Large scale automatic extraction of 3D roof segments from aerial imagery

P2: graduation proposal

Martijn Vermeer

June 16, 2017

Student number:	4487362
Address:	-
Postal code:	-
Place of Residence:	-
Telephone:	-
Email:	-

Main mentor:	Hugo Ledoux
Second mentor:	Tom Commandeur

1 Introduction

Over the past two decades the 3d reconstruction of roofs has been a well studied topic within Geomatics Haala and Kada (2010). The main reason for this is that roofs form the basis for 3d city models which are used in a great variety of domains such as urban planning, emergency response, navigation systems, augmented reality and gaming to name just a few. Since 3d city models are often concern very large data sets that are optimally updated regularly methods for generating these models desirably have a high degree of automation (Demir and Baltsavias, 2012).

One of the most common ways to model roofs is as a polyhedron consisting of planar 3D roof segments (Rottensteiner, 2010). Interestingly these planar 3D roof segments are of direct interest as well, since they can be used for several cases such as asbestos detection and photo-voltaic potential calculations. Based on the photo-voltaic potential of roof segments making up the roof the optimal configuration and efficiency of solar panels can be determined. The applications for solar energy and asbestos detection make this a highly relevant topic. This is especially true in the Netherlands, since the Dutch government passed a law stating that asbestos roofs will be illegal from 2024. However, retrieving 3D roof segments the buildings blocks for roofs is far from a trivial task.

The usual data sources exploited for the extraction of roof segments are LIDAR and/or aerial imagery (Demir and Baltsavias, 2012). Depending on which sources are used three main approaches can be identified. The first and most straightforward one is to use LIDAR data only, the LIDAR pointcloud can be used directly to find planar segments. The second one uses aerial stereo images only, in this case the pointcloud needs to be created first by image matching. The third approach exploits both data sources, usually the LIDAR segmentation is enhanced with color information from aerial images.

LIDAR and aerial stereo imagery have their own strengths and shortcomings. Aerial imagery generally has a higher resolution compared to LIDAR and contains spectral information in addition to geometric information. On the other hand image matching is difficult in low texture areas especially when only two overlapping images are available, resulting in inaccurate and incomplete data (Hirschmüller, 2005). LIDAR is invariant to such lighting/spectral conditions.

Most studies present in literature make use of LIDAR data. In addition to the shortcoming of the data itself described above LIDAR is expensive to gather and therefore not widely available and frequently updated (Gehrke et al., 2010). Studies that do only use aerial stereo images are in practise often not applicable on a large scale since they make use of time consuming region growing algorithms (Rottensteiner, 2010)(Taillandier and Deriche, 2004).

The goal of this study is to efficiently extract 3D roof segments on a large scale using only Aerial Imagery. This will be done by determining the optimal image matching algorithm and exploring which clustering and segmentation steps are most efficient and necessary.

First an overview of the related work will be given in section 2. Second section 3 addresses the research questions. The methodology is described in detail in section 4. Subsequently the preliminary results are presented in section 5. Finally section 6 shows the time planning and the tools and datasets that were used are discussed in section 7.

2 Related work

2.1 roof reconstruction

During the mid 90s the first studies were conducted on the automatic reconstruction of buildings and in particular roofs. An important driver behind this research was the newly available

LIDAR technology that provided more accurate and increasingly dense height measurements. A detailed overview of the early attempts on building reconstruction is given by Brenner (2005) and a somewhat more recent update in Haala and Kada (2010). Since the resolution of height data was limited the earliest studies such as Haala (1994) often used a very model driven approach. In the model driven approach certain assumptions on the parametric shape (model) of the building are made beforehand, the parameters of this model are then found based on the data. In the data driven approach the parametric shape is directly determined from the data. Note that the distinction between the two approaches is not discrete but rather continuous. For this research the focus is on a more data driven approach. The reason for this is that the goal is to retrieve very detailed roof shapes which are hard to model. Furthermore there is no necessity for the segments to be perfectly aligned or watertight, which is a strength of a more model driven approach.

