ROYAL INSTITUTE
OF TECHNOLOGY

Combining Street View and Aerial Images to Create Photo-Realistic 3D City Models

Caroline Ivarsson

Master of Science Thesis in Geoinformatics
TRITA-GIT EX 14-006

School of Architecture and the Built Environment
Royal Institute of Technology (KTH)
Stockholm, Sweden

June 2014

Preface

This master thesis work is the last project for obtaining a Master's Degree in Geomatics Engineering at the Royal Institute of Technology, Stockholm, Sweden.

I would like to thank my supervisors at Blom Sweden AB, Helén Rost and Hans Strandberg, for the opportunity of carrying out my master thesis at the company and for useful advice throughout the work. I would also like to thank my supervisor and examiner at KTH Geoinformatics, Yifang Ban, for helping me finding this topic and supporting me throughout the work with comments and suggestions.

Also I would like to thank all people who has helped out during the thesis work at Blom, Acute3D, Cyclomedia and TerraSolid and to the persons who have been interviewed for the purpose of the thesis work.

Finally I would like to thank family and friends who have supported me with encouraging words and comments when motivation has been low.

Stockholm 2014-06-24

Caroline Ivarsson

Abstract

This thesis evaluates two different approaches of using panoramic street view images for creating more photo-realistic 3D city models comparing to 3D city models based on only aerial images. The thesis work has been carried out at Blom Sweden AB with the use of their software and data.

The main purpose of this thesis work has been to investigate if street view images can aid in creating more photo-realistic 3D city models on street level through an automatic or semi-automatic approach. Two different approaches have been investigated in this thesis work: using the street view images for texturing already generated 3D building models and using the street view images directly for reconstructing 3D city models. Data was collected over the study area of KTH, Stockholm, Sweden and the models were created with the two software TerraPhoto, used for texturing, and Smart3DCapture, used for reconstruction. The created models were analyzed and compared with the models based on only aerial images and the two approaches were compared to each other.

Through using also street view images when creating city models the models are shown to become more photo-realistic representations on street level compared to the models based on only aerial images. The two tested approaches create very different 3D city models in terms of what is visible in the final model, a textured buildings model or a fully reconstructed environment, and also contains different amount of involved manual work. The textured models contain only the buildings and tend to look very flat because the street images are being projected onto flat building walls whereas the reconstructed models reconstructs everything being visible in the images, trees, cars etc., and create a full scale 3D city model.

There are however limitations associated with using street view images when modeling cities since they only contain information on the ground level and no information about roofs or higher parts of the city environment and the street view images therefore need to be used in combination with aerial images. Using the street view images for reconstructing city models showed some complications in terms of wavy facades, bumpy roads and objects such as trees being inaccurately modeled. The texturing approach creates less visually pleasing models since objects such as trees and lighting poles are being projected onto the facades. The models based on aerial images look more visually appealing compared to when also using the street view images but lack the resolution be considered as photo-realistic on street level. This thesis work has shown that there is potential in using street view images when creating photo-realistic 3D city models on street level even though it is not yet a semi-automatic or automatic approach.

Sammanfattning

Denna rapport beskriver två olika tillvägagångssätt för att använda panorama gatubilder för att skapa mer fotorealistiska 3D stadsmodeller jämfört med att enbart använda sig av flygbilder. Arbetet har genomförts i samarbete med Blom Sweden AB som har tillhandahållit programvaror och data i form utav bilder.

Det huvudsakliga syftet med studien var att se om gatubilderna kan användas för att skapa mer fotorealistiska 3D stadsmodeller på gatuplan genom ett automatiskt eller halv automatiskt tillvägagångssätt. Två olika tekniker undersöktes: använda gatubilderna för texturering av befintliga 3D byggnadsmodeller och att använda gatubilderna för att direkt rekonstruera 3D stadsmodeller. Data samlades in över studieområdet KTH, Stockholm, och anpassades innan 3D modellerna skapades i programvarorna TerraPhoto, användes för texturering, och Smart3DCapture, användes för rekonstruering. Modellerna blev sedan analyserade och jämförda med modeller baserade på enbart flygbilder och sedan var de två teknikerna jämförda med varandra.

Genom att även använda gatubilder vid skapandet av 3D stadsmodeller har det visats att modellerna får ett mer fotorealistiskt utseende jämfört med modeller som är enbart baserade på flygbilder. De två tillvägagångssätten skapar olika typer av modeller i avseendet av vad som finns synligt i modellerna, en texturerad byggnadsmodell eller en helt rekonstruerad miljö, samt att olika mängder av manuellt arbete är involverade i de olika tillvägagångssätten. Den texturerade modellen innehåller enbart byggnader och tenderar till att se väldigt platt ut eftersom gatubilderna blir projicerade på byggnadsfasaderna medan den rekonstruerade modellen innehåller allt som finns synligt i bilderna, träd, bilar etc. och rekonstruerar en 3D stadsmodell som även innehåller den omkringliggande miljön.

Det finns dock begränsningar i användandet av gatubilder och den främsta är att de är tagna på gatunivå vilket innebär att det inte finns några bilder över varken tak eller högre delar av stadsmiljön och gatubilderna behöver därför användas i kombination med flygbilder. Att använda gatubilder för rekonstruktion visade sig innebära problem i form av ojämna fasader, gropiga gator och att objekt som träd blir felaktigt modellerade. Den texturerade modellen skapar visuellt mindre tilltalande modeller som ser väldigt platta ut och med objekt projicerade på fasaderna. Stadsmodellerna som enbart är baserade på flygbilder ser visuellt bättre ut jämfört med när även gatubilder används men den låga upplösningen gör att modellerna inte kan ses som fotorealistiska på gatuplan. Det här arbetet har visat att det finns potential i att använda gatubilderna för att skapa fotorealistiska 3D stadsmodeller på gatuplan även om det ännu inte är ett halvautomatiskt eller automatiskt tillvägagångssätt.

Table of Contents

List of Figures

1 Introduction

1.1 Background

During the last few years the evolution of 3D city models have received an increasing interest with an increased number of applications that use 3D information of topographic objects (Elberink & Vosselman, 2006). 3D city models are no longer only being used by companies and municipalities, but regular people want to view and use 3D models and 3D maps to get a better understanding of the environment they are living in.

A 3D city model can be defined as a digital representation of the earth's surface and its related objects such as buildings, trees, vegetation and manmade features belonging to the urban area (Singh et.al, 2013). An accurate and immersive virtual reality city model will enable the users to better visualize their living spaces in a more interactive and efficient way than before (Tan et.al, 2012). All these factors depend on the accurate reconstruction of both the 3D terrain and building models as well as the texturing of model surfaces. 3D city models can aid applications such as urban planning, training, disaster simulation and virtual-heritage conservation (Früh & Zakhor, 2003).

Most available 3D city models are based on aerial images, either as the basis for reconstruction or used as textures, which creates low resolution models that do not appear photo-realistic when it comes to viewing the model from a street level perspective. Using aerial images for city modeling is good since the images cover a large area but is restricted in the sense that smaller buildings are harder to recognize in images with low resolution (Amat et.al, 2010).

Services such as Eniro and Google Street View are using 2D panorama images at static viewing points within their map services to provide a more realistic view of the street level of a city but without having to generate detailed photo-realistic 3D city models (Google Street View, 2014; Eniro, 2014). These services help people getting a better understanding of what the city environment looks like at different locations even if not being at the specific location. Creating street level city models is one of the most fundamental applications of 3D city models since many of the daily activities such as navigation and travelling takes place on street level and it has received an increasing interest from both industry and academic fields (Mao, 2011).

The company Blom Sweden AB are using the technology of Dutch Cyclomedia to collect street view images that are being used in e.g. their application BlomSTREET and in the Eniro street view service (Blom, 2014a; Eniro, 2014). The images are high-resolution street view panoramic images and a large library of images are available throughout the country of Sweden. This thesis will investigate how this library of street view images can be used to create photo-realistic 3D city models through a semi-automatic or automatic approach.

1.2 Existing Problems

Most commercially available models takes months to create and usually require significant manual intervention. This process is not only expensive but also unsuitable in applications where the goal is to monitor changes over time (Frueh & Zakhor, 2001). Municipalities using 3D models want the models to be generated quickly since they want the models to be up-to-date when being delivered, it is not desired to wait for a long time for a 3D model that is not up-to-date (Eskilstuna Kommun, 2014). Cities are constantly changing and the models need to be generated in a fast and automatic approach in order to be up-to-date. Creating street level city models from aerial images are not suitable since the aerial images cannot produce photo-realistic models at ground level (Xiao et.al, 2009).

Most approaches existing to create high level of detail models from ground view images involve enormous amounts of manual work although the results are visually appealing. The manual work can include imports of geometries obtained from construction planes or selecting primitive shapes and corresponding points for image-based modeling. Attempts have been made to use image-based or 3D laser scanner approaches to collect data in an automatic approach but the acquisition process is slow since it only scales to a few at the same time because they acquire data in a slow, stop-and-go fashion (Früh & Zakhor, 2003).

Blom Sweden AB delivers 3D city models based both on laser scanning- and photogrammetric techniques. Up until today they have only used aerial images for texturing and reconstruction 3D city models they deliver to their customers. The existing library of street view images are currently being used in e.g. the BlomSTREET application (Blom, 2014a). The company now wants to look at the possibilities of using these panoramic street view images in combination with the aerial images to deliver more photo-realistic 3D city models to their customers.

1.3 Objectives

The objective of this thesis is to investigate how the street view images available at Blom Sweden AB can be used in a semi-automatic or automatic approach to increase the level of detail of the 3D city models they deliver. The improved appearance of the models will be compared to models based on solely aerial images. Two different approaches will be used, both having the goal of adding photo-realistic detail to 3D city models through the use of street view images in a semi-automatic or automatic approach.

- The first approach deals with using street view images to texture already existing 3D building models.
- The second approach will investigate if the street view images are suitable for reconstructing 3D city models directly.

Both approaches will be studied with use of software being available at the Blom Sweden AB office, texturing with TerraPhoto and reconstruction with Smart3DCapture. Once the models have been created it will also be discussed how such models can be used and what limitations they are associated with.

2 Literature Review

As start for this thesis work, a literature review was done in order to see what applications, definitions and techniques exist within the field of 3D city modeling. The goal was also to see what research has been performed in using street view images to create photo-realistic 3D city models on street level. It should be mentioned that a lot of work has been performed within the field of 3D city modeling based on images, including street view images, and this literature review aims at focusing on the techniques relevant to this thesis work. The goal has not been to create a complete review of the research field.

First the use of 3D models today will be described followed by an insight to what the use of 3D city models with street view images could be used for. Then a brief introduction to some existing software to create 3D city models based on or textured with images will be given and finally some research is presented that relates to creating photo-realistic 3D models.

2.1 3D City Models

When talking about 3D city models one commonly refers to the detail of the model as the level of detail (LOD). There are various ways of defining the different LODs and in this thesis the definition from Blom Sweden AB will mainly be used. They deliver 3D models in four different LODs, ranging from wire frames to fully textured models (Blom, 2014b):

- LOD 1 Building blocks minus roof details
- LOD 2 Building blocks with large scale roof detail with approximation of color derived from LOD4 details
- LOD 3 Building blocks with large scale roof detail plus standard format library textures based on color and representation of real features
- LOD 4 Building blocks with large roof details plus Photo-realistic textures from aerial images

The article *Virtual 3D City Modeling: Techniques and Applications* gives a summary of some techniques for creating 3D city models and existing applications of such models. 3D city models are being used in a wide range of fields such as planning and design, infrastructure and facility services, commercial sector through marketing and promotion and learning information on cities (Singh et.al, 2013). What technique is being used to create the city models depends on the application and the required level of detail of the resulting 3D city model. The number of applications have increased rapidly the recent years with examples being found in location based services, virtual reality tasks, visualization for city planning etc. (Frueh & Zakhor, 2001).

Landmarks act as parts of 3D city models and a landmark can be defined as an outstanding virtual 3D model of a building (Schulze-Horsel, 2012). 3D landmarks can be e.g. churches, famous historical objects or other important buildings that are popular in navigation and web-mapping applications. The landmarks can be created from close-range photogrammetry, terrestrial laser scanning or 3D modeling, the same techniques that can be used for modeling entire cities. To make the landmarks look photo-realistic, texture mapping is usually performed after the modeling. Applications of such landmark models are urban planning, architecture visualization, real estate management, navigation systems and web-based visualization tools for e.g. tourism. Adding multiple such landmarks creates a 3D city model.

A wide field of users of 3D city models are municipalities. Patrik Johansson, working as a 3D model engineer at Eskilstuna municipality, explains that 3D city models are very important in the dialogue with the citizens of the city of Eskilstuna. They have a digital model over the city as well as a physical model in scale 1:400 covering the whole city. Both models are used for the same purposes even though they serve different citizens. The physical model serve well for blind people who can feel the city structures with their hands as well as people not so experienced with computers that can interpret a physical model more easily. In the municipality they use the 3D models for visualizing what the urban environment would look like if a new building was placed at a certain location. Also sunlight-, shading- and sight- analyses are performed by means of the 3D city models. Some analyses are easier to perform with the digital 3D model, but both models serve as important tools for the municipality. The digital city model that Eskilstuna is using today is based on 2D drawings where the buildings have been extracted to a height values based on laser data. In the areas where they have had new construction projects in the city, some buildings have received textures that have been acquired by a digital camera and then added manually to the building models. It is a tedious process and the municipality is looking at buying a fully textured 3D model to use within their organization. They see a need for a fully textured city model to improve the dialogue with the citizens. Such a model would also help the municipality save a lot of time and it is desirable to have a high-resolution textured model covering the whole city. The model should preferably be as detailed as possible and fully textured, something that they at the moment not have the time to create themselves. At the moment what they are looking for is a model that can work for visualizing and they do not want to have to wait for a long delivery (Eskilstuna Kommun, 2014).

