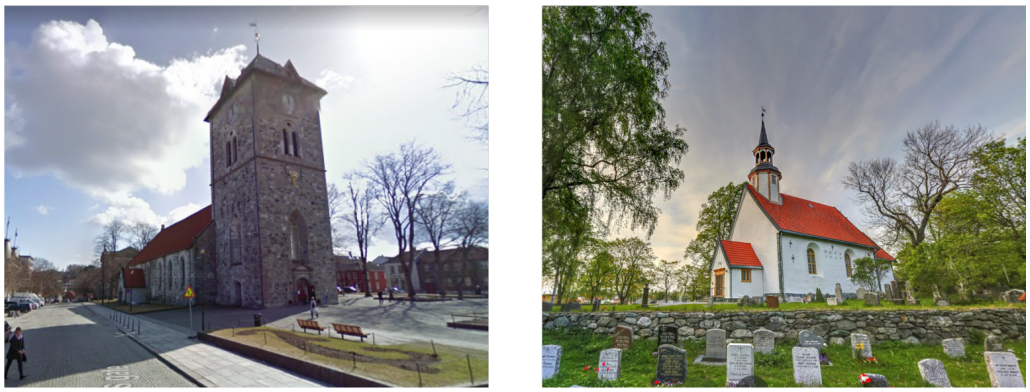
The lack of knowledge regarding the physical metrics on the facade and its surrounding elements limits the extension of supporting rules. Knowing the outline of the facade and its relative width and height could considerably improve the task of defining the facade, similar to the work done by Fan et al. 2021. The mathematical relations between the facade objects and the facade outline could then be utilized to determine a very accurate size of the facade by applying standardized sizes of facade objects such as windows and doors.

The fact that the facade height of a building may vary given the angle of observation and the shape of one particular facade is a case that introduces potential misinterpretations. This is especially true regarding buildings with gabled roofs, where facades are pairwise the same height, causing the height estimation to differ with a whole floor dependent on which facade side has been detected.

Residential and commercial buildings predominantly follow standardized facade patterns, which fit and endow good results with our applied method. However, our method struggles with complex facades that deviate from the traditionally applied architectural principles, particularly concerning modern or complex facade arrangements (Figure 16).

Further on, service and industry buildings such as churches, hospitals and warehouses usually have a very distinct and unique facade structure (Figure 32), in addition to abnormal floor-to-floor heights. This is a difficulty for facade parsing methods, as no obvious facade structure can be detected. Therefore, by exploiting building types, one could bypass the floor segmentation task when no facade objects were observed by giving a fixed height value based on the building type to yield more accurate results.

Following the concept of building typology from Chapter 2, perceiving a building in relation to its placement in the hierarchical network of the urban scene could therefore be beneficial, as there exist evident correlations between building structures and their surrounding enclave or neighbourhood. The consistency of similar characteristics and features among neighbouring buildings could be advantageous in cases where the absence of height data occurs on one or multiple buildings in a given area, illustrated through an example in Figure 33.



(a) (b)

Figure 32: This figure gives an example of how two churches share similar features within their unique facade pattern, consequently leading to similar facade height. ((a): Google, "Streetview," digital images, Google Maps (GoogleMaps 2021), photograph of Kongens gate 5, 7011, Trondheim, captured: Aug 2020; (b): Google, "Streetview," digital images, Google Maps (GoogleMaps 2021), photograph of Jarleveien 44, 7041, Trondheim, captured: Aug 2020))

Calculating statistical variables can prove to be beneficial for increasing the coverage of the method, as regarding city blocks and enclaves can help estimate missing or erroneous height data by exploiting the dispersion and deviation of a given urban block typology. Moreover, by examining confined homogeneous areas, we see that implementing specific rules could accordingly consider erroneous or lack of height data where the presence of an architectural standard is of high probability. Depending on the scale of the confined area and the corresponding average height variance (Table 5), setting building heights of missing data to the mean will contribute to an overall better model for infrastructure analysis.