The method presented by Vosselman (1999) using LIDAR was one of the first more data driven approaches. Another study by Vosselman and Dijkman (2001) additionally uses ground plans to get more detailed results. Many more LIDAR based studies are present such as Huang et al. (2013) and Tarsha-Kurdi et al. (2007), which specifically focuses on the extraction of 3d roof planes using hough transform and extended RANSAC algorithms. Although these studies provide algorithms for the extraction of 3D roof segments from pointclouds, color information that is present in aerial imagery is not exploited.

Studies such as Rottensteiner and Briese (2003) and more recently Demir and Baltsavias (2012) combine LIDAR data with aerial imagery to achieve better reconstruction results. Furthermore Jarzabek-Rychard and Maas (2017) improves LIDAR resuls by edge refinement using aerial imagery. These hybrid approaches could be considered optimal since they take advantages of strengths of both data sources.

That LIDAR has been the dominant data source becomes clear from several overview papers Brenner (2005), Haala and Kada (2010) and Rottensteiner et al. (2014). The reason for this is likely the high quality of LIDAR in terms of accuracy and completeness as compared to image matching. The downside of LIDAR is that it often not available or up to date and on top of that expensive to gather (Gehrke et al., 2010). Therefore methods that only rely on aerial imagery as data source are of special interest for large scale applications. It could be argued that some of the methods presented above are still applicable by simply substituting the LIDAR pointcloud for a pointcloud retrieved from image matching. Caution should be taken when doing so, since the noise accuracy and completeness properties of these two pointcloud types can be very different.

Two early studies Suveg and Vosselman (2004) and Taillandier and Deriche (2004) that addressed the problem of reconstruction just based on aerial images show the difficulty of this task. Both methods are at least partly model driven, and are thus limited in processing any building type. In Rottensteiner (2010) pointcloud and aerial imagery information is combined. It is specifically mentioned that the pointcould can be retrieved by image matching. Initial segmentation is based solely on radiometric information, the pointcloud is subsequently used to reconstruct and retrieve roof planes. A disadvantage of the presented method is that it is likely to be not very efficient and since it makes use of a region growing algorithm like Nex and Remondino (2012) and previously mentioned studies in this section. A very recent study using only aerial imagery is Malihi et al. (2016), the methodology is model-driven and only a single building is analyzed. This study will thus focus on filling the research gap that remains for the large scale efficient extraction of 3D roof segments using only aerial stereo imagery.

2.2 image matching

As the image matching is such a crucial step in the process of retrieving 3D roof segments from aerial imagery, an overview will be given of the different matching algorithms that exist. A distinction can be made between stereo algorithms based on local, and global correspondences. An overview of local algorithms is presented by Scharstein and Szeliski (2002). In these algorithms local image patches are compared to find the optimal match. In the global approach additionally a global optimization function is defined Hirschmüller (2008). Local algorithms are generally fast, but perform poorly in areas with little texture or repeated patterns. As a result matching results are often incomplete and inaccurate. Global algorithms on the other hand can partly deal with these problems, however this comes at the cost of much longer computation times.

The semi-global-matching algorithm (SGM) proposed by Hirschmüller (2005) combines the advantages of both approaches. In SGM optimization is done along several 1D paths instead of the whole image. Therefore it is much faster than global approaches, whilst still achieving a similar accuracy. A more recent algorithm for efficient and accurate stereo matching is LIBELAS introduced by Geiger et al. (2010). The method makes use of the triangulation of support points that can be robustly matched. A similar study by Sinha et al. (2014) present an algorithm using local plane sweeps achieving accurate results in an efficient way as well.

Another branch of stereo matching algorithms is based on deep learning using convolutional neural networks. A recent study by Luo et al. (2016) showed that the results are very promising. However, training data for aerial images is currently nonexistent, since much of the research originates from the self-driving car industry. As a result these algorithms can not yet be exploited for aerial stereo matching without creating training data yourself.