There are other types of businesses such as visualization companies working with 3D city models. One such company located in Stockholm is SightLine Vision AB. Their main focus group is construction- and real estate- companies wanting to visualize new constructions or building sites. It could involve attracting buyers for apartments or to convince residents to vote in favor of a project. In their delivery products they usually work with high LOD models in the center of the visualization, where the main focus of the model lies, and when moving further away from the center the LOD of the model decreases. Right at the center are high resolution textured building models based on images they acquire themselves with a digital camera. This is tedious work but it is the only way they know to receive the high level of detail they need in the city models they deliver. If such a detailed textured model could be created automatically they would like to use it but they see problems with automatic or semi-automatic generated 3D models and the presence of trees and other objects in the models. When they create models they want clean building models with empty streets and clean facades and then they want to add trees etc. to the models themselves. It is also important that the textures are acquired perpendicular to the facades, to avoid perspectives being present in the textures, and that there is nothing besides the wall texture on the building façade in the model. However, as they move away from the center of the visualization they want to use a model with a lower LOD. They want the focus to lie on the center and what is lying around is not as interesting and should not take focus from the center, they think that an automatically created model could serve as the level outside of the visualization center. Since a city not only consists of buildings SightLine also has objects like bridges, roads and tunnels in their models and they are not as easy to create or texture manually through only using digital images. Also in that case it would be interesting to use an automatically generated model to receive such objects. SightLine could see themselves working with a fully automatic generated textured 3D city model, their only concern is how detailed the model would be and if it would be suitable within their business and the accuracy that they need when they deliver the models to their customers (SightLine Vision AB, 2014).

The company Agency9 is working with web and mobile solutions for 3D visualizations of geographical data. They do not create any 3D models themselves but rather work with visualizing city models for municipalities and companies through their own visualization software. They are under the impression that customers want the city models to be as clean as possible and with a high level of detail and resolution so that one can focus on what is being visualized in the model and not the 3D model used for visualizing. But they also think it depends on the purpose of the model. If it is being used only internally within the company a lower level of detail is probably acceptable but if it’s being used to communicate with citizens a model with low resolution might lower the impression and reliability. Johan Göransson, product manager at Agency9, has seen street view images being used for texturing 3D city models before but says that the results have looked very flat, plain and not so lively. He says that it is important to think of how trees and other structures will be visualized in the model, when using images to texture it would be important and desirable to filter out or remove trees from the building textures and also it is important to think about illumination and shadows in the images. He is also worried that creating high resolution models will increase the model size needed to save and render a 3D city model, it is important not to lavish with geometry triangles. Johan Göransson explains that many of their clients want to perform shading and sunlight analyses and in those cases it is not good to have those effects already in the used city model (Agency9, 2014).

2.2 Different Approaches to Create Digital 3D Models

There are various ways of creating 3D city models and the main three approaches can be read about in the article written by Singh et.al. (Singh et.al, 2013). The methods used to create 3D city models are generally being divided into three different groups based on the production technique; automatic, semi-automatic and manual generation (Kobayashi, 2006). An automatic approach consists of extracting buildings, trees etc. from aerial or satellite images by using technologies of image processing and pattern recognition in artificial intelligence. The semi-automatic approach is to create the 3D objects one by one with support of technologies such as photogrammetry and 3D vision. The manual approach is to create all geometries of an object one by one by using a software, e.g. CAD. Since there are many methods of 3D modelling and many applications of 3D models it is important to choose the most suitable method for the specific application (Kobayashi, 2006). Besides these three approaches to creating 3D models another important technique is that of texturing 3D models with high resolution images to generate models with more photo-realistic appearance (Schulze-Horsel, 2012).

2.2.1 Extrude 2D Maps

Creating 3D models based on 2D maps is considered being a manual approach since the geometries are being created manually one by one. A 2D map can for example be a CAD drawing that gets extruded based on an assigned height value for each respective building polygon or by using laser data to extrude the buildings (Elberink & Vosselman, 2006). This can be a time-consuming procedure that involves full manual control of the creation process. But it is at the same time a common approach since many municipalities and companies have access to 2D maps covering their area of interest as base information which they are making use of when generating their 3D city models (Eskilstuna Kommun, 2014).

This technique is for example the basis of ESRIs module City Engine, used to create 3D city models (ESRI, 2014). In the software a 2D map is imported and used and by applying created rules to the 2D drawing and its adherent attributes a 3D city model is generated. This procedure requires much manual work at the start of the modeling process when writing the rules and attributes for generating the buildings, but after writing all the rules and creating all attributes the modeling becomes a more or less automatic process when applying it to the same type of buildings. How complex the buildings turn out depends on how complex the written rules are.

2.2.2 Laser Scanning Method

LiDAR (Light detection and ranging) is a common technology used to extract 3D geometries and it is based on the principle of the LiDAR instrument transmitting light to a target and measuring it by using the reflected signals (Kobayashi, 2006). The point clouds extracted from the LiDAR data are then converted into polygons. An advantage of using laser scanning methods is that no camera estimation parameters or feature detection are needed in order to obtain 3D geometries (Früh and Zakhor, 2003).

Both aerial and terrestrial laser scanning techniques can be used for creating 3D city models. Aerial laser scanning methods enable geometric acquisition of terrain surface including objects such as trees and buildings. To create a 3D city model it is however not enough with only using the aerial laser scanning model, but also an interpretation of the surface model, such as finding planar surfaces, is required to obtain a usable 3D city model (Haala & Brenner, 1997).

Terrestrial laser scanning approaches are slow since it does not scale to more than a single or few buildings because they acquire data in a slow, stop-and-go, fashion (Früh and Zakhor, 2003). The terrestrial laser scanning data points though make it possible to extract complicated geometries like architectural components on the building walls which might not be visible from aerial laser scanning (Kobayashi, 2006).

2.2.3 Photogrammetric Method

Photogrammetry is the concept of deriving metric information about an object through measurements made on photographs of an object. All types of images, satellite, aerial or close-range, can be used. The fundamental task is to establish the geometric relationship between the images and the object as it existed during the capturing moment (Mikhail et.al. 2001).

The photogrammetric approach to generate 3D models means using images to reconstruct 3D surfaces and in order to do so at least two images covering the object to be modeled are needed. The photogrammetric approach can be considered as semi-automatic since it involves manual digitizing on images and then the height determination is performed automatically (Amat et.al, 2010).

The different types of images have different advantages and disadvantages. Close-range images can be acquired directly with a regular digital camera and are usually not expensive to acquire. There are restrictions when it comes to acquiring close-range images due to security and privacy issues and many images are needed in order to cover the same area as an aerial image would do. Having several smaller images also makes it harder to match images with different color balances (Kobayashi, 2006). Aerial or satellite images are advantageous since only a few images are needed to cover a large area but they are usually more expensive to acquire (Kobayashi, 2006). Using aerial images it is more difficult to recognize small buildings and other small objects due to the resolution of the images (Amat et.al, 2010).

2.2.4 Texturing 3D Models

Textures are patterns that do not belong to the property of a scene but is a property of the imaged surface (Mikhail et.al. 2001). Texturing can be done by e.g. adding a pseudo texture or by assigning a photo to the 3D model by using an affine texture mapping. Using pseudo textures provides a better result than using no texture but it does not portray the true characteristics of the building façade (Tan et.al, 2012). Texturing a model with images creates a photo-realistic model which looks much like the reality which the model is representing. When using images for texturing they should preferably be acquired perpendicular to the façade in order to avoid distortions being present in the images. Optimally the images are also rectified and cleaned from objects such as trees before being used for texturing (Schulze-Horsel, 2012).

How the texturing is done depends much on the type of images used and also on the software used. Commonly one image is not enough to cover the whole building façade and image mosaics or stitched images are needed. Some different techniques are described in the article *Polygon-Based Texture Mapping for Cyber City 3D Building Models* written by Tsai & Lin (Tsai & Lin, 2006). The most novel technique is said to be that of using video sequences to create wall textures. Another popular technique is using panoramas made from stitched images. Problems existing when texturing 3D buildings are that the images are not necessarily acquired during the same viewing conditions (e.g. viewing point, zoom factor, looking angle), they are of assorted perspectives, scales, brightness, contrasts, color shadings and other properties that need to be adjusted before they can be used for creating mosaics and panoramas (Tsai & Lin, 2006).

2.3 Existing Software for 3D Models

There are various existing software available to create 3D models based on photogrammetric methods or software that can be used for texturing 3D models. In this section some software available for the two different approaches will be presented.

2.3.1 Reconstructing 3D Models from Images

There exists several software for generating 3D city models based on images by only using photogrammetric techniques. Some of them are: Saab Vricon Systems (Saab, 2014), 123D Catch (123D Catch, 2014), Nverse (Nverse, 2014), nFrames (nFrames, 2014) and ShapeQuest (ShapeQuest, 2014). They all build on the principles of photogrammetry and requires only images of different types as input.

Another software used for reconstruction of 3D city models is Smart3DCapture, a commercial software from the French company Acute3D. The product is currently being used by companies such as Nokia, Blom and Asia Air Survey (Acute 3D, 2014a). The software uses high resolution images as input and without strong human interaction it generates 3D city models. Directly when producing the 3D model it is being textured with the images used to reconstruct the geometries. Smart3DCapture also has the ability to remove objects that are not visible in the majority of the images covering the same spot. Moving cars will for example be removed from the model whereas parked cars will remain in the final model. As input the software takes images with corresponding orientations and positions. If no orientation and position data are available the software estimates these parameters during the aero triangulation. It is also possible to manually add control points with known coordinates in the image set in order to improve the location and orientation.

Figure 1 Screenshot of model of KTH, reconstructed from aerial images in Smart3DCapture.

2.3.2 Texturing 3D City Models

There exists several software that can be used to texture already existing building models, examples of such software are: Sketchup (SketchUp, 2014), ESRI CityEngine (ESRI, 2014) and Photoshop (PhotoShop, 2014), a few of the software also have other functions besides texturing. Some software perform the texturing process more or less automatic whereas in some software it is necessary to manually attach the texture to the corresponding wall.

Another software that can be used for texturing is TerraPhoto. TerraPhoto is a software belonging to the TerraSolid software family (TerraSolid, 2014). The TerraSolid software are mainly used for processing airborne and mobile LiDAR data and images with most of the applications being built on top of Bentley Software. TerraPhoto can be used to automatically project textures onto building walls with help from the adherent orientations and positions, it is also possible to manually change and add textures to the building walls.

2.4 Related Work

The problematic involved in creating a photo-realistic 3D city model has been addressed by many researchers through different projects. This section will present some of the related work aiming at producing and generating photo-realistic 3D city models, some of them through the use of street view images.

2.4.1 Integration of Aerial and Close-Range Images

The authors Amat et.al. describe an approach to integrate aerial and close-range photogrammetric methods in the article *Integration of Aerial and Close-Range Photogrammetric Methods for 3D City Modeling Generation* (Amat et.al, 2010). In the article the authors describe the problems of 3D city models being based on aerial images only reaches, what they defined as, LOD 2 and they propose a method of producing 3D buildings based on close-range photogrammetric methods to be used in combination with a model based on aerial images. Limitations in the aerial images lie in detecting smaller buildings and to accurately reconstruct geometries and architectural instalments like windows, doors or balconies which are commonly not visible in aerial images. In the paper aerial images are used in order to produce a DTM as base for the 3D city model whereas the close-range images are used for the 3D buildings reconstruction. Also objects like streets, trees, parking lots and temporary buildings are extracted from aerial images. The close-range images used in the project were captured by a digital camera which was calibrated before the capturing in order to define camera parameters and to provide accurate acquisitions. The images were captured with an angular separation of 90 degrees and with overlap between the images in order to be able to choose reference points between the images. The authors used the Software PhotoModeler for processing of the images and to create the 3D building models. When the 3D buildings were digitized, the models were scaled and oriented into the correct coordinate system before they were exported. The result of the project show that the use of close-range photogrammetry increases the LOD of buildings from LOD 2 to LOD 3. It is also concluded that using data from close-range photogrammetry is needed in order to develop photo-realistic 3D city models. The authors also point out that the accuracy of the close-range photogrammetry depend on the camera resolution, quality of camera calibration, geometry of camera position and the precision of marking location on the images.

2.4.2 Combing Laser Scanning Technologies with Photogrammetry

The authors Früh & Zakhor investigate the use of aerial images and ground views in the article *Constructing 3D models by Merging Aerial and Ground Views (*Früh & Zakhor, 2003). The article is a continuation of their earlier paper describing 3D mapping on street level based on ground view images and laser scanning data. Problems with the earlier research is that the proposed approach did not have the ability to model the building roofs since only ground images with adherent laser data were used and this paper intends to use both aerial and ground view images in a combination to produce a model accurate at both aerial and ground level. Two different models with different resolutions are created, each model being based on one type of images only. The aerial model contains roof information from a DSM (Digital Surface Model), being based on aerial laser scanning data, and also information regarding the terrain. The ground model is based on laser data and ground images acquired by a camera and two scans mounted on top of a car. The car uses an algorithm to estimate relative vehicle positions according to the horizontal scans performed. This is done in favor of using GPS (Global Positioning System) due to the sometimes limiting accuracy of GPS systems in the city environments. The two models with different resolutions are then registered by using a Monte Carlo localization of the laser scans. Since the ground based model has higher resolution it is given priority compared to the aerial model when choosing what model to use when they overlap, the aerial model is given priority when it comes to roofs and the terrain. The two models do not match completely at all times and therefore the authors have developed a procedure of blending the two meshes to fit together. The combination of the aerial and ground based images give photo-realistic models but the disadvantage of the approach is the limiting view of the ground based images. For some facades a combination of both aerial and ground based images are used which is visible as a line on the façade. The suggested approach also has limitations when it comes to handling foreground objects such as cars and trees.

2.4.3 Texturing Buildings in a Large Scale

The authors of the article *Large Scale Texture Mapping of Building Facades* show that it is possible to texture buildings in a large scale using oblique images (Tan et.al, 2012). The buildings were extracted from satellite images but the images have too low resolution to serve for texturing the generated buildings. For texturing oblique close-range images were acquired with a commercially available digital camera. In the project the authors have been focused on transforming images from the image space to building model space. This has been done through a reverse mapping method that transforms the building façade to fit within the image according to the image's extrinsic and intrinsic parameters. After the texturing of the buildings has been carried out the texture mosaics were color balanced in order to make the textured model look visually correct across borders between different images. The color balancing of the close-range images was performed through histogram matching of the images used in the mosaics. The authors conclude the study by saying that texturing the building models with close-range photogrammetry improves the visual effect efficiently and accurately, but using oblique images will cover larger areas but at the same time involve more skew distortions which will need to be considered when using them to texture buildings.