(a) (b)

Figure 33: This figure illustrates an example of where using the estimated height of the surrounding neighbouring buildings would give an accurate result on a building with no estimated height due to a non-existing or bad quality street view image. Image (a) is the generated 3D building model missing the estimated height of a building, and image (b) is the corresponding Google Maps image of the actual street. (Google, "Streetview," 3D model, Google Maps (GoogleMaps 2021), 3D model of Frode Rinnans veg 50, 7050, Trondheim, captured: May 2021)

Evaluating the large-scale viability of our method applied in the study area, particularly comparing the estimated heights and the ground truth heights gathered from LiDAR data from the Norwegian Mapping Authority, we can see a clear correlation between the height estimations and the ground truth heights (Figure 23). The apparent systematic height difference can be explained because the height estimations represent the facade height, contrary to the ground truth data, which is based on the relative height from the DTM to the highest point on the roof, often deviating from the facade height. In addition, when ascertaining the height difference results, we calculated the average height difference to be 2.91 meters when excluding the left and right tails, and only considering the height values $y \, \epsilon \, (2.5, 20.0)$. This underlines the fact that the height estimations are prone to a systematic height variation, as the roofs and various construction components are included in the ground truth data.

Furthermore, interpreting the results in Figure 25, we see by comparison that the height deviates on average with a value of 3.018 meters for the focused area. Given that we initially chose 2.5 meters as the standard floor height and assuming that the roof height was the same as the floor-to-floor height, it could be beneficial to increase our stated floor-to-floor height by a couple of decimeters to achieve a closer value to the ground truth height. Floor thickness, foundation height and other minor attributes influencing the total facade height could be factoring features causing the

deviating value. Additionally, our ground truth data may be inadequate in some cases, as the LiDAR approach will struggle with adjacent trees and other obstacles occluding the roof outline, causing erroneous ground truth values as seen in Figure 34.

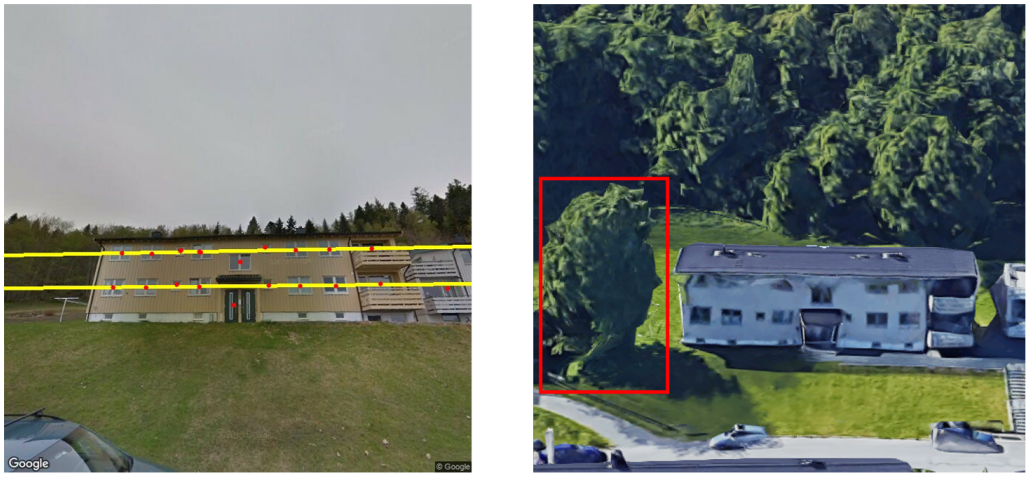


Figure 34: This figure illustrates an example showing the segmented facade (a) with an estimated height of 5 meters, and a corresponding 3D model view (b), demonstrating how the ground truth value (of 16 meters), can be assumed to be a consequence of an error as the height of the tree (red quadrilateral) was encompassed in the building footprint. (Google, "Streetview," 3D model, Google Maps (GoogleMaps 2021), 3D model of Aaengveien 9, 7020, Trondheim, captured: May 2021)

5 Conclusion and Future Work

In this chapter, we conclude our research and experiments in light of our research goals, in addition to highlighting the work that could be interesting to pursue in the future.

5.1 Conclusion

This thesis aimed to investigate the utilization of floor segmentation and rule-based height estimation on street-level imagery to enable large-scale analysis of an urban environment. A key prospect of our approach was to respond to the shortcomings of state-of-the-art methods by answering the stated research questions **Q1**, **Q2** and **Q3**.

Q1: Will the number of floors enable an adequate estimation of the facade height by using floor segmentation on street-level imagery?

Q2: Will extensive knowledge about the features of a facade aid in the estimation of the facade height?

Q3: Will extensive knowledge about the attributes of a building aid in the estimation of the facade height?