3 Research questions

As argued in the introduction and the related work section the main goal of this research is to automatically extract 3D roof segments on a large scale using only aerial stereo imagery as data source. The corresponding main research question is therefore:

- *How to automatically extract 3D roof segments on a large scale using only aerial stereo imagery?*

subquestions related to this main question are:

1. *What is the optimal stereo matching method for this purpose?*
2. *How to efficiently retrieve roof segments from pixels with radiometric (RGB) and geometric (XYZ) information?*
3. *How to deal with partial information in the segmentation process?*
4. *What is the optimal clustering/segmentation method for this purpose?*

Regarding the scope of this research it is important to mention that the goal of this research is not to create new matching, clustering or segmentation algorithms, but to come up with an efficient workflow for the extraction of 3D roof segments mainly using existing algorithms. Which approaches and algorithms will used is decided based upon existing literature rather than via implementing and testing different algorithms and approaches. More explicitly this applies to determination of the optimal matching algorithm and segmentation/clustering approach and algorithms. Furthermore the research aims at developing an efficient method since it should be applicable on a large scale. Instead of trying to find the optimal implementation of algorithms and actually bench-marking them, the search for an efficient approach must be seen as more theoretical.

4 Methodology

The flowchart in figure 1 gives an overview of the entire workflow to retrieve 3D roof segments from a rectified pair of aerial stereo images. First the images are matched to create a disparity image, also known as depth map. A two step approach is used to extract the roof segments from the radiometric and geometric information in the aerial and disparity image respectively. In the first step Watershed segmentation is used to segment one of the images based only on its RGB values. Subsequently planes are fitted through the diparity values of the resulting color segments. Secondly, based on their orientation these segments can now be clustered and merged to form the final roof segments. The reason for this two step approach is that it is theoretically more efficient than a one step approach in which a orientation values for each point would need to be determined based on a neighborhood of points. The final roof segments can easily be reconstructed to world coordinates based on the plane that is fitted through the disparity values. After the roof segments are presented in world coordinates the BAG outline could be used to cut segments falling outside the building polygon. In the following sections all the individual steps are described in more detail.

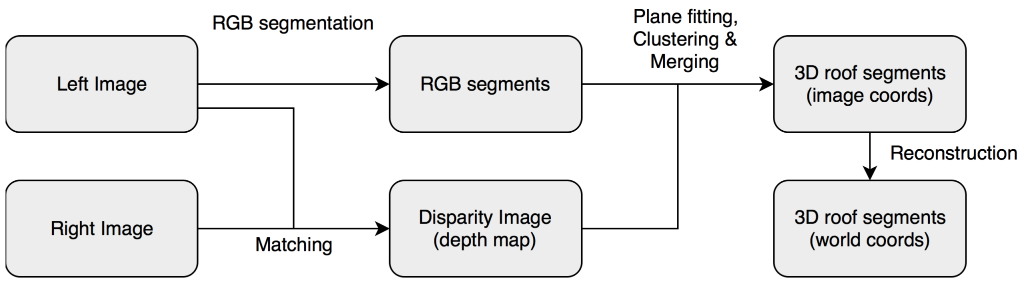

Figure 1: Workflow to extract 3D roof segments from aerial imagery

4.1 rectification

In order to be able to match two images they should be rectified first. A detailed description of the algorithm used for the rectification can be found in Fusiello et al. (2000). During the rectification process the two images are transformed to the same image plane. As a result of the rectification process, corresponding points in the two images will always have the same y coordinate. Finding matching pixels in the two images becomes therefore more trivial since only a single row need to be checked to find the corresponding pixel.

4.2 matching

All matching algorithms construct a disparity map from an rectified image pair. The disparity map contains the disparity (difference in x coordinate) between all matched pixels. Since the research goal is extraction of 3D roof segments on a large scale the matching algorithm should be efficient whilst still providing an accuracy as high as possible to retrieve satisfactory results. Implementations of two algorithms that are openly available in matlab satisfy these criteria. The semi-global matching algorithm presented by Hirschmüller (2005) and the LIBELAS algorithm introduced by Geiger et al. (2010). The two algorithms are compared to determine which one is optimal.