Tsai & Lin suggest a polygon-based approach for texture mapping of 3D building models in their paper from 2006 (Tsai & Lin, 2006). In their paper they deal with the problem of color changes in images as well as the problem of projecting texture from image space to the object space. Their proposed method suggests a hybrid approach of image adjustment, merging, blending and mapping transformation. After image acquisition the images are pre-processed with the goal of transforming a group of photographs pertaining the same object façade to a common texture space. The images are also polygonized into separate areas of interest. The images were then registered through a manual matching process of defining tie points and finally adjusting for irregular-shaped façade texture images. The developed method is a robust method but with the disadvantage of long processing time and low accuracy of the created image mosaics. Another bottleneck the authors mention is the identification of tie points between the images and that the method is yet not robust for modeling larger scale city models.

2.4.4 3D Modeling Based on Street View Images

Many researchers have investigated the possibilities of using street view images in order to create photo-realistic 3D city models. Here are three different papers described.

Xiao et.al. propose an automatic approach to generate street-side 3D models in the article *Image-Based Street-side City Modeling* (Xiao et.al, 2009). Their approach suggests a multi-view segmentation method that recognizes and segments each image at a pixel level into five different areas through a supervised learning method. The whole sequence of segments are then partitioned into independent buildings blocks using vertical and horizontal lines visible in the images, an assumption is made that all buildings have rectilinear facades. The façade modeling is proposed to be done by an inverse patch-based orthographic composition and structure analysis method that regularizes the missing and noisy 3D data. The approach shows promising results in modeling textured photo-realistic building blocks but has limitations in that it is based on assumptions that all facades and windows are rectilinear. This is not always the case when it comes to more complex buildings or landmarks. Another limitation is said to be that the images used for this paper are not always covering the top of the building facades and neither the roofs.

Luhmann & Tecklenburg discuss the potential of using street panorama images directly for 3D reconstruction of objects in the paper *3-D Object Reconstruction from Multiple-Station Panorama Imagery* (Luhmann & Tecklenburg, 2004). It is suggested that by using panoramic images less images are needed for reconstruction and fewer tie points are needed when performing the bundle adjustment. It is however important for reconstruction purposes in this case that not all panorama images are acquired from the same height since that leads to horizontal epipolar lines which cannot be used for measurements of horizontal object features. The main benefit of using panorama images is that less images are needed for reconstruction and thereby also fewer camera stations when acquiring the images.

In the article *Piecewise Planar City 3D Modeling from Street View Panoramic Sequences* the authors Micusik and Kosecka investigate the possibilities of using panoramic street view sequences to create photo-realistic city models (Micusik & Kosecka, 2009). Like Xiao et.al. they suggest a priori image segmentation as well as presence of dominant scene orientations and piecewise planar structures for the reconstruction process. Panoramic images were constructed from four perspective images with 127 degrees of vertical field of view since the top does not contain much information. The images are segmented into so called superpixels, a larger semantically meaningful object, based on a depth algorithm using five consecutively images. Each superpixel receives a depth value and a normal giving minimal photo consistency error when projecting it onto other views while considering smooth changes of the depth and normal of the neighboring superpixels. Lastly they introduce a depth fusion algorithm that combine advantages from volumetric- and viewpoint-based fusion methods by choosing a middle image in a sequence as the reference and the projecting the filtered 3D points into it. In the article two different experiments are performed with the results of textured triangulated surface meshes. The drawback of the approach here lies in, as with the method developed by Xiao et.al. the use of the superpixels is limited at places not described by local planar patches.

3 Methodology

The aim of this thesis is to look at the possibilities of creating photo-realistic 3D city models through semi-automatic or automatic approaches using street view images that can be used by municipalities and companies etc. in their organizations and activities. To investigate the possibilities of creating such models the following methodology is proposed.

As a first step data will be collected to be used in the process of creating the 3D city models. The data used is data available and already collected by Blom Sweden AB. The study area for this thesis is mainly a small area covering the L-building and its surroundings at The Royal Institute of Technology (KTH), Stockholm. Other small areas are to be used as a complement to this main study area. The two types of images that need to be collected are street view images and aerial images. When data is collected 3D city models will be created using the two proposed approaches with help from software being available at the company of Blom Sweden AB.

The first approach involves investigating if the panoramic street view images are suitable for texturing existing 3D building models and what limitations this approach withholds. The 3D city model being textured is a buildings model created from laser scanning data in the software TerraScan. The texturing approach will be tested with TerraPhoto, a software belonging to the TerraSolid family.

The second approach to be investigated is if the street view images are suitable for reconstruction of the 3D geometries directly. This will be investigated through use of the image matching software Smart3DCapture.

The results of the two approaches will be visualized and analyzed in solitude as well as in comparison to each other when it comes to accuracy, how much detail the street view images add to the city models and the limitations that each approach is associated with.

3.1 Data Collection

The data used in this thesis work mainly consists of two different types of images, panoramic street view images, so called cycloramas, and aerial images.

The street view images are high-resolution, georeferenced 360 degree panoramic images captured at street level by Cyclomedia (Blom, 2014a). The images are acquired by two fisheye cameras mounted on a specially equipped car and it captures images approximately every 5 meters. Each camera acquires an image with 185 degrees viewing field and the two images are acquired at the same location, the front-looking camera acquiring its image before the back-looking camera, and stitched together into a 360 degree panoramic image, see figure 2. The cycloramas are high-resolution images of 4800 x 2400 pixels and they are all oriented and georeferenced into existing geodetic reference system in plan and elevation. The resolution of the images depend on the distance to the object, at 10 m distance the approximate resolution is 0.013 m and at 50 m distance the resolution is approximately 0.065 m (Kart & Bildteknik, 2011). The position of the vehicle is updated through a GPS system with an approximate accuracy of 0.10 m. Blom are currently using these images in their product BlomSTREET in which measurements, road inventory etc. can be made (Blom, 2014a). The images are also being used in the Eniro street view service (Eniro, 2014).

Figure 2 Example of a Cyclorama. Source: Cyclomedia

Since neither of the two software to be used could use the cycloramas directly central perspective images were extracted from the cycloramas, see example of a central perspective image in figure 3. From each cyclorama eight central perspective images with horizontal and vertical opening angles of 90 degrees were extracted, which leads to 50% overlap between each image taken from the same acquisition location and in each direction. The image directions were chosen based on the heading of the images, separated by 45 degrees to receive eight images. For the KTH area this image set consisted of approximately 1330 central perspective images captured along Drottning Kristinas väg. All images were extracted with the metadata of orientation (heading, pitch and roll), positioning (latitude, longitude and height), image size and image name.

Figure 3 Example of a street view image, central perspective image. Source: Cyclomedia

The street view images used were all acquired during the summer of 2013 which means that trees and vegetation are fully bloomed, which can be seen above in figures 2 and 3. This means that trees and vegetation are covering some parts of facades which might not be desired when using the images for texturing and creating 3D models. The images are acquired this time of the year due to creating a more lively and realistic environment that most customers want when viewing and using the images in BlomSTREET.

The aerial images used in this thesis are also acquired by Blom Sweden AB and consists of both vertical and oblique images acquired with a 5-camera system (Blom, 2014c). The oblique images also provide information of the sides of the buildings since they are taken with an angle of approximately 45 degrees and can therefore be useful when texturing since they also cover large parts of the building facades. The images were not acquired for the purpose of reconstructing or texturing 3D models but to be the basis of the Blom application BlomOBLIQUE, used to monitor, make measurements etc. in georeferenced aerial images. The image set used consists of 308 images covering the area of KTH and Östermalm and were all acquired in August 2013.

3.2 3D Modeling Using Street View Images

To create the models based on the aerial and street view images the two different approaches as described above will be used. In this section the approaches are described more in detail.

3.2.1 Textured 3D Model

The textured model was created using a vector buildings model based on aerial laser scanning data that was textured with the two different image sets. The model was first textured with only the aerial images automatically based on the input orientation and location data of the images. The camera positions of the acquiring camera were adjusted through defining tie points in the image dataset. After the automatic texturing process not all wall textures looked visually well or matched the rest of the textures and in those cases the textures were changed and modified to receive a better visual appearance of the entire textured 3D model.

When this was done the street view images were used for texturing the same buildings model as the one textured with the aerial images. Since the software does not take panoramic images as input the central perspective images were used for texturing. The street view images do not cover the entire area of the buildings model and therefore the aerial images were used as basis for texturing with street view images. Meaning that only the walls that had a corresponding street view image was textured with them, the rest of the facades continued having the aerial image as wall façade texture.

3.2.2 Reconstructed 3D Model

The reconstructed models were created based on aerial and street view images with adherent position and orientation information. A reconstructed model based on only aerial images was first created over the entire study area. Thereafter a model based only on street view images was reconstructed through the same procedure as with the aerial images. Central perspective images extracted from the panoramic images was used also in this approach since the cycloramas could not be used directly.

Finally a model was reconstructed through combining the aerial and street view images directly in order to see what a combination of the two image sets would look like.

3.3 Used Software

The software used to investigate the two approaches are software being used by Blom Sweden AB within their 3D city modeling projects. These two software have therefore been the only two used for creating the 3D city models during this thesis work.

3.3.1 TerraPhoto

TerraPhoto is specifically developed for processing images captured together with laser data during a survey mission. But the software also has the capability of texturing with images that were not acquired during a survey mission as long as the camera parameters and the position and orientation of each image are known. It is also possible to refine the relative orientation between the input images through defining tie points in the image set and the absolute orientation of the image dataset can be adjusted through the use of known control points (TerraSolid, 2014).

The software offers possibilities for texturing 3D building models with both aerial and mobile images. Texturing with images can be done through connecting the images with tie points in order to refine the relative orientation among the images in an image set. After this often tedious process, depending on the amount of images, the building facades can be textured automatically. The software chooses only one image that covers the largest part of each façade and that is acquired the closest to the building façade, it takes no consideration of the angle from which the image has been captured. After this automatic process there are various possibilities of how to change and refine the texture that the software has chosen.

One of those possibilities is the use of depth maps when texturing. Depth maps are created from classified laser data, e.g. vegetation, buildings, ground, and the software then calculates depth maps that show the distance from the camera to the closest point in that cell, see example of a depth map in figure 4. It is also possible to create depth maps including building polygons. The use of depth maps can help to removes parts of the textures covering a wall that are being covered by e.g. trees and replace that part with a different texture not covered by trees, see example in figures 5 and 6. This kind of stitching is however only possible if there is a part of the façade missing in one texture, the software always tries to choose only one image that covers the whole façade when texturing.

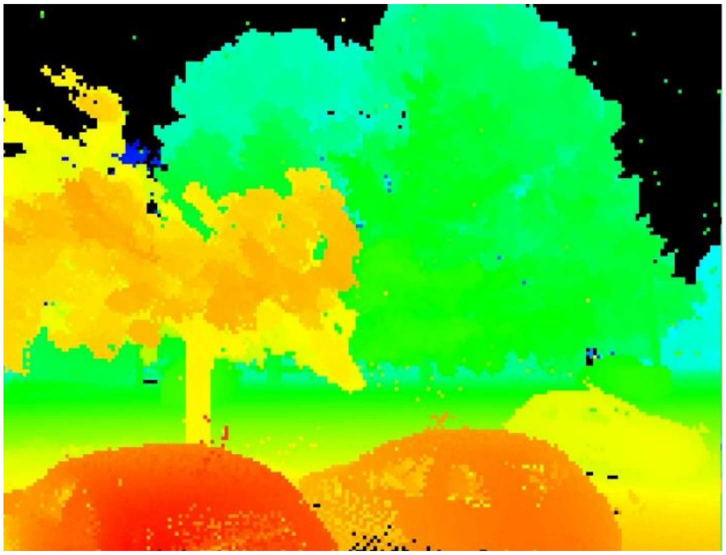

Figure 4 Depth map created in TerraPhoto

Figure 5 Screenshot in TerraPhoto, hole in the façade due to the use of depth maps

Figure 6 Screenshot in TerraPhoto, stitched texture of the façade

TerraPhoto offers a wide variety in what can be done when texturing building facades with either aerial or street images. The same functions are available for both image types, but the images cannot be used as a combination directly for texturing the buildings since they are collected by different cameras with different parameters, the texturing therefore needs to be performed in two steps. It should also be mentioned that this texturing procedure is only done for texturing facades of buildings and not for texturing roofs. Texturing the roofs of a model in TerraPhoto is done through the use of orthophotos that can be generated in TerraPhoto and the orthophotos can also be draped on a ground surface model in which the buildings can be visualized within the software.

3.3.2 Smart3DCapture

The software Smart3DCapture can be used to automatically reconstruct 3D models based on only images. The images can be either aerial or street view images even though the software today is mainly developed for the use of aerial images. The images are imported as blocks through an XML-file that specifies filename, coordinate system, position and orientation of each image. All this adherent information is not required, but it gives a better approximation when performing the aero triangulation. If the positioning and orientation information is not provided it is estimated during the aero triangulation and based on the image matching. The positioning information can also be used through the use of GPS tags that exists in the Exif metadata of the image or through the use of control points (Acute3D, 2014b).

The images are matched globally by detecting pixels that correspond to the same physical point among all images. Within each block several photo groups may exists which are acquired by different cameras and therefore also have different internal parameters. The two internal camera parameters focal length and sensor- or lens size are required as input for all photo groups that are used for reconstruction. When all images are imported an aero triangulation is performed where all camera positions are estimated and adjusted, this can be done, as mentioned above, either with or without input positions and orientations (Acute3D, 2014c).

Tie points are generated in the aero triangulation corresponding to all the pixel correspondences between the images. These tie points are then used during the reconstruction that creates a triangular mesh to be textured for final use or an editable model. The model can be produced as tiles or in one comprehensive block, there is also an option of creating the model based on different LODs. When reconstructing and texturing the 3D model the software always tries to use the image with the highest resolution. The software also has the ability to remove objects that are not visible in all images covering the same area. For example, if a moving cars is not visible in the same location in the majority of the images it will not be reconstructed in the 3D model (Acute3D, 2014b).

It is also possible to use polygon constraints when reconstructing 3D models in Smart3DCapture. It can e.g. be problematic to reconstruct water surfaces due to reflections and movements but when importing the water surface as a polygon the software reconstructs the surface with the shape of the polygon and only objects deviating from the surface will be reconstructed, e.g. boats (Acute3D, 2014b).