In this work, we have presented a method exploiting the properties of man-made building principles such as regularity and symmetry within facades. Based on the analysis conducted in our experimental study, we can conclude that our method provided promising height estimation results in our study area, given its cultural, temporal and topographic footprint.

The results indicate that segmenting the facade into distinct floors contributed to a realistic representation of the building and could be exploited to estimate the height and the scale of the urban scene. Therefore, to answer research question **Q1** we can conclude that our work definitely supports this presumption.

Moreover, the application of urban rules can assist in the task of increasing the quality of the height estimation. When applying relevant rules, we observed that it converged to the ground truth data. By exploiting the features and attributes inherent in facades, we gained knowledge of the building facade as a whole, yielding promising results. To answer research question **Q2**, we observed that in most cases, the height was approximated closer to the ground truth when we applied the rules. We can conclude that such knowledge can aid in improving the height estimation. However, we observed that rules were highly dependent on the quality of object detection and the correct segmentation of floors, therefore providing variable results.

Given the choice of building information source, we experienced that some building attributes were sparsely set throughout the study area, culminating in limited data and the inability of extracting extensive knowledge. Subsequently, we deduce that the results were inconclusive and are therefore unable to answer the research question **Q3**. However, our results indicate the potential that the aforementioned attributes may have had, given a higher presence of such data.

We conclude that our method for segmenting floors and estimating heights is a sufficient tool for representing the scale and character in an urban environment. However, the accuracy of the output is highly influenced by the quality of building information and street view imagery, emphasizing the significance of proper quality assurance of the pre-processing step. In addition, the availability of data and the current trends within modern architecture pose challenges that need to be considered in the future.

5.2 Future Work

In this work, we tested the method in a relatively homogeneous and concentrated area. Future work should be able to transfer the method to other regions without a significant decrease in estimation accuracy. As such, climate, culture and the inherent topography of an area must be accounted for to provide a more robust method. Furthermore, our method is unable to generate generic models where data

is non-existing, limiting the method as an open alternative in areas where cadastres are incomplete.

Therefore, future works should aim to combine open source and official cadastres to increase coverage or enable georeferencing of generic models that provide adequate estimations without relying on existing datasets. Furthermore, the utilized detection method primarily relies on the relations of doors and windows to estimate the height, providing the rudimentary architectural structure and describing the most prominent patterns. However, given the complexity of the interaction between building components, it is evident that much more information can be extracted and used to derive an extended set of rules. A good starting point would be detecting the facade outline, as it would be beneficial in properly distinguishing individual buildings. Thus, calculating mathematical properties on the facade, such as ratio, presence, and relative distance, would ease the task of describing the facade.

Bibliography

[1] Yusuf Cihat Aydin and Parham A. Mirzaei. 'A Novel Mathematical Model to Measure Individuals' Perception of the Symmetry Level of Building Facades'. In: *Architectural Engineering and Design Management* (2021), pp. 1–18. DOI: 10.1080/17452007.2020.1862042.

[2] Emmanuel P. Baltsavias. 'A Comparison Between Photogrammetry and Laser Scanning'. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 54.2 (1999), pp. 83–94. ISSN: 0924-2716. DOI: https://doi.org/10.1016/S0924-2716(99)00014-3.

[3] Susanne Becker and Norbert Haala. 'Grammar Supported Facade Reconstruction from Mobile LIDAR Mapping'. In: *Proceedings of the ISPRS Workshop on City Models, Roads and Traffic: CMRT09; Paris, France, September 3-4, 2009* 38 (Sept. 2009), pp. 229–234.

[4] Filip Biljecki et al. 'Applications of 3D City Models: State of the Art Review'. In: *ISPRS International Journal of Geo-Information* 4 (Dec. 2015), pp. 2842–2889. DOI: 10.3390/ijgi4042842.

[5] Rainald Borck. 'Will Skyscrapers Save the Planet? Building Height Limits and Urban Greenhouse Gas Emissions'. In: *Regional Science and Urban Economics* 58 (2016), pp. 13–25. ISSN: 0166-0462. DOI: https://doi.org/10.1016/j.regsciurbeco.2016.01.004.

[6] Maria Brovelli and Giorgio Zamboni. 'A New Method for the Assessment of Spatial Accuracy and Completeness of OpenStreetMap Building Footprints'. In: *ISPRS Int. J. Geo-Inf.* 8.7 (2018), p. 289. DOI: https://doi.org/10.3390/ijgi7080289.