4.3 RGB watershed segmentation

Roughly speaking there are three approaches to segmenting an image into segments with similar colors. First, a region growing algorithm can be used to grow segments. Second, a clustering algorithm can be used cluster/group pixels with similar colors. Third, coloredges can be detected and subsequently watershed segmentation can be applied. Generally region growing algorithms can deal better with color gradients within segments compared with clustering algorithms. However region growing has as drawback that it is relatively slow since it is iterative. Furthermore both region growing and clustering form regions on color edges especially when these are wide and not distinct. Watershed segmentation overcomes most of these problems since it is efficient, can deal with color gradients within segments and forms boundaries at the center of the color edges. For these reasons it was decided to do the initial color segmentation with a watershed algorithm.

4.4 plane fitting

All disparity values within each color segment are used to fit a plane through the color segment. Plane fitting is done using the MSAC a variant of the RANSAC algorithm. This algorithm is iterative and therefore more robust against outliers than a single least squares approach. It should be noted that more robustness goes at the cost of efficiency.

4.5 clustering

The color segments are grouped together using the mean shift clustering alogrithm. This particular clustering algorithm was selected since it is both efficient and accurate. Furthermore the amount of clusters does not need to be specified advance which is the case for example with k-means clustering. Since the amount roof segments for a building is unknown in advance this is an import property. The clustering is done based on the color, orientation and location of the color segments.

4.6 merging

A final merging step is required since a clustering algorithm also takes into account color and location which could be varying especially in larger roof segments. This merging step merges neighboring regions only based on whether the plane parameters of the two segments are similar to a certain degree.

4.7 reconstruction

The reconstruction of roof segments from image coordinates to world coordinates is done in two steps. First a polygon is retrieved that forms the outline of the particular segment. The disparity values of the vertices of this polygon can be determined based on the planed that is fitted through the disparity values of the particular segment. Second the vertices in image coordinates (row,column,disparity) are transformed to world coordinates (X,Y,Z).

4.8 post-processing

Several final steps can be added to clean and finalize the roof segments. Most significant is the BAG that can be used to cut segments falling partly or entirely outside the building polygon. Furthermore useless segments can be filtered out. An example of such segments are those with a slope higher than a certain value, which generally represent walls instead of roofs.

5 Preliminary results

The preliminary results will be presented by showing the different processing steps for a single building. The selected building has a trivial shape and contains little roof objects such as chimneys, dormers and roof lights. Therefore this is a relatively easy building to extract the 3D roof segments from.

5.1 rectification

Figure 2 shows the rectified image pair for the example building. The actual images are much larger, for this selection the BAG outline was used to cut out the area of interest including a buffer. Note that this buffer is necessary to be sure to capture the entire building on the image due to the perspective differences.

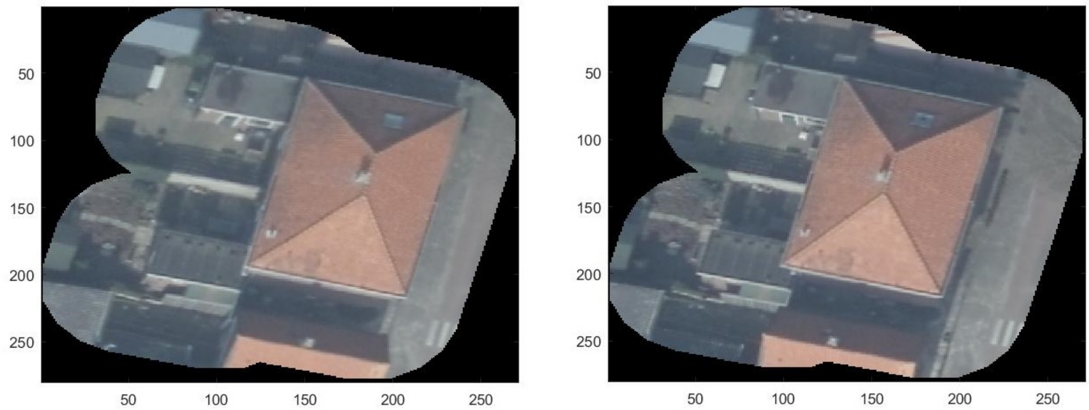

Figure 2: Left and right view of a single building cut out from a rectified pair of stereo images