The models can be created in various 3D formats such as OBJ, OSGB and Collada but also in their own format S3C that can be used in the Smart3DCapture viewer, a software used to visualize and view the generated models. The models can be automatically exported into different LODs, a less accurate LOD is loaded when viewing a larger part of the model and a more detailed LOD when viewing the model in a more zoomed mode. The models can be exported and edited in external software and then again imported into the software where the edited 3D model will be textured (Acute3D, 2014b).

4 Results and Analysis

In this section visual results and analysis of the different created models will be displayed and commented. Thereafter the two different approaches of using street view images will be compared and finally some limitations and recommendations will be given.

4.1 Model Textured in TerraPhoto

Viewing the model textured with only aerial images gives a visually good impression when viewing it from a bird's perspective, see figures 7 and 8. The terrain model has been draped with an orthophoto. When viewing the model from a slightly tilted perspective it can be seen which buildings are created in 3D and which are only visible on the orthophoto that is projected on the surface model, see figure 8. Since the laser data and the two image sets were not acquired at the same time (laser data during 2007 and images during 2013) there exists several discrepancies between the datasets which led to that not all buildings were created and textured.

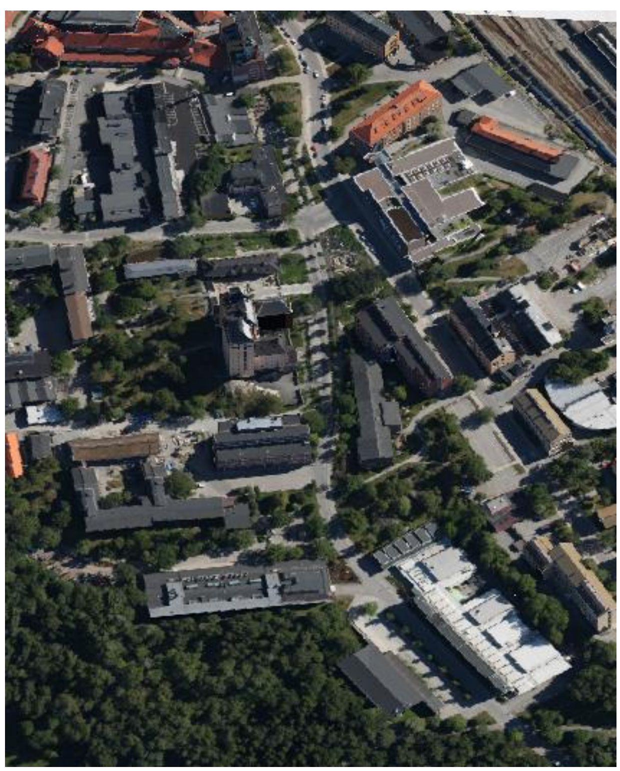

Figure 7 Screenshot in TerraPhoto, aerial textured model viewed from above

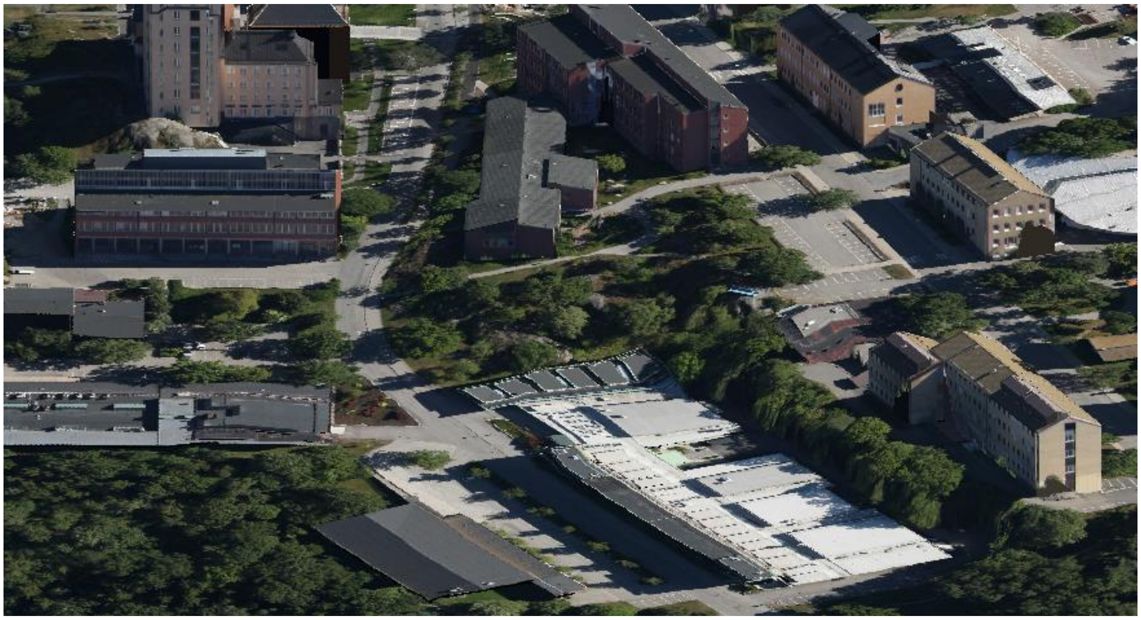

Figure 8 Screenshot in TerraPhoto, aerial textured model viewed from above with a tilted perspective

Figures 9 and 10 show two images of a building at KTH when zooming in on the model textured with aerial images. The textures are correctly placed on the building polygons and have not been adjusted or changed due to holes because of the use of depth maps or because other images are better matches.

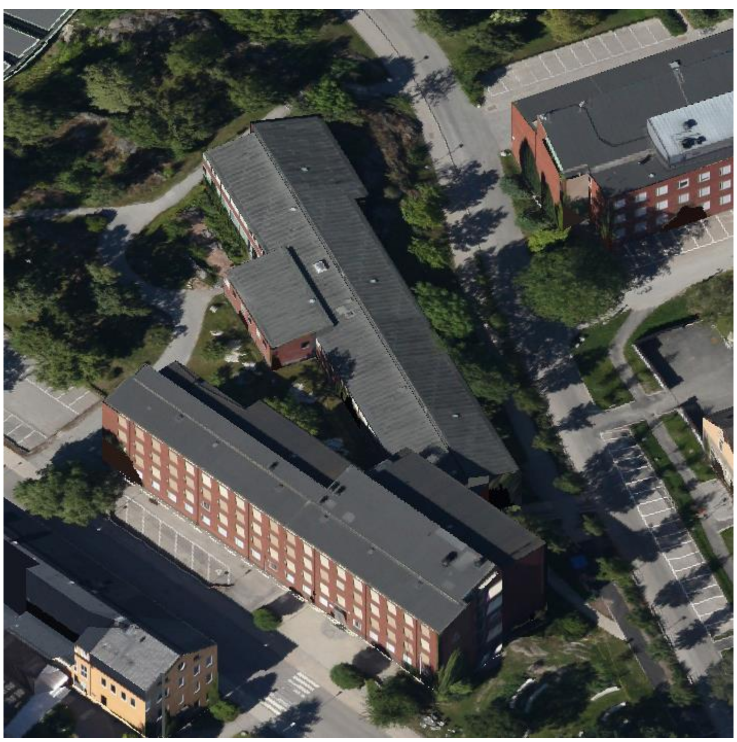

Figure 9 Screenshot in TerraPhoto, close-up on the aerial textured model

It can generally be said that the models textured with aerial images are photo-realistic models when viewing the models from a bird's perspective since that is where the images have been acquired. But as one views the model from a street view perspective, see figures 10 and 11, it is clear that the aerial images do not contain the high resolution needed to create a true photo-realistic model on street level. The buildings also tend to look very planar, which is an appearance that buildings do not have in reality. Architectural instalments such as doors and windows look flat even though they in reality might have some depth changes. Especially in figure 10 it can be seen how the window blinds have been projected onto the facades instead of standing out from the façade. The flat appearance is due to the fact that the buildings created are based on only aerial laser data which lack detail on building facades especially when approaching the street level.

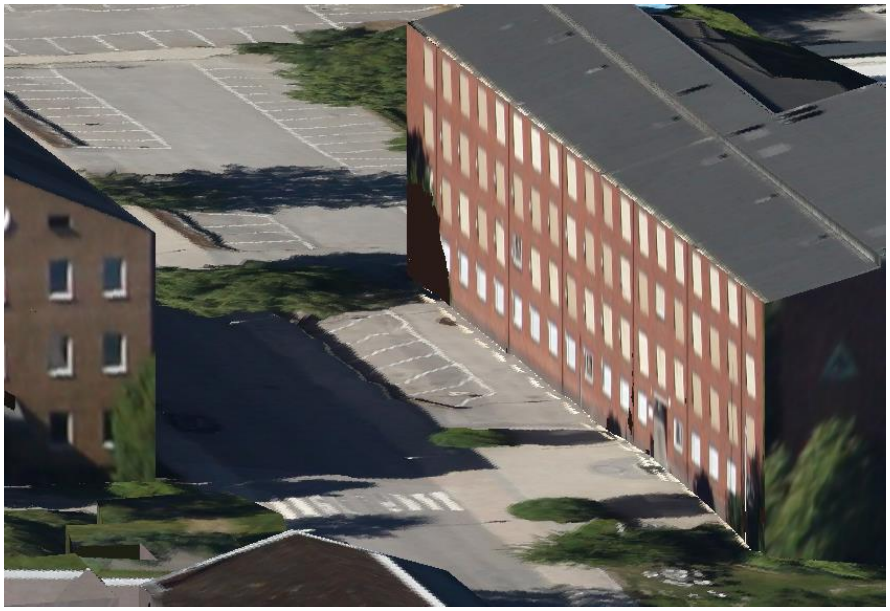

Figure 10 Screenshot in TerraPhoto, close-up on textured aerial model

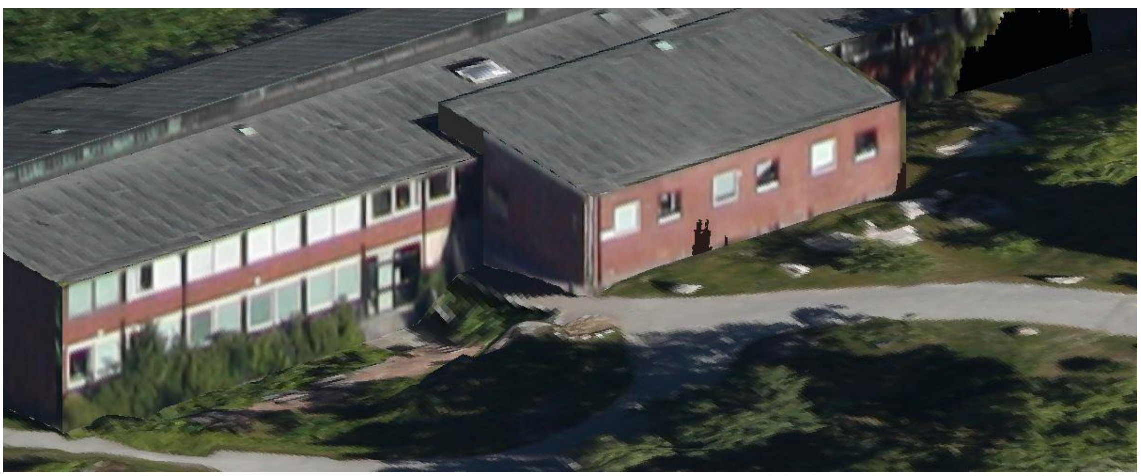

Figure 11 Screenshot in TerraPhoto, close-up on textured aerial model

In figure 12 the model textured with street view images can be seen from a perspective slightly seen from above. It can be seen how the street view images have added higher resolution to the model in the parts where street view images have been used, compared to the same building in figure 11. In the same figure it can be seen that the building façade in the right corner has not received a street view image as texture (see the red rectangle), this depends on that there was no street view image that suited the model better than the aerial image. That façade is lying in a corner of the building where the car was not able to drive to acquire any street view images.

Figure 12 Screenshot in TerraPhoto, model textured with street view images

Another issue with texturing with street view images that can be seen in figure 13 is that of texturing long building facades. The software tries to find one single image that can be used to texture the whole wall façade instead of stitching several images together. This can lead to that an image taken from a perspective is chosen in favor of several images acquired with a more perpendicular angle to the façade. In the right part of the façade in figure 13 the texture has a high resolution, but moving to the left along the building façade it can be seen how the resolution is decreasing due to the acquisition perspective. There exists an option in the software of saving all images that cover a certain façade and stitching them together in an external software and then choosing that stitched image for texturing. This can however be a tedious process if it needs to be performed for several building walls, which could be the case when texturing city models.

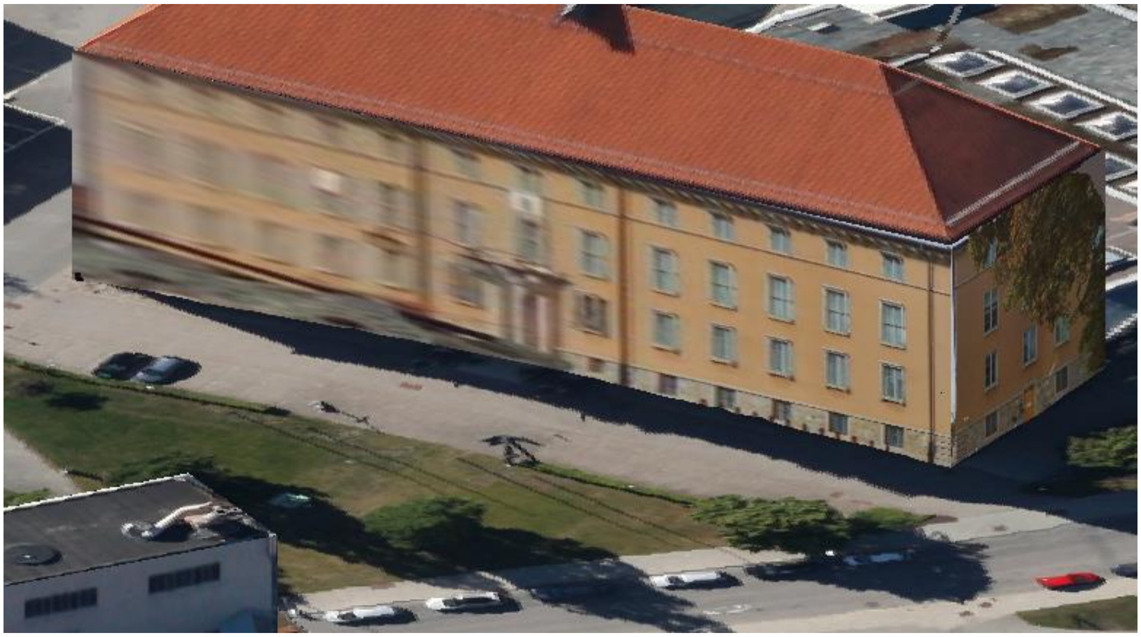

Figure 13 Screenshot in TerraPhoto, long façade textured with street view image

Using depth maps does not work as well for street view images as for aerial images since there are usually not as many laser points acquired on the building walls, where the images will be used as textures. This can for example be seen in figure 14 below. An image has been chosen for the left part of the building that also contains textures for the extension of the building, see the red rectangle. Optimally this would not happen since there is a building covering the façade from the angle where the image was acquired. A big tree in the left corner has also been projected onto the façade since there are no depth maps created that contain this tree, perhaps were no laser points acquired of this tree.