[7] Fei Chen. 'Urban Morphology and Citizens' Life'. In: *Encyclopedia of Quality of Life and Well-Being Research*. Ed. by Alex C. Michalos. Dordrecht: Springer Netherlands, 2014, pp. 6850–6855. ISBN: 978-94-007-0753-5. DOI: 10.1007/978-94-007-0753-5_4080.

[8] Bumseok Chun and Jean-Michel Guldmann. 'Two- and Three-Dimensional Urban Core Determinants of the Urban Heat Island: A Statistical Approach'. In: *Journal of Environmental Science and Engineering* B1 (Jan. 2012), pp. 363–378.

[9] Alexis Comber et al. 'Using Shadows in High-Resolution Imagery to Determine Building Height'. In: *Remote Sensing Letters* 3.7 (2012), pp. 551–556. DOI: 10.1080/01431161.2011.635161.

[10] Bath & North East Somerset Council. *Bath Building Heights Strategy Part 3: Building Heights Strategy*. URL: https://www.bathnes.gov.uk/sites/default/files/sitedocuments/Planning-and-Building-Control/Planning-Policy/Evidence-Base/Urban-Design-Landscape-and-Heritage/BathBuildingHeightsStrategyPart3BuildingHeightsStrategy-lowres.pdf (visited on 5th May 2021).

[11] Chengri Ding. 'Building Height Restrictions, Land Development and Economic Costs'. In: *Land Use Policy* 30 (Jan. 2013), pp. 485–495. DOI: 10.1016/j.landusepol.2012.04.016.

[12] Direktoratet for Byggkvalitet. *Byggteknisk Forskrift (TEK17)*. 2017. URL: https://dibk.no/regelverk/byggteknisk-forskrift-tek17/ (visited on 21st Apr. 2021).

[13] Edward Duncan. 'Urbanization and 3D City Modelling for Developing Countries'. In: *Electronic Journal of Information Systems for Developing Countries* 54 (Aug. 2012), pp. 1–20.

[14] T. Esch et al. 'Towards a Large-Scale 3D Modeling of the Built Environment—Joint Analysis of TanDEM-X, Sentinel-2 and Open Street Map Data'. In: *Remote Sensing* 12.15 (2020). ISSN: 2072-4292. URL: https://www.mdpi.com/2072-4292/12/15/2391.

[15] Hongchao Fan, Gefei Kong and Chaoquan Zhang. 'An Interactive Platform for Low-cost 3D Building Modeling from VGI Data Using Convolutional Neural Network'. In: *Big Earth Data* 5.1 (2021), pp. 49–65. DOI: 10.1080/20964471.2021.1886391.

[16] David Frantz et al. 'National-Scale Mapping of Building Height using Sentinel-1 and Sentinel-2 Time Series'. In: *Remote Sensing of Environment* 252 (2021). ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2020.112128.

[17] P. Gong et al. 'ICESat GLAS Data for Urban Environment Monitoring'. In: *IEEE Transactions on Geoscience and Remote Sensing* 49.3 (2011), pp. 1158–1172. DOI: 10.1109/TGRS.2010.2070514.

[18] Google. *Google Street View Static API*. URL: https://developers.google.com/maps/documentation/streetview/overview (visited on 27th Apr. 2021).

[19] GoogleMaps. *GoogleMaps*. URL: https://maps.google.com (visited on 20th May 2021).

[20] Bin Jiang. 'A Complex-Network Perspective on Alexander's Wholeness'. In: *Physica A: Statistical Mechanics and its Applications* 463 (2016), pp. 475–484. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2016.07.038.

[21] N. Kadhim and M. Mourshed. 'A Shadow-Overlapping Algorithm for Estimating Building Heights From VHR Satellite Images'. In: *IEEE Geoscience and Remote Sensing Letters* 15.1 (2018), pp. 8–12. DOI: 10.1109/LGRS.2017.2762424.

[22] Jay Kappraff. 'The Proportional System of the Parthenon'. In: *Symmetry-Culture and Science* 27.4 (2016), pp. 369–391.

[23] Doug Kelbaugh. 'Typology — An architecture of limits'. In: *Architectural Theory Review* 1.2 (1996), pp. 33–52. URL: https://doi.org/10.1080/13264829609478288.

[24] Gefei Kong and Hongchao Fan. 'Enhanced Facade Parsing for Street-Level Images Using Convolutional Neural Networks'. In: *IEEE Transactions on Geoscience and Remote Sensing* (2020), pp. 1–13. DOI: 10.1109/TGRS.2020.3035878.