5.2 matching

Figure 3 shows the disparity map of the left image of the selected building. White values indicate a high disparity and dark values a low disparity. Objects that are closer to the camera have a higher disparity, since the aerial images were taken from above this means that disparity gives an impression of the height. Higher disparity values correspond to higher elevation and lower disparity values to lower elevation. In the image the roof is therefore characterized by high disparity values and the ground by low disparity values. Furthermore for some pixels no match is found, examples of such pixels are the black spots on the right side of the roof. The reason that no match is found is these areas becomes clear when looking at the rectified images in figure 2. This Area in particular has very little texture and is therefore hard to match. Additionally occluded areas can't be matched, a clear example of this can be seen on the side of the left roof segment.

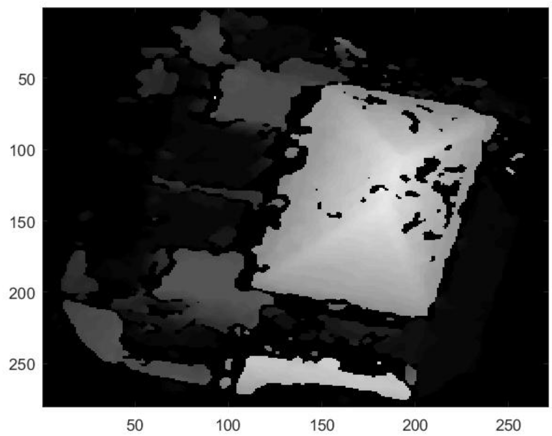

Figure 3: Disparity retrieved by semi-global-matching(SGM)

5.3 watershed segmentation

Figure 4 shows the two steps within the watershed segmentation. The left image shows the result of the color edge detection. The right image shows the actual watershed segmentation. The colors of the segments are completely random. Note that edges of the roof segments are already visible, however the image is still over-segmented.

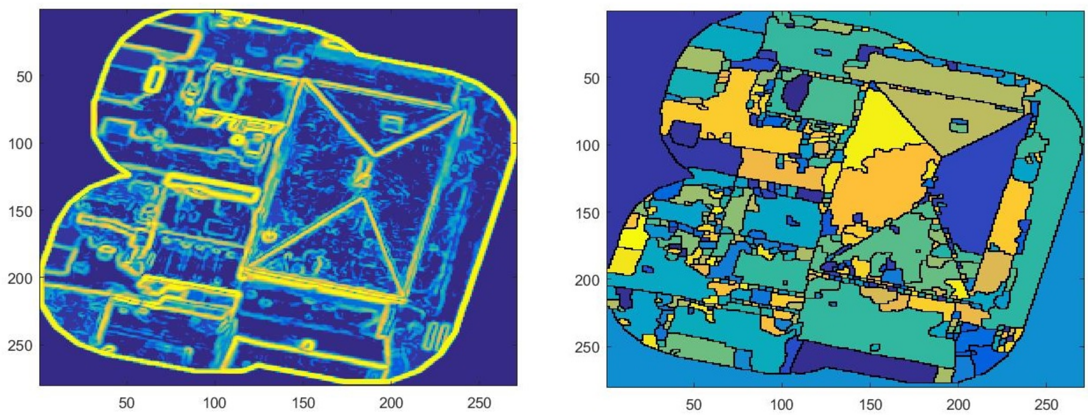

Figure 4: Watershed segmentation showing the coloredges (left) and resulting segmentation (right)

5.4 clustering & merging

The result of the clustering and merging steps is presented in Figure 5. As can be seen in the left image most of the over-segmentation is already dealt with during the clustering step. However some larger areas still need to be merged which is done in the second merging step (right image). After these steps larger areas that have the same orientation representing planar roof segments remain. At this point the four main planes of the roof can clearly be distinguished in the image.