Figure 14 Screenshot in TerraPhoto, badly chosen texture due to perspective

Sometimes it can be difficult to choose which texture is best suited for a wall façade and it might be especially hard for the software to automatically choose a texture that looks visually pleasing. In figure 15 the automatically chosen texture is on the wall and in figure 16 the texture has been changed to an image that is taken from what looks like a more perpendicular perspective, it also helped removing the tree from the wall façade. When changing the texture of a façade the user gets options on different images that fit the wall and to what percentage it covers the building façade.

Figure 15 Screenshot in TerraPhoto, automatically chosen texture for the building façade

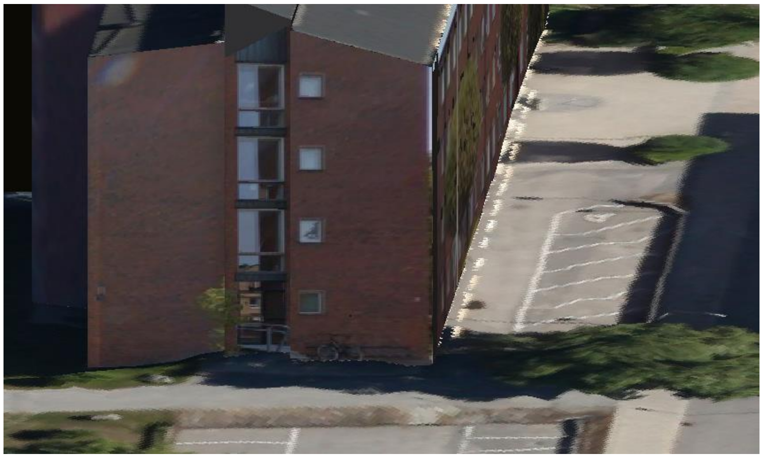

Figure 16 Screenshot in TerraPhoto, manually changed texture for the building façade

When adding street view images for texturing the model looks more photo-realistic comparing to the model textured with only aerial images, compare figures 11 and 12. However both the models tend to look very flat since also images acquired at a distance from the building walls are being projected onto the facades containing everything which is visible in the images. An example of this can be seen below in figure 17. The traffic sign has been projected straight onto the building facades even though the sign is actually standing on the pavement and not very close to the building façade. This is a problem with using images for texturing buildings since everything visible between the camera acquisition point and the building wall will be part of the texture and projected onto the building façade.

Figure 17 Screenshot in TerraPhoto, traffic sign projected onto the building façade

There are no other 3D objects than buildings visible in these models since only 3D building polygons are being textured and this gives the models a very clean and sparse appearance. To create more realistic city models with other types of objects they either need to be created from laser data in the process of creating the 3D model or they can be created separately and inserted into the model. Such objects can for example be trees, bridges or fences, but all these objects might not be suitable for being textured with images.

A limitation of texturing with TerraPhoto is that the default is that only one image is used to texturing each wall. The image is chosen based on percentage of coverage and the distance from acquisition point to wall. Using the central perspective images taken perpendicular to a wall in most cases means using images that do not cover the whole facades due to the narrow opening angle chosen to avoid distortions in the images. It is neither possible to directly use the aerial and street view images in combination when texturing the model. For tall building facades it would be desirable to have the top part textured with aerial images and the lower with street view images and then create a smooth transition between the images texturing the building façade.

The street view images come with the limitation that images are not acquired of all building facades due to limitations in the acquisition process. This makes it a more suitable approach to create a model based on aerial images and then switching the textures of the facades that do not fit well or look visually bad. It is a manual and tedious process, but it also creates a fully textured model based on both aerial and street view images. When using the street view images they do not cover building roofs and ground and therefore an orthophoto based on aerial images is draped on the roofs and ground to get those textures. The extraction of the street view images also needs to be performed in such a way that the vertical opening angle is wide enough to cover the whole building façade. In this case images with a 90 degree vertical opening angle were chosen with the angle being larger than needed, but it depends on the city environment that is being modeled. With smaller houses a smaller vertical opening angle would be enough but if texturing models of skyscrapers a wider opening angle would be needed. In such cases more than one image placed vertically would be desirable in order to avoid distortions due to a too wide vertical opening angle. When extracting the images from the panoramas not all extracted images are needed for texturing the buildings model, only the images that cover some part of building facades would need to be used. For this thesis this selection of images was performed manually but in cases of texturing larger cities this would mean much manual work.

When exporting the models from TerraPhoto it is only possible to export the building models with adherent wall- and roof- textures and not the terrain model that can be visualized in the software. This can be a disadvantage since the models need to be visualized in another software with complementary information of the terrain, one solution is to visualize the 3D building models in Google Earth.

4.2 Model Created in Smart3DCapture

The city models reconstructed from aerial images in Smart3DCapture create a photo-realistic 3D model that looks much like the world it's representing. What is being reconstructed in this model depends on what is being visible in the images, the resolution of the images and what pixels are being matched as tie points during the aero triangulation. Figures 18 and 19 show the reconstructed model based on only aerial images in Smart3DCapture as seen from a bird's perspective. When viewing the model from above it looks good and the viewer of the model perceives a realistic model and can get a good overview of the area which the model is representing. The geometries appear to be planar and the textures of high enough resolution to be viewed at a distance.


Figure 18 Screenshot in Smart3DCapture, overview of model reconstructed from aerial images


Figure 19 Screenshot in Smart3DCapture, overview of model reconstructed from aerial images

When viewing the model more closely and approaching the street level the viewer can experience certain limitations of the model. An example is the low resolution of the images used that sets limits to how photo-realistic the model is perceived at the street level, see figures 20 and 21.

Figure 20 Screenshot in Smart3DCapture, street view of the model reconstructed from aerial images

Figure 21 Screenshot in Smart3DCapture, street view of the model reconstructed from aerial images

Another example of the limiting low resolution can be seen in figure 22. The building has been reconstructed with the right geometries but the texture lacks resolution in comparison to the reality when viewing the model from a close perspective. The shape of the windows do not appear as straight and the facades have received a blurry façade color.

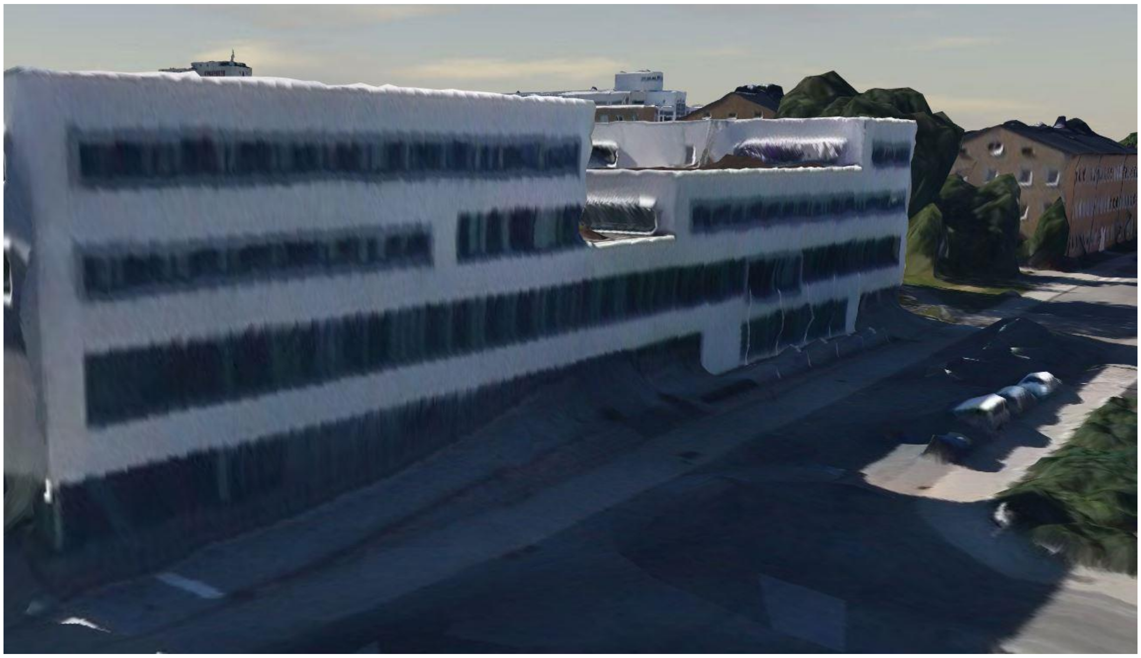

Figure 22 Screenshot in Smart3DCapture, blurry façade in model reconstructed from aerial images

Reconstructing the 3D city model from images of course has the advantage that everything visible in the images can be used for reconstruction. It does not only involve buildings, which is the case when texturing building facades, but also objects like trees and flag poles. An example of a tree reconstructed from aerial images can be seen in figure 23. The tree does not have the typical shape of a tree and no tree trunk is visible. This depends on that there are no images acquired that cover the sides of the trees from the ground since the images are acquired from an airplane.

Figure 23 Screenshot in Smart3DCapture, tree from model reconstructed from aerial images

The model reconstructed in Smart3DCapture based on only street view images was reconstructed with images of horizontal- and vertical opening angles of respective 90 degrees. This square format includes all parts of the buildings and at the same time a lot of information of the surrounding environment. Having a too large vertical opening angle though showed to create some complications as can be seen in figure 24 below.

Figure 24 Screenshot in Smart3DCapture, model reconstructed from street view images

Having a too large vertical opening angle of the central perspective images led to much sky and clouds being part of the images which the software also tries to use in the reconstruction process by finding matching points also in the sky. As seen in figure 24 above this leads to the model being shaped as a tunnel where the sky is a part of this tunnel. This is not desired since it is not a true representation of the world and it is not suitable when viewing the city model from above, which is an important application.

Despite the tunnel perspective in the model it can be seen that using the high resolution street view images improves the photo-realistic view of the model prominently, this can be seen in figure 25 below. The sky is still a part of the model but one can see how the street view images aid to create a more photo-realistic model when looking at the building façade.

Figure 25 Screenshot in Smart3DCapture, building reconstructed from only street view images

Using only street view images means that the whole environment is being reconstructed based on only these images. It can be problematic since this surrounding environment also includes the roads which the car acquiring the images is driving on. In figure 24 as well as figure 26 a consequence of this can be seen, namely the holes in the roads and the bumpy roads. Figure 26 displays a part of the road Sibyllegatan, located in the area of Östermalm in Stockholm. The roads has been reconstructed based on only street view images that have been acquired by a car that does not acquire images covering the entire roads and it is also difficult for the software to find matching points on roads that do not have features standing out such as road lines or pedestrian crossings. Also images of the car acquiring the images can be seen on the road in figure 26, red-white striped parts which is part of the car acquiring the cycloramas for Cyclomedia. This can be compared to figures 20 and 21 that are screenshots of the model reconstructed from aerial images where no holes and a flat street appears.

Figure 26 Screenshot in Smart3DCapture, reconstruction of Sibyllegatan, street view images

To deal with the problem of the holes the possibility of using polygon constraints was tested in Smart3DCapture. The polygon constraint functionality has been implemented for improving the reconstruction of water surfaces. The software takes a three dimensional polygon of the surface that is to be kept according to the shape of the polygon during the reconstruction. For this street the result is shown in figure 27 below, the street has now been flattened during the reconstruction process. But even though it helps with reconstructing a flat street there are still some problems when it comes to the textures that are being projected onto the flat street, in figure 27 below it can e.g. be seen how the street line and pavements are not reconstructed as straight features.

Figure 27 Screenshot in Smart3DCapture, reconstructed road through using polygon constraint

Another way to get rid of the holes in the street and the tunnel appearance of the model is to clean the triangular mesh resulting from the reconstruction from all triangles incorrectly modeled or fill holes due to missing tie points. Such a cleaned model can be seen in figure 28, also displaying Sibyllegatan in Stockholm. The triangular mesh was exported from Smart3DCapture in a different format (.OBJ) and edited in an external editing software, MeshMixer, and then imported into Smart3DCapture and being textured with the images used for reconstructing the model. Triangles belonging to the sky as part of the tunnel appearance were removed, holes in the triangular mesh were filled and the street was flattened to a smoother surface.

Figure 28 Screenshot in Smart3DCapture, edited reconstructed model

The model now has a very realistic appearance with flat streets and no sky as a roof in the model. The street also looks more planar compared to when using the polygon restriction. This though comes with the disadvantage of manual work. The editing needs to be performed manually and with large cities comes larger amounts of work. A problem still existing in figure 28 is that of cars blocking the facades. It can be seen that between the cars and the building walls the pavements are not flat but a bit wavy, this is due to the fact that there are no images covering this area and the software then matches based on what is visible in the images.