[25] Z. Li et al. 'Automated Building Generalization Based on Urban Morphology and Gestalt Theory'. In: *International Journal of Geographical Information Science* 18.5 (2004), pp. 513–534. DOI: 10.1080/13658810410001702021.

[26] Gregoris Liasis and Stavros Stavrou. 'Satellite Images Analysis for Shadow Detection and Building Height Estimation'. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 119 (2016), pp. 437–450. ISSN: 0924-2716. DOI: https://doi.org/10.1016/j.isprsjprs.2016.07.006.

[27] Hantang Liu et al. 'DeepFacade: A Deep Learning Approach to Facade Parsing'. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*. 2017, pp. 2301–2307. DOI: 10.24963/ijcai.2017/320.

[28] Peter Mooney and Marco Minghini. 'A Review of OpenStreetMap Data'. In: *Mapping and the Citizen Sensor*. Sept. 2017, pp. 37–59. DOI: 10.5334/bbf.c.

[29] Hourakhsh Ahmad Nia and Rokhsaneh Rahbarianyazd. 'Aesthetics of Modern Architecture: A Semiological Survey on the Aesthetic Contribution of Modern Architecture'. In: *Civil Engineering and Architecture 8 (2)* 76 (2020).

[30] Roland Olbricht. *Overpass API,* 2021. URL: https://dev.overpass-api.de/overpass-doc/en/index.html (visited on 12th Apr. 2021).

[31] OpenStreetMap. *OpenStreetMap*. URL: https://www.openstreetmap.org/ (visited on 20th May 2021).

[32] OpenStreetMap. *OpenStreetMap Wiki,* 2021. URL: https://wiki.openstreetmap.org/wiki/ (visited on 24th Mar. 2021).

[33] F. Pedregosa et al. *Scikit-learn: Machine Learning in Python*. Version 0.24. 2021.

[34] Feng Qi, John Z. Zhai and Gaihong Dang. 'Building Height Estimation using Google Earth'. In: *Energy and Buildings* 118 (2016), pp. 123–132. ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2016.02.044.

[35] Qiao, Yiqing. 'City, Urban Planning and the Creation of Urban Culture-Taking the Ancient City of Xi´an as an Example'. In: *MATEC Web Conf.* 100 (2017), p. 05072. DOI: 10.1051/matecconf/201710005072.

[36] Joseph Redmon and Ali Farhadi. *YOLOv3: An Incremental Improvement*. 2018.

[37] Eirik Resch et al. 'Impact of Urban Density and Building Height on Energy Use in Cities'. In: *Energy Procedia* 96 (2016). Sustainable Built Environment Tallinn and Helsinki Conference SBE16, pp. 800–814. ISSN: 1876-6102. DOI: https://doi.org/10.1016/j.egypro.2016.09.142.

[38] Nikos A Salingaros. 'Symmetry Gives Meaning to Architecture'. In: *Symmetry Cult. Sci* 31 (2020), pp. 231–260.

[39] M. Schmitz and H. Mayer. 'A Convolutional Network for Semantic Facade Segmentation and Interpretation'. In: XLI-B3 (June 2016), pp. 709–715. DOI: 10.5194/isprs-archives-XLI-B3-709-2016.

[40] Beril Sirmacek et al. 'Performance Evaluation for 3-D City Model Generation of Six Different DSMs From Air- and Spaceborne Sensors'. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5.1 (2012), pp. 59–70. DOI: 10.1109/JSTARS.2011.2178399.

[41] United Nations Development Programme - UNDP. *Sustainable Urbanization Strategy: UNDP's Support to Sustainable, Inclusive and Resilient Cities in Developing World*. 2016. URL: http://www.undp.org/content/dam/undp/library/Sustainable%5C%20Development/Urbanization/UNDP (visited on 18th Apr. 2021).

[42] C. A. Vanegas et al. 'Modelling the Appearance and Behaviour of Urban Spaces'. In: *Computer Graphics Forum* (2010). ISSN: 1467-8659. DOI: 10.1111/j.1467-8659.2009.01535.x.

[43] M. Zuliani, C.S. Kenney and B.S. Manjunath. 'The MultiRANSAC Algorithm and its Application to Detect Planar Homographies'. In: 3 (2005), pp. III–153. DOI: 10.1109/ICIP.2005.1530351.

NTNU

Norwegian University of
Science and Technology