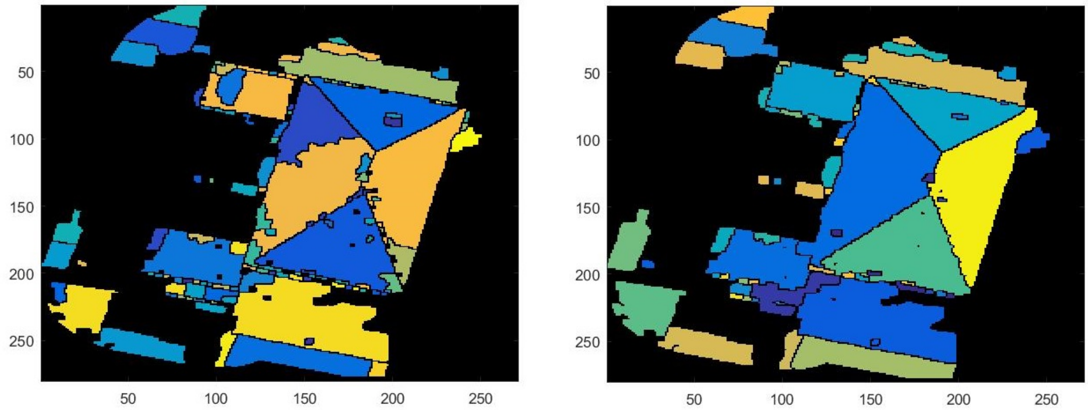

Figure 5: Clustering (left) and merging (right) of the color segments to retrieve roof segments

5.5 reconstruction

Figure 6 shows the reconstructed segments after filtering and cutting with the BAG outline. Note that holes in the segments are not yet considered and only the outer ring of the segment is reconstructed.