When experimenting with the images it also became clear that overlap is needed between images acquired from the same viewing point. This even though it does not add any extra detail to the model since the same part is visible from a different perspective acquired from a different viewing position. The overlap is though important in the aero triangulation because it helps the aero triangulated block to become more consistent with an increasing number of links between the images (Acute3D, 2014c). But without the overlap between the images from the same positions the model suffers from something that looks like distortions and wavy facades, see figure 29 below. It is clearly seen how the building façade is stitched together with images that are taken from different perspectives due to its wavy appearance. This model is based on images having a horizontal opening angle of 45 degrees but with no overlap between the images. These effects in the images depend on that the images and points that have been used in the image matching process have been acquired from different perspectives.

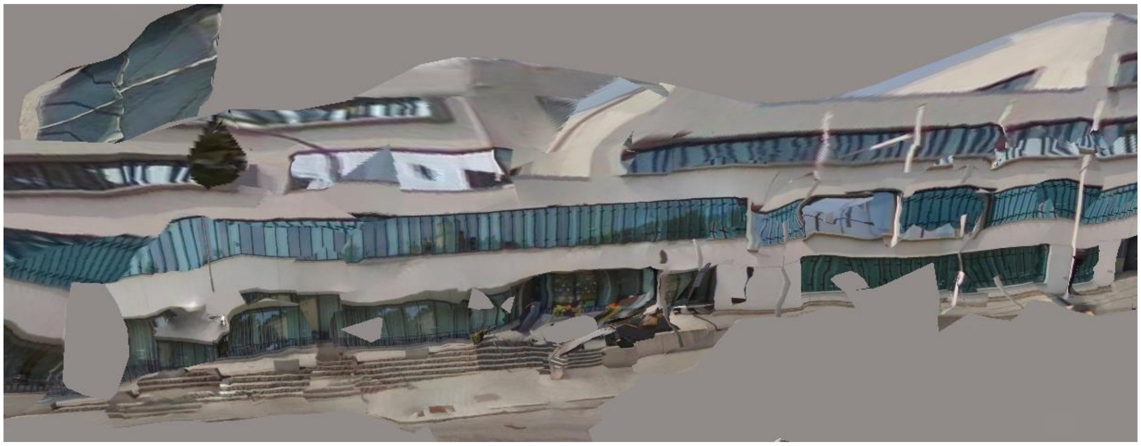

Figure 29 Screenshot in Smart3DCapture, façade reconstructed from images with no overlap from same camera location

Apart from the tunnel appearance seen in figure 24 above it is clear that the street view images contribute to create more photo-realistic models on street level and that it helps with reconstructing the building geometries. It should though be clarified that the street view images only help with reconstructing what is being visible in the images. Where there are no images nothing will be reconstructed and it is therefore desirable to use the street view images in combination with aerial images.

To overcome the limitations existing when only using the street view images a model reconstructed directly from both street view images and aerial images was created. As seen in figure 30 below the problem of having the sky as part of the model disappears when adding the aerial images in the reconstruction step. Using the aerial and street view images in combination is needed to create a model that contains detailed information of both the street level and from the sky. When looking at the model from this perspective it does not look much different than the model reconstructed from only aerial images, compare with figure 18 above.

Figure 30 Screenshot in Smart3DCapture, reconstructed model with both aerial and street view images seen from above

But when zooming in at the model it can be seen in figure 31 how the type of used image is changed in the border between roof and wall façade, since the street view images do not cover the roofs. The software always uses the image with the highest resolution when choosing what images to use as textures. It is also clear that the aerial images have been chosen for reconstructing the roofs since there are no street view images covering the roofs.

Figure 31 Screenshot in Smart3DCapture, model reconstructed from aerial and street view images

However, when looking at the model from a street view perspective, see figure 32, it is apparent that the use of street view images can aid in creating more photo-realistic 3D models on the street level. One of the biggest improvements are the wall façade textures, see figure 33. Using the high resolution street view images adds a photo-realistic touch to the model that cannot be created from only using low resolution aerial images and it looks more or less like an image acquired of the building wall.

Figure 32 Screenshot in Smart3DCapture, street level of the reconstructed model

Figure 33 Screenshot in Smart3DCapture, façade reconstructed from street view images

There are also other positive effects of using the street view images for reconstruction and one of them is the reconstruction of trees, see figure 34. With the help from images acquired also from the sides of the tree on the ground it is possible to get trees with tree trunks and a more realistic shape of the head of the tree. There is though a problem with the sky being a part of the head of the tree.

Figure 34 Screenshot in Smart3DCapture, tree reconstructed from street view and aerial images

All trees do however not turn out as good as the ones in figure 34 above. Another example of a tree is the one displayed in figure 35 below, left being the reconstructed tree and right being one of the input images of the same tree.

Figure 35 Faulty tree based on street view images (left), screenshot Smart3DCapture and original image of the tree (right), Cyclomedia

This tree has not been reconstructed with either the right shape of the tree trunk nor been textured with the correct texture, especially visible at the tree trunk. This can depend on several reasons, but the main one is that the tree is standing close to a building façade and that there are no existing street view images takes in between the tree and the building. That is also the main difference in comparison to the tree in figure 34 above, where images were existing in all directions around the tree since the tree is standing relatively free. Having street view images acquired in all directions is of course problematic since the images used here are acquired by a car and cars are not allowed or capable to drive everywhere in order to get the images needed for reconstruction.

The figures from the reconstructed model based on both street view and aerial images have so far been displayed from a model where the street view images were listed first in the XML-file that was used for import of the images into the software. It was decided to try and see if there is any difference in the resulting model when listing the aerial images in the XML-file first and then the street view images. Acute3D themselves say it should not matter because a global approach is being used for image matching in the aero triangulation and reconstruction meaning that all images are being compared to all other images and this should lead to that the input order should not matter. They do however add that some algorithms have a bit of a random selection which might lead to different reconstructions from one model to another and it turned out that in this case the input order of the images affected the result of the reconstructed model (Acute3D, 2014c).

It can be seen in figure 34 above that the tree contains some sky parts which was the case for most of the trees being reconstructed with the street view images first, see also figure 32. When instead importing the aerial images first the same tree got reconstructed as in figure 36 below. The head of the tree contains less sky parts but there is instead a large difference in how the tree trunk has been reconstructed. It is reconstructed wider and with a building façade as part of the texture.

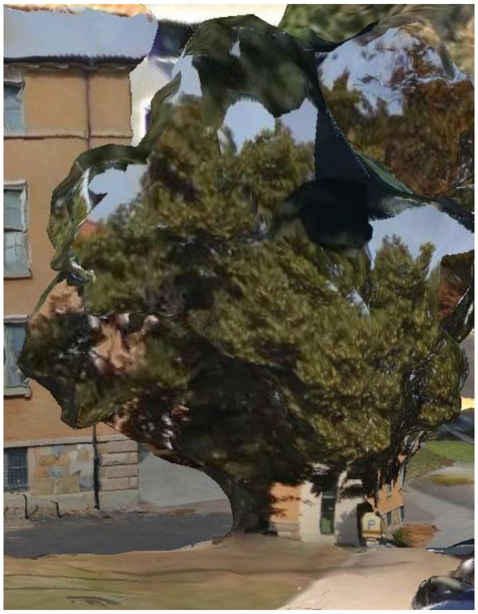

Figure 36 Screenshot in Smart3DCapture, tree reconstructed when importing aerial images first

Another difference in the model when importing the images in a different order in the software is the transition between aerial and street view images as can be seen in figure 37. The transition in this case is not as distinct as in figure 31 above and there are almost no sky parts in the border between wall and roof. It seems as though the first image set being imported into software through the XML-file, in this case the aerial images, works as basis for the geometry reconstruction and the rest of the street view images are being used as textures on these geometries, this since the street view images have a higher resolution compared to the aerial images. Care should therefore be taken when importing the images into the software since it does have an influence on the resulting model and the objects being reconstructed.

Figure 37 Screenshot in Smart3DCapture, model reconstructed with images imported in reversed order

Since Smart3DCapture could not use the panoramic images as input directly central perspective images were first extracted from the cycloramas. It is not enough to have images covering only the facades since the whole city environment needs to be reconstructed. It was though shown that having a too wide vertical opening angle created problems when creating models based on only street view images. This is a problem since the extraction of the street view images should preferably be done automatically with the same image sizes. That is however not applicable since the city environment is not uniform throughout a city and this leads to that images with e.g. different vertical opening angles are needed in different areas. This makes the extraction not as automatic and it needs to be considered when reconstructing 3D city models based also on street view images.

4.3 Comparison of the Two Approaches

There are many differences between the two tested approaches for using the street images to create more photo-realistic 3D city models on street level, both when it comes to the methodology of the approach used but also in the resulting models. In this section the main differences will be discussed.

One of the largest differences is the amount of images required for each approach. For reconstructing a whole city model of an area in Smart3DCapture many images are needed that cover the whole environment being reconstructed, it is not enough with only building facades since then only the buildings would be reconstructed. When reconstructing with Smart3DCapture only what is visible in the images will be reconstructed and it is therefore important that everything that needs to be modeled is visible in the image dataset. Using the street view images for reconstruction means dealing with a large number of images which generates a lot of processing. The models created in this thesis are all based on eight central perspective images from each acquisition location which are located approximately five meters apart along the streets. For the KTH area this resulted in around 1300 street view images used for reconstruction in comparison to around 800 street view images used for texturing in TerraPhoto. Using that many high resolution images for reconstruction means that a lot of tie points will be generated in the matching process of the images. Many tie points also produce many triangles being reconstructed in the final model. This is a question related to if the computer used is able to deal with rendering all of these triangles in an efficient manner. This is generally not a problem when using the TerraPhoto software for texturing, since each building wall is only built up of one or a few triangles that are being textured. It is a more efficient storing of the model but at the same time the model does not receive all characteristic architectural instalments such as windows, balconies and door openings modeled in a detailed way.

For texturing 3D models the way it was performed in TerraPhoto it is enough with images that cover only the building facades and images of the surrounding environment are not needed. For the study area of KTH approximately 800 of the central perspective images contained some part of a façade. It is still good to have overlap between the images in order to be able to stitch the images together for using them as façade textures. It is enough to have images covering only the building facades but selecting only images that contain some building façades is a tedious process if done manually after extraction from the panoramic images. When extracting the central perspective images from the cycloramas it could have been done through only choosing images oriented in a certain viewing angle in relation to the driving direction, such as 90 degrees left and right relative to forward driving direction. This is however only useful when driving on streets that are parallel to the building orientations and not when turning or when the buildings facades are not parallel to the streets. If extracted in such a way neither images of the sides of the buildings not facing the streets would be extracted. The extraction of images containing some part of building facades should therefore be considered further since it could lead to less processing. This is especially good when it comes to using the street view images in TerraPhoto where the relative orientation between all images need to be adjusted with tie points to receive a better texturing. With fewer images less manual work is needed for processing and this is also the case for other texturing software.

Another difference is the amount of manual work behind the two approaches. Texturing facades in different software usually involve much manual work, this is also the case for texturing with street view images in TerraPhoto. Perhaps not when texturing the facades automatically, but to visually check and adjust the results that have been automatically generated. Especially in the case with 3D city models with street view images where the number of images used is large and there are many possibilities of what texture to choose for a specific building façade. In this case the texturing process in TerraPhoto was done automatically after the tie point selection and adjusting the camera positions, but to refine the textures and switch to a different texture where the automatically generated textures did not match was done manually. Also the process of switching to a street view image texture was done automatically. There is a possibility of stitching images together in an external software to create panoramic images that can be used for texturing in TerraPhoto, but it results in more manual work when it comes to creating the panoramic textures to use. For this approach of texturing a 3D city model it is also required that there is an existing buildings model that can be textured with the street view images. A buildings model can for example be generated through the use of laser data, as done in this thesis, and the detail of this buildings model then determines the detail of the finished model in terms of geometries.

The technique used in the Smart3DCapture software is a more automatic approach once all the images have the required parameters. Both position and orientation data came from the extraction of the central perspective images from the cycloramas so what needs to be created is the XML-file to import these images with adherent information into Smart3DCapture. But in this specific case it might be argued that the approach is too automatic since the user has very little control of how the reconstruction is performed. There is no way to tell the software what type of images should be given priority in the reconstruction process. As shown above street view images yield better results when reconstructing tree trunks but aerial images are better for reconstructing the heads of the trees. As the software is built today there is no option how to deal with the problem of combining the two sets of images, they should be given the same priority due to the global approach but the listing order of the image sets seems to matter and the software always choses the image with the highest resolution for texturing. It might not always be the best to give priority to images with higher resolution, the lower resolution images could yield better results in some cases. For example would it be a good idea to always use aerial images when reconstructing and texturing objects from above, such as roofs and top of trees. This might also lead to no sky being part of heads of the trees since the aerial images would then be used as textures. Maybe could also the technique suggested by Früh & Zakhor of blending meshes bases on different sources be of use in this case (Früh & Zakhor, 2003).

Through simply viewing the models created with the different approaches and the different image sets large differences can be seen between the models, see figures 38-41. It can be seen that through the use of street view images the models receive a more photo-realistic texture that creates the feeling of being on the street. In the case of reconstructing a city model the whole environment including e.g. trees is being modeled which creates a full realistic model. The environment is only visible in the textured model due to a draped orthophoto on a terrain model. Viewing the model from a street view perspective the reconstructed models gives the impression of being in the city environment which the images are representing. In the case that a cleaner model is desired it is of course possible to edit and remove all objects that are not desired in the final reconstructed 3D city model, such objects can be trees or fences. But an easier way might be to use the textured model if such a clean model is desired.

The textured model creates a cleaner model which does not give the same realistic appearance due to the fact that no other objects than buildings are textured. The buildings model that is being textured in TerraPhoto has a more flat appearance and there are no other objects than buildings visible as 3D objects in the resulting model. To receive a more realistic city environment more processing and data is needed in order to model also the surrounding environment. But these clean models are also desired by visualization companies who wants to model the environment by adding trees etc. themselves. In the case cleaner models are desired the textured model might be useful directly.

Figure 38 Screenshot in TerraPhoto of model textured with aerial images

Figure 39 Screenshot in TerraPhoto of model textured with street view images

Figure 40 Screenshot in Smart3DCapture of model reconstructed from aerial images

Figure 41Screenshot in Smart3DCapture of model reconstructed from aerial and street view images

As both the visualization companies SightLine Vision AB and Agency9 mentioned, some clients do not want their models to contain trees or other objects, they instead want to be able to place these objects themselves in the final model (SightLine Vision AB, 2014; Agency9, 2014). In such cases a clean and sparse textured buildings model might be more suitable. Delivering such a plain model however becomes problematic when using a photogrammetric approach since objects that are visible in the images will be part of the reconstructed model and to remove all such objects much manual work could be needed. Care should therefore be taken when choosing to use also street view images for creating 3D models for a specific purpose, the high level of detail might not always be desired.