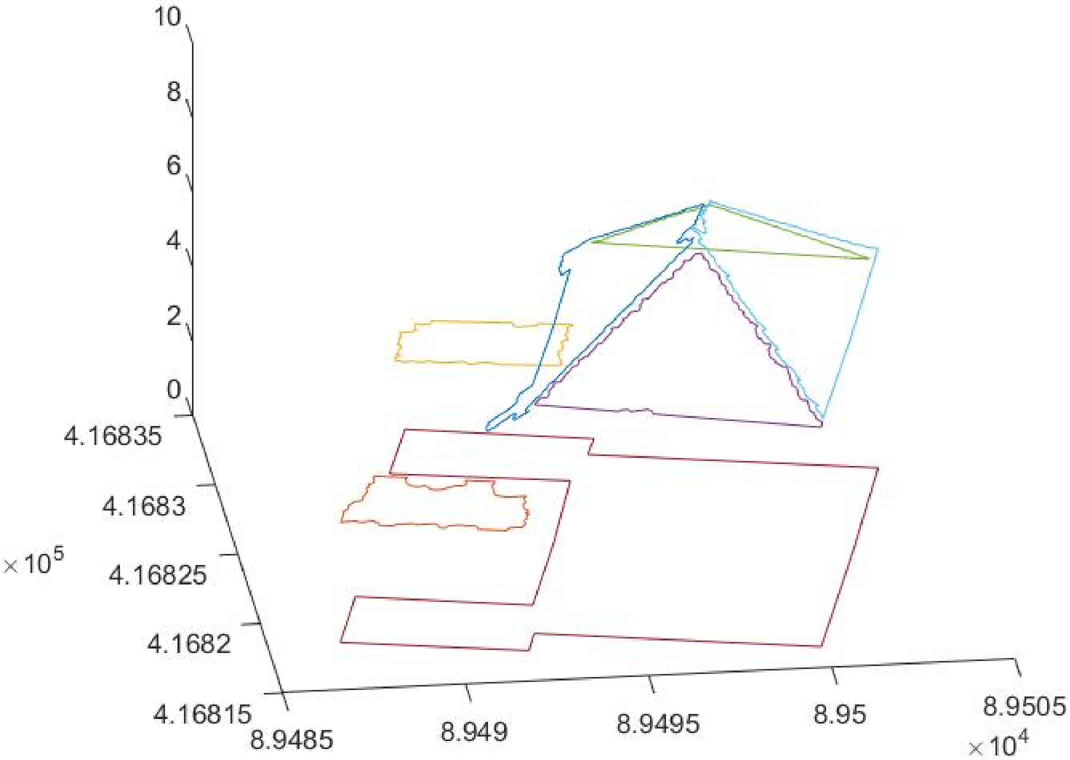

Figure 6: Reconstruction of the roof segments visualized over the BAG footprint

6 Time planning

Figure 7 shows the planning for the graduation project. Note that even though results are already present for the different steps in the workflow these steps are far from finished at this point. For the matching the planning is to implement some results with the SURE software package by nframes. This is a licensed state of the art software package giving a good indication of what would be the optimal result. For the clustering and merging testing is required to determine which parameters should be used and with which weights. Furthermore alternative algorithms (spectral clustering) could yield better results. Reconstruction is quite straight forward, however the program should be able to treat holes. Post-processing, based on the final workflow different filtering and cleaning steps are required in the end to finalize the results.

7 Tools and datasets used

7.1 Tools

MATLAB: The main tool used during this project is MATLAB. The company READAAR generally develops in MATLAB enabling this project to be integrated easily. Furthermore this project is in the field of photogrammetry in which is often dealt with large matrix calculations. Since MATLAB is particularly designed for this purpose it is a good choice as well. Examples of advantage are quick visualization, efficient matrix computation and wide variety of built in libraries/functions.

PostgreSQL & PostGIS: The database management systems used to store BAG data and resulting roof segments and any additional data is PostgreSQL with the PostGIS extension for spatial data.

QGIS: The main usage of QGIS is for visualization. Furthermore it is used for trivial tasks such as clipping when selecting areas etc. Possibly it will be used for some spatial analyses later on in the project, but that is not clear yet.

Python: Some scripts are in Python when more convenient libraries are available. In particular communication with the database occurs via Python.

7.2 Datasets

Stereo10: The main dataset used for this project is the nationwide aerial stereo imagery of the Netherlands, which has a resolution of 10 centimeter. This dataset is known as stereo10 and yearly updated. The dataset is not open source and currently managed and sold by the company Cyclomedia.

BAG: The second dataset used in this project is the BAG. This is the national register of adresses and buildings. The most important part of this dataset that is used is the footprint of the buildings.

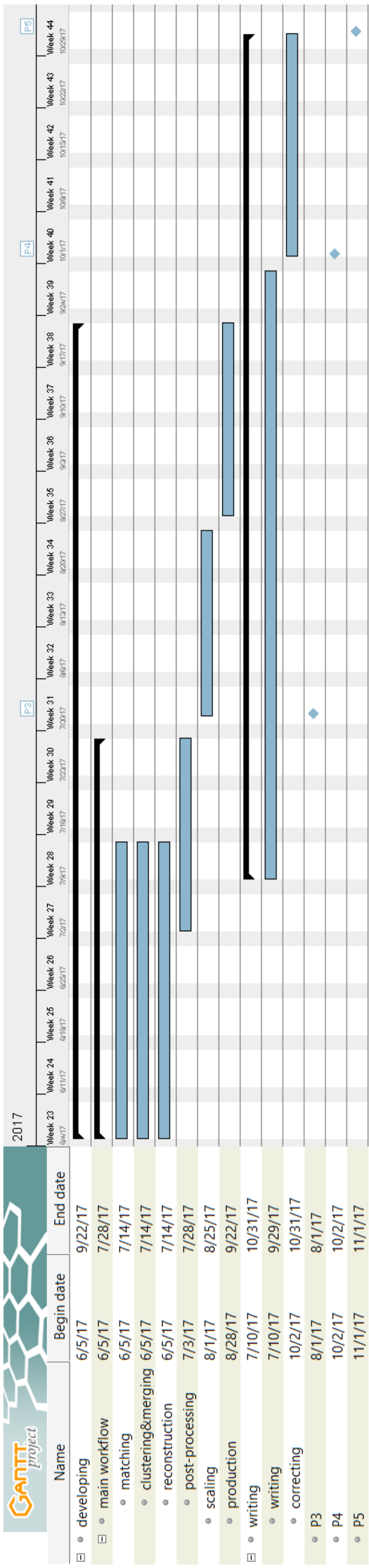

Figure 7: GANTT chart showing the time planning for the graduation project

References

Brenner, C. (2005). Building reconstruction from images and laser scanning. *International Journal of Applied Earth Observation and Geoinformation*, 6(3):187–198.

Demir, N. and Baltsavias, E. (2012). Automated modeling of 3d building roofs using image and lidar data. *Proceedings of the XXII Congress of the International Society for Photogrammetry, Remote Sensing*, 25.

Fusiello, A., Trucco, E., and Verri, A. (2000). A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22.

Gehrke, S., Morin, K., Downey, M., Boehrer, N., and Fuchs, T. (2010). Semi-global matching: An alternative to lidar for dsm generation. *Proceedings of the 2010 Canadian Geomatics Conference and Symposium of Commission I*, 2.

Geiger, A., Roser, M., and Urtasun, R. (2010). Efficient large-scale stereo matching. *Asian conference on computer vision*, pages 25–38.

Haala, N. (1994). Detection of buildings by fusion of range and image data. *Spatial Information from Digital Photogrammetry and Computer Vision: ISPRS Commission III Symposium*, pages 341–346.

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

Hirschmüller, H. (2005). Accurate and efficient stereo processing by semi-global matching and mutual information. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hirschmüller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*, 30(2):328–341.

Huang, H., Brenner, C., and Sester, M. (2013). A generative statistical approach to automatic 3d building roof reconstruction from laser scanning data. *ISPRS Journal of photogrammetry and remote sensing*, 79:29–43.

Jarzabek-Rychard, M. and Maas, H.-G. (2017). Geometric refinement of als-data derived building models using monoscopic aerial images. *Remote Sensing*, 9(3).

Luo, W., Schwing, A. G., and Urtasun, R. (2016). Efficient deep learning for stereo matching. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5695–5703.

Malihi, S., Zoej, M. J., Hahn, M., Mokhtarzade, M., and Arefi, H. (2016). 3d building reconstruction using dense photogrammetric point cloud. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

Nex, F. and Remondino, F. (2012). Automatic roof outlines reconstruction from photogrammetric dsm. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(3):257–262.

Rottensteiner, F. (2010). Roof plane segmentation by combining multiple images and point clouds. *Proceedings of the Photogrammetric Computer Vision and Image Analysis Conference*, 38:245–250.

Rottensteiner, F. and Briese, C. (2003). Automatic generation of building models from lidar data and the integration of aerial images. *ISPRS*, 19.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J. D., Breitkopf, U., and Jung, J. (2014). Results of the isprs benchmark on urban object detection and 3d building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93:256–271.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1/2/3):7–42.

Sinha, S. N., Scharstein, D., and Szeliski, R. (2014). Efficient high-resolution stereo matching using local plane sweeps. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1582–1589.

Suveg, I. and Vosselman, G. (2004). Reconstruction of 3d building models from aerial images and maps. *ISPRS Journal of Photogrammetry and remote sensing*, 58(3):202–224.

Taillandier, F. and Deriche, R. (2004). Automatic buildings reconstruction from aerial images: a generic bayesian framework. *Proceedings of the XXth ISPRS Congress*, pages 12–23.

Tarsha-Kurdi, F., Landes, T., and Grussenmeyer, P. (2007). Hough-transform and extended ransac algorithms for automatic detection of 3d building roof planes from lidar data. *Proceedings of the ISPRS Workshop on Laser Scanning*, 36:407–412.

Vosselman, G. (1999). Building reconstruction using planar faces in very high density height data. *International Archives of Photogrammetry and Remote Sensing*, 32(3):87–94.

Vosselman, G. and Dijkman, S. (2001). 3d building model reconstruction from point clouds and ground plans. *International archives of photogrammetry remote sensing and spatial information sciences*, 34:37–44.