Which of the two approaches is the best to use when creating 3D city models much depends on what the desired result is and what the created city model will be used for. Using the models for visualizing how a new building would blend in with the surrounding environment a model like the one reconstructed in Smart3DCapture would be a better choice but if showing how a certain architectural building will look, with the focus on how the buildings looks, then the more cleaned textured model might be more suitable.

4.4 Limitations and Recommendations

Using the street view images as basis for reconstructing or texturing 3D city modeling apart from the high resolution come with some limitations that will be discussed in this section.

Neither of the two software could use the cycloramas directly and therefore central perspective images were extracted according to a specified size and vertical resp. horizontal opening angles. The creation of the central perspective images was done by Cyclomedia, the company whose cars are collecting the street view images. The creation of the street view images reduces the fisheye effect that can be seen in the panoramic images, also known as barrel distortion, that come from acquiring the images with a fisheye lens camera, see figure 2. For both approaches each cyclorama was divided into eight central perspective images with the same opening angles. For the texturing approach images were extracted with 90 degrees horizontal and vertical opening angles and with the pitch angle being 0 degrees, which means the camera appears to be looking horizontally relative to the driving plane. This is good for images used for texturing since the 0 degree pitch angle makes it look as if the images were acquired perpendicular to the façade. For reconstruction in Smart3DCapture the images were exported with the same opening angles, 90 degrees, but with a pitch of 15 degrees, the camera appearing to be tilting slightly upward. This angle was chosen so that less part of the streets would be covered by the images to be included in the reconstruction process so that this information would rather be taken from the aerial images. This helps in removing holes in the streets when reconstructing based on both aerial and street view images. It also helps reducing the risk of receiving parts of the acquisition car in the images, as was seen in figure 26 above.

One aim when creating 3D city models is that it should be done in an as automatic procedure as possible since it usually involves large areas and large amounts of data. For this thesis work the extraction of the central perspective images was performed automatically since all the images were extracted with the same size and opening angles within the whole area. The city environment does not look the same throughout a city which would require central perspective images with different sizes and opening angles. An issue is for example the vertical opening angle of the images that determine how much of the building façade will be covered in the image set which is not an easy task since all buildings do not have the same height. How much of the building façade that will be covered in the images is also related to width of the street and where on the street the camera is acquiring the images. It is not always possible for the car to choose where on the street it can drive, it has to drive on the assigned road lane and this creates longer distances to some building facades than to others. A suggestion of how to deal with this kind of problem is to split the city environment into different areas where each area would represent certain characteristics such as building types and building heights. For each area different image sizes could then be extracted that suits the particular environment contained within the area. An example could be that of using the building height in relation to the distance between building and car to calculate a suitable image size that can be extracted from the cycloramas. Having images with a larger vertical opening angle than what is needed to cover the building facades is not needed since it adds no extra information when it comes to building facades, the rest is taken care of by the aerial images.

Using all the street view images extracted is of course good since many images covering the same object from many directions are available. But the fact that there are images available in all directions also creates images having very long viewing distances. Those images might contain distortions and might not be suitable for reconstruction or texturing. For reconstruction it might be desirable to set a limit to how far a street view image can view to be able to be used for reconstruction and the same for texturing. Since there are images available in all directions from many acquisition locations this means that there are images acquired at closer distances covering all objects in the area being modeled. Using images acquired closer and perpendicular to a façade is also desirable when using the images for texturing. Instead of choosing one image that covers a whole façade it would be better to use several images viewing the façade at an angle close to perpendicular and then stitching them together. In the software TerraPhoto it is not possible to do this automatically, the stitching has to be done in an external software and then again imported into the texturing software.

It is not only a matter of extracting images with correct images sizes but also to extract the correct images, especially in order to reduce the number of images for texturing. It was mentioned before that it would be desirable for texturing purposes, as done in TerraPhoto, to only have street images that cover some part of a building façade. It could be easily done through an algorithm that looks at image orientations relative to driving direction but some building textures could then be lost if there are building facades not parallel to the driving direction. An approach could be to use an image processing algorithm to look for planar surfaces in the images and then only extract those images from the cycloramas making the assumption that those planar surfaces are buildings and that buildings only consist of planar surfaces. This is of course a big pre-processing step but it might help to reduce the actual texturing process even though all building walls are not planar and would then not be included in the extracted image set. In the case of using a texturing software like TerraPhoto it would though be useful since it would reduce the number of images that need to be matched through tie point selection.

Using street view images results in a large amount of images needed that makes the texturing and reconstruction process difficult and more complex. Using all street view images for texturing in TerraPhoto means defining tie points in all the images which can be a tedious process. The selection of the images from the large extracted dataset that contain some building facades can be done manually but it would be preferable to have an automatic selection of the images that do contain some parts of a building façade. Having many images for reconstruction makes the process more time-consuming since it tries to find corresponding points between all images, and with more images more time is needed.

It would have been desirable to use the panoramic images directly since it comes with many advantages, the most important one being that only one image is needed instead of eight smaller images covering the same area, and this reduces the number of scenes to be processed when texturing and reconstructing. This was also discussed in the study by Luhmann and Tecklenburg where they used panoramic images directly for reconstruction with the main advantage being that less images were needed compared to having central perspective images (Luhmann and Tecklenburg, 2004). The collection procedure of the panoramic street view images allows for all images to be acquired at the same time with the shading being the same across a larger area. Having several images of the same object acquired from different angles can also create the opportunity to get images of parts of facades that otherwise would be covered by e.g. trees when acquired from only a perpendicular perspective. But the biggest advantages if the cycloramas could be used directly is of course that those images already exist and in such a case no extractions from the cycloramas would need to be performed. But using these cycloramas for reconstruction and texturing would also involve more complex reconstruction and texturing since the images have been captured with fisheye cameras, the images being different than just regular panoramic images. Only using the cycloramas would neither provide the required overlap for reconstruction in Smart3DCapture.

The high-resolution of the street view panoramic images is of course an advantage when they are being used for texturing but it also comes with a few shortcomings. For example, the size of the images increases with higher resolution which is necessary to be able to observe features more clearly in the images. This might not always be a good thing since objects not necessarily interesting might appear in the images and in that case need to be filtered out or processed. It is also important to realize that everything visible in the scene when acquiring the images will be visible in the images. Especially for images with high resolution this means that many more features will be visible in the images that will be projected onto the facades or be reconstructed if visible in more than one image. An example of how it can be problematic when reconstructing all visible features can be seen in figure 42. In this scene four flagpoles are standing in front of the façade of this building and Smart3DCapture tries to use those pixels in the reconstruction process which ends up with parts of the façade being extruded from the actual geometry of the wall. In the case of using an image only for texturing this wall these flagpoles would instead be projected onto the façade, they would not mess with the geometry of the object but would be visible on the façade as they exist in the image. It is also possible to pre-process the images used for texturing so that the flag poles would not be as apparent when used for texturing. Apart from the high resolution creating issues with more features being visible in the images it is also problematic since high resolution images are usually large in file size which can generate a more time-consuming processing.

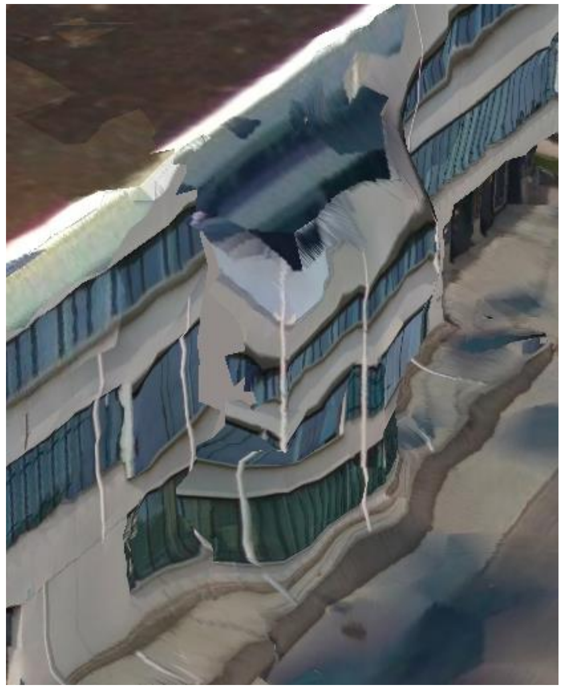

Figure 42 Screenshot in Smart3DCapture, flagpoles are destroying the building geometry

A problem existing in both the aerial and street view image datasets is shadows which can create visual problems when using them in city models. To avoid shadows in images they should all optimally be acquired when no shadows are present or be acquired in less sunny conditions so that buildings and other objects do not cast shadows in the city environment. An example of shadows in a model can be seen in figure 43 below. Avoiding shadows during the acquisition is of course not always possible due to different reasons. The aerial images are acquired when it was possible to get a flying permission and preferably with no or few clouds to get good visibility and quality of the images. The street view images used in this thesis were not acquired for the purpose of reconstructing 3D city models and during the acquisition process no care was taken in order to acquire the images under good conditions. It is difficult, if not even impossible, to acquire street view images covering an entire city without capturing any shadows. The shadows are not the only problem when it comes to the street images, also sun reflections occur in the images and can create sunspots in the images. To avoid these artefacts in the images one would need to plan the acquisition of the street images so that they are acquired during optimal conditions. Care should also be taken when combining the street view images with aerial images or combining street images acquired during different time epochs, if not acquired at the same time during the day shadows might fall from different directions causing conflictions in the city model. Shading is also problematic since some customers want to perform sunlight and shading analyses in the 3D city models and it is then not optimal to already have the effects of shadows and sunlight in the existing model (Agency9, 2014).

Figure 43 Screenshot in Smart3DCapture, tree casting shadows in the street images

Acquisition time is also important to consider when combining aerial and street view images. To combine them in an optimal manner the images should preferably be captured at the same instance so that non-permanent objects, e.g. construction sites, are existing at the same locations in both image sets. This may not be a problem when only using the images for texturing buildings since the objects will then only be projected onto the façade and there is a possibility to choose which image should be used, but when it comes to reconstructing surfaces based on images it is more important. If the object is only existing in some images but not in others it is difficult for both software and humans to know what is permanent and what is just a temporary change in one of the image sets. For the datasets used in this thesis the two image sets were not acquired at the same time, roughly a month apart, and no conflict between the image datasets were discovered during this thesis work.

Apart from shadows being present in the images they can also suffer from objects being displayed with different colors in different images due to acquisition from different perspectives. This can create problems when matching the images or stitching them together. It has not been discovered to be a distinguished problem in this thesis work, but it could be useful to have a color correction algorithm implemented in order to prevent this from creating issues in the matching process. This especially becomes a problem when viewing a model close, as seen in figure 44 below. This is a façade that has been textured in TerraPhoto through the use of two different images and it can clearly be seen how the images do not have the same color balance even though covering the same façade. It would be desirable to smoothen this change in color over the border so that the change is not as sharp. It might also be a good idea to adjust the colors in a pre-processing step before inputting them in a matching software like Smart3DCapture that uses all images for matching. If acquired from different perspectives it might be difficult for the software to find point correspondences because of different color balances.

Figure 44 Screenshot in TerraPhoto, color matching problem

The acquisition time during the year is also an important issue to consider when working with the images for visualization purposes. For viewing the images as well as using them for reconstruction it is of course more visually appealing if the trees are green since it makes the model look more alive and realistic. But having the trees in fully bloom means that the heads of the trees will be projected on the facades when texturing and that might not be desired. In some cases it might not be possible to find an image of the façade without vegetation and it would then maybe be better to have an image with only branches instead of trees in fully bloom.

The perhaps most important limitation of the use of the street view images is that they can never be used in solitude to create 3D city models. They are restricted to images acquired at street level and do not contain any information about building roofs and sometimes not of higher parts of building facades, depending on the vertical opening angle. The images used have been acquired by a camera mounted on a car and that restricts the coverage to where the car is able and allowed to access and drive. In other projects Cyclomedia has also acquired images by a camera mounted on a boat or on a moped. However, it is still limiting since there are no street images captured close by when it comes to areas only being accessible to pedestrians. There are also security and privacy issues associated with collecting this kind of data. Where is it ok to drive with the camera and how close is it ok to acquire images of someone's property and home? This is a data collection problem but it leads to problems when texturing and reconstructing 3D models. If there are no images of the buildings or the environment there is nothing to use for texturing and reconstruction. The aerial images here serve as a good complement to the areas where there are no street view images available, even though being acquired with a lower resolution.

5 Conclusions and Future Research

5.1 Conclusions

It has been shown that using street view images can help create more photo-realistic city models on street level, regardless if the street images are being used for texturing or reconstruction. Using the street view images as textures increases the level of detail to LOD 4. In this thesis work only the possibilities of these approaches have been evaluated through the use of smaller test areas and there is still more work to be done in investigating the techniques before they can be used to automatically create photo-realistic 3D city models in a larger scale. The two tested approaches still involve some manual work which would need to be reduced if they were to be used in large scale 3D city modelling in an automatic or semi-automatic approach. It has also been shown that the street view images can never be used in solitude but has to be combined with e.g. aerial images in order to create full scale 3D city models.

A problem when creating 3D city models is the surrounding environment, which has also been shown in this thesis work. Inaccurately visualized trees, bumpy roads etc. are some problems in the reconstructed models and projected objects onto facades are possible problems in the textured models. Some of the defects have been solved and corrected in this thesis work, bumpy roads through the use of road polygons or cleaning the triangular meshes and the trees have been shown to be reconstructed more accurately when acquiring street view images in all directions around the tree. It shows that there is great potential for creating photo-realistic city models through either of the two approaches tested.

The two suggested approaches deliver two very different 3D city models, the textured model being only the textured buildings model with a draped orthophoto on a surface model to represent the terrain and the reconstructed model being a full representation of the real world with all existing objects. Which approach to use depends on the purpose of the 3D city model and what it finally will be used for. If a very clean model is required for visualization then a textured buildings model might be better than a full representation of the environment as the case is with the reconstructed model.

There are some limitations with both approaches that have been evaluated in this thesis work and even though both approaches aid in creating more photo-realistic 3D city models on street level they do not appear as visually appealing as the models based on only aerial images. The question is if the street view images used in this thesis have too high resolution to be suitable for the task. There will always be improvements to make in this kind of modelling approaches since the demands are constantly increasing. It is though important to remember that 3D city models are just models representing the reality, they will never look exactly like the reality.

5.2 Suggested Improvements

In this section some general limitations and improvements of the methodology and the software used in this thesis will be discussed.

A recommendation when collecting the street view images and using them for texturing would be to collect them together with terrestrial laser scanning data. This would avoid discrepancies between the dataset used for creating the building models and texturing them and it would also lead to that there are laser points showing the details that are visible in the street view images. Having also terrestrial laser points would make it possible to use the depth map functionality in TerraPhoto which could help with removing projected objects on the building facades. As the images have already been collected without any terrestrial laser scanning data a possible solution now might be to perform some kind of image pre-segmentation to see what does not belong to the building facades and implement an algorithm that removes this part of the texture and chooses an image acquired from a different direction to fill that gap.

The texturing approach evaluated in this thesis with the software TerraPhoto also show that more functions need to be implemented in order to deal with the advantages that the many high-resolution street view images come with. One such function is that more than one image needs to be used for texturing each façade, all existing images need to be used to make full use of them.

The reconstruction approach in this thesis showed that different results are obtained based on if aerial or street view images are being listed first in the import list. It would be desirable as a user to get better control of the reconstruction process when using both aerial and street view images, in which urban environments should the aerial images be given priority and vice versa. Giving the aerial images priority when viewing objects from the top would be favorable when reconstructing roofs, streets and it might also help with the sky being present in the tree-tops, if aerial images are chosen no sky would be included since the sky is not present in the aerial images. The way the algorithm is implemented in Smart3DCapture now the street view images are always given priority as long as there is a street view image, which might not always be the best option. An idea could also be to have an existing buildings model based on e.g. laser data to serve as a rough basis when reconstructing the city model based on the images. The reconstructed model in Smart3DCapture could then be compared with the buildings model in order to obtain the vertical planar surface characteristic for a building but at the same time get windows and balconies characteristics as they differ from completely planar building walls. How this should be performed is a more difficult task but it could increase the accuracy of the building geometries where the images are not covering the areas well.

For reconstruction purposes it would also be desirable to have a ground model imported into the software so that the ground level is known when reconstructing e.g. streets. In this thesis a polygon constraint was used to create flat streets as well as editing the triangular mesh but it could require much manual work to create all these road polygons if they are not already existing before starting to create the 3D city model. The editing of a whole city model also includes much manual work. The use of a ground model based on e.g. laser data would therefore be preferable since no extra pre-processing would need to be performed.

A limitation of both the tested approaches within this thesis work is that it has only been tested on smaller study areas. In order to see if the approaches are suitable for large scale 3D city modeling both of them need to be further evaluated.

5.3 Future Applications and Future Work

A photo-realistic 3D city model based on both aerial and street view images has great potential in many fields. Having a detailed photo-realistic 3D city model on all levels is important and can be used for several different purposes. Using street view images on the ground level makes it easier for city planners and municipalities to visualize and communicate with its citizens. Using a model that does not represent the reality in a realistic way can make it difficult for municipalities and citizens to truly understand what is happening and how it would look in the real world. It is also important that focus lies on what is being visualized and not on the model itself when using digital 3D city models for visualization.

An advantage of using Smart3DCapture for reconstructing is that it can be an almost completely automatic approach to creating digital 3D city models. The only manual process is the input of the images and visually inspecting and editing the results. This can involve much manual work, all depending on the requirements of the final 3D model. Even though the reconstruction approach does not lead to perfect visual results, it can still be very useful considering the amount of work being involved in the generation process.

The city models created with street view images in this thesis work have some limitations due to the environment which they are representing. The models have been created through semi-automatic approaches and even though some more work is needed in order to create better models the results show more detailed and photo-realistic models in comparison to city models based on only aerial images. The visualization companies SightLine and Agency9 both see a use for this type of model but they are worried about the detail of the created models (SightLine Vision AB, 2014; Agency9, 2014). SightLine sees this kind of a model working somewhere in between their very detailed manually created and textured models and the model only based on aerial images. Agency9 still thinks that a model lacking in accuracy can be used internally for visualization, but to use it externally some more improvements would need to be made. Before starting to work with the street view images to create large scale 3D city models it would be useful to perform a more detailed survey to see what the exact requirements of the visualization companies and municipalities are.

For rendering and viewing the 3D city models it is also important to think about when the different views should be used. A model with street view images usually requires more storage space since it includes more high resolution images and it might therefore not be necessary to render these images when viewing the model from a distance. When viewing the model from a bird's perspective maybe the aerial images should be rendered and then when approaching street level the model should automatically be switched to the street level model. Another option would be for the user to have the option of choosing which model to be displayed when viewing it, based on the application. For example, when viewing the whole city a model based on aerial images might be enough but when looking at a single street it would be better to view the model based on street view images. This has not been investigated within this thesis work but it is an important matter to deal with in order to be able to visualize 3D city models in an efficient manner. For reconstructing models with the software Smart3DCapture this has not been an issue since the model is reconstructed into different LODs automatically when it is being created. This means that the model is generalized when viewing it from a distance but as the viewer zooms in on the model the viewed area becomes more detailed.

This thesis has not been complete in terms of investigating all existing software that can be used to create 3D city models with the use of street view images and this has neither been the purpose of the thesis work. There are not many software available that are suitable for reconstruction only based on images, most of them require also laser data for the reconstruction of geometries. And maybe the way to proceed with the street view images is to capture them together with ground based laser data so that the geometries are being captured during the acquisition and can just be textured with the acquired street view images. There also exists other software available for texturing already existing building models and some of those software might be better suitable for texturing with street view images than what TerraPhoto is. The most useable feature would be the possibility of being able to use more than one texture on each façade through automatically stitching images together.

References

123D Catch, 2014 – Turn Ordinary Photos into Extraordinary 3D Models with 123D Catch http://www.123dapp.com/catch (Accessed 2014-06-01)

Acute3D, 2014a – Acute3D Capturing Reality http://www.acute3d.com/ (Accessed 2014-02-10)

Acute3D, 2014b – *Smart3DCapture Product Sheet*, http://www.linkfast.com.tw/file/Acute3D/S3CProductDatasheet.pdf (Accessed 2014-03-03)

Acute3D, 2014c – Sylvain Lotteau, Application Engineer. Email 2014-05-22

Agency9, 2014 – Johan Göransson, Product Manager. Interview 2014-03-06

Amat N. et.al, 2010 - *Integration of aerial and close range photogrammetric methods for 3D city modeling generation*. Geoinformation Science Journal, Vol. 10, No. 1, 2010, pp: 49-60. Online at: http://eprints.utm.my/27785/1/HalimSetan2010_IntegrationofAerialandCloseRangePhotogrammetric.pdf (Accessed 2014-04-03)

Blom, 2014a – *BLOMSTREET*. http://www.blomasa.com/main-menu/products-services/data-models-blomurbex/blomstreettrade.html (Accessed 2014-02-12)

Blom, 2014b – *BLOM3D*. http://www.blomasa.com/main-menu/products-services/data-models-blomurbex/blom3dtrade.html (Accessed 2014-01-20)

Blom, 2014c – *BLOMOBLIQUE.* http://www.blomasa.com/main-menu/products-services/data-models-blomurbex/blomoblique.html, (Accessed 2014-02-12)

Elberink A.O. & Vosselman G., 2006 – *3D Modelling of Topographic Objects by Fusing 2D Maps and Lidar Data.* IEEE Conference on Computer Vision and Pattern Recognition, Kauai, USA, 2001, pp. II-31-38, vol. 2.2. Online at: ftp://jetty.ecn.purdue.edu/jshan/proceedings/topology_ucl2005/3Dmodellingoftopographicobjects_Goa_final.pdf (Accessed 2014-04-03)

Eniro, 2014 – Kartsök. http://www.eniro.se/hjalp/kartor/ (Accessed 2014-06-01)

Eskilstuna Kommun, 2014 – 3D model engineer Patrik Johansson. Interview 2014-03-18

ESRI, 2014 – *Esri CityEngine*, http://www.esri.se/Produkter/ArcGIS/3D-GIS/Esri-CityEngine/ (Accessed 2014-04-03)

Frueh C. & Zakhor A., 2001 – *3D Model Generation for Cities Using Aerial Photographs and Ground Level Laser Scans*. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, USA, 2001, pp. II-31-38, vol. 2.2.

Früh C. & Zakhor A., 2003 – *Constructing 3D City Models by Merging Aerial and Ground Views*. Univeristy of California, Berkley. Online at: http://128.59.11.212/~allen/PHOTOPAPERS/frueh.cga.pdf (Accessed 2013-12-06)

Google Street View, 2014 – Street View. http://www.google.se/help/maps/streetview (Accessed 2014-06-01)

Haala N. & Brenner C., 1997 - *Generation Of 3D City Models From Airborne Laser Scanning Data.* University Stuttgart. Online at: http://www.ifp.uni-stuttgart.de/publications/1997/tallin.pdf (Accessed 2014-04-04)

Kart & Bildteknik, 2011 - *Panoramabilder för fotogrammetrisk stereomätning och datainsamling*. Number 2011:4. Online at: http://www.kartografiska.se/images/stories/kart_o_bildteknik/KB_4_2011_web.pdf (Accessed 2014-02-08)

Kobayashi Y., 2006 - *Photogrammetry and 3D city modeling*. Arizona State University, USA. Online at: http://yoshikobayashi.com/?page_id=16 (Accessed 2013-12-06)

Luhmann T. & Tecklenburg W., 2004 – *3-D Object Reconstruction from Multiple-station Panorama Imagery*, University of Oldenburg. Online at: http://www.isprs.org/proceedings/xxxiv/5-w16/papers/panows_dresden2004_luhmann_b.pdf (Accessed 2014-05-08)

Mao B., 2011 – *Real Time Visualization of 3D City Models in Street Level View Based on Visual Salience*. International Journal of Geographical Informaton Science (Submitted 2011)

Micusik B. & Kosecka J., 2009 – *Piecewise Planar City 3D Modeling from Street View Panoramic Sequences*. Computer Vision and Pattern Recognition 2009. CVPR 2009 p. 2906-2912.

Mikhail E. et.al. 2001– *Introduction to Modern Photogrammetry*. First edition. John Wiley & Sons Inc.

nFrames, 2014 – nFrames, http://nframes.com/ (Accessed 2014-06-16)

Nverse, 2014 – Nverse Photo, http://precisionlightworks.com/nversePhoto.html (Accessed 2014-06-01)

PhotoShop, 2014 – PhotoShop, http://www.photoshop.com/ (Accessed 2014-06-01)

Saab, 2014 – *Vricon System*, http://www.saabgroup.com/en/Campaigns/Rapid-3D-Mapping/

(Accessed 2014-02-10)

Schulze-Horsel, 2012 – *3D Landmarks – Generation, characteristics and applications*. Online at: http://www.isprs.org/proceedings/XXXVI/5-W47/pdf/schulze-horsel.pdf (Accessed 2014-06-01)

ShapeQuest, 2014 – ShapeQuest, http://www.shapecapture.com/home.html (Accessed 2014-06-01)

Singh S.P. et.al, 2013 – *Virtual 3D City Modeling: Techniques and Applications*. Online at: http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-2-W2/73/2013/isprsarchives-XL-2-W2-73-2013.pdf (Accessed 2014-02-10)

SightLine Vision AB, 2014 –SightLine Vision AB. Interview 2014-03-04

SketchUp, 2014 – SketchUp, http://www.sketchup.com/ (Accessed 2014-06-01)

Tan Y.K.A. et.al, 2012 – *Large Scale Texture Mapping of Building Facades*. http://www.isprs.org/proceedings/XXXVII/congress/5_pdf/121.pdf (Accessed 2014-04-04)

TerraSolid, 2014 – *TerraPhoto– Ortho-rectification of Images and Data Visualization*. http://www.terrasolid.com/products/terraphotopage.html (Accessed 2014-02-12)

Tsai F. & Lin H.C., 2006 – *Polygon-based Texture Mapping for Cyber City 3D Building Models.* University of Singapore. Online at: http://gcl.csrsr.ncu.edu.tw/publication/bldgTexture.dist.pdf (Accessed 2014-04-04)

Xiao J. et.al, 2009 – *Image-based Street-side City Modeling*. Online at: http://web.mit.edu/jxiao/Public/publication/2009/TOG/paper_high-res.pdf (Accessed 2014-05-08)

Reports in Geodesy and Geographic Information Technology

The TRITA-GIT Series - ISSN 1653-5227

14-001 **Muhammad Zaheer.** Kinematic orbit determination of low Earth orbiting satellites, using satellite-to-satellite tracking data. Master of Science thesis in Geodesy No. 3130. Supervisor: Milan Horemuž. March 2014.

14-002 **Razmik Agampatian.** Using GIS to measure walkability: A Case study in New York City. Master of Science thesis in Geoinformatics. Supervisor: Hans Hauska. April 2014.

14-003 **Solomon Dargie Chekole.** Surveying with GPS, total station and terresterial laser scaner: a comparative study. Master of Science thesis in Geodesy No. 3131. Supervisor: Milan Horemuž and Erick Asenjo. May 2014.

14-004 **Linda Fröjse.** Unsupervised Change Detection Using Multi-Temporal SAR Data - A Case Study of Arctic Sea Ice. Master of Science thesis in Geoinformatics. Supervisors: Yifang Ban och Osama A. Yousif. June 2014.

14-005 **Can Yang.** Multi-Agent Simulation to Evaluate Sustainability of Cities. Master of Science thesis in Geoinformatics. Supervisors: Yifang Ban and Mats Dunkars. June 2014.

14-006 **Caroline Ivarsson.** Combining Street View and Aerial Images to Create Photo-Realistic 3D City Models. Master of Science thesis in Geoinformatics. Supervisors: Yifang Ban and Helén Rost. June 2014.

TRITA-GIT EX 14-006

ISSN 1653-5227

ISRN KTH/GIT/EX--14/006-SE