AUTOMATIC BUILDING DETECTION USING OBLIQUE IMAGERY

Jing Xiao

Examining committee:

Prof.dr.ir. M. Molenaar	University of Twente
Dr.ir. R.N.J. Veldhuis	University of Twente
Prof.dr.-ing. Chr. Heipke	Leibniz Universität Hannover
Prof.dr.ir. P.J.M. van Oosterom	Delft University of Technology

ITC dissertation number 224
ITC, P.O. Box 6, 7500 AA Enschede, The Netherlands

ISBN 978-90-6164-353-1
Cover designed by Benno Masselink
Printed by ITC Printing Department

UNIVERSITY OF TWENTE.

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

AUTOMATIC BUILDING DETECTION USING OBLIQUE IMAGERY

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof.dr. H. Brinksma,
on account of the decision of the graduation committee,
to be publicly defended
on Wednesday 15th May 2013 at 16.45 hrs

by
Jing Xiao

born on 6th January 1984

in Wuhan, China

This thesis is approved by
Prof. dr. ir. M.G. Vosselman, promoter
Dr. M. Gerke, assistant promoter

To my family

Acknowledgements

The four years PhD life is a challenging, fruitful, enjoyable, painful and unforgettable journey. Without the support from many people, I could not finish this journey by my own. At the end of it, I would like to express my most sincere appreciation for all their help and collaboration.

My greatest gratitude and highest respect in the first are expressed to my promoter, Prof. George Vosselman. He provided me this PhD opportunity and opened my eyes to Photogrammetry. It has always been an honour to work with him.

I am indebted to Markus Gerke, my daily supervisor for his close guidance, valuable comments, moral support and his ever accommodating nature. Thankfulness is expressed also to my colleague and friend Adam Patrick Nyaruhuma. We started work on a similar topic around the same time, and we have had nice discussions and communications along the way.

I feel lucky to work with all the talented people in EOS. Thanks to Prof. Stein for his encouragement and advice. Teresa Brefeld makes the life at ITC nice and joyful. Also many thanks are given to all the colleagues for their constructive suggestions and assistance.

Of course friendship is another fruit I have gained through the study. I am immensely grateful to my close friends Pablo, Yaseen, Coco, Biao, Qiuju, Xiaogang and all the Chinese community. I would also like to thank Wang Tiejun and Michael Weir, who were always kind and helpful through my entire life in the Netherlands.

My PhD study has been funded by the Chinese Scholarship Council. Thanks for CSC to offer me this wonderful opportunity to do a PhD at ITC.

Most importantly, I can never express my deepest gratitude to my parents for their unconditional love, attention and support over the years. Thanks a lot to my husband who gives me the freedom and trust to explore my life. This book is dedicated to my family.

"There is no such thing as darkness;
only a failure to see."
- Malcolm Muggeridge

Table of Contents

List of figures

List of tables

Abbreviations

ALS	Airborne Laser Scanning
CIR	Colour Infrared
CityGML	City Geography Makeup Language
EFS	Pictometry® Electronic Field Study™
EuroSDR	European Spatial Data Research
GCP	Ground Control Point
GPS	Global Positioning System
GSD	Ground Sample Distance
IMU	Inertial Measurement Unit
Lidar	Light detection and ranging
LoD	Level of Detail
MDL	Minimum Description Length
MRF	Markov Random Field
PMVS	Patch-based Multi-view Stereo Matching
RJMCMC	Reversible Jump Monte Carlo Markov Chain
RMS	Root Mean Square
RMSE	Root Mean Square Error
SIFT	Scale-invariant feature transform
SenP	Slagboom en Peeters

Chapter 1 Introduction

1.1 Problem domain

With the rapid development of electronics, computer technology in recent years, many of the 3D geo-information applications envisaged in the past has gradually become a reality, such as urban emergency response, virtual tours, intelligent transportation, urban planning and design. Therefore, various efforts have long been made for the 3D information generation, especially on 3D city modelling.

OGC CityGML develops an international standard on 3D geo-information, issuing data acquisition, data maintenance and the specific applications. Stoter et al. (2011) also published a pilot research on establishing national 3D standard in the Netherlands (3D standard NL), which can also be valuable for the other countries to building their own standard.

The input data for building 3D geo-information includes the traditional cadastral maps, documentations from field work, as well as the objects extracted from remotely sensed data. Obtaining the latter has been a challenging task in the fields of photogrammetry and computer vision for almost the last two decades (Grün et al., 1997; Grün et al., 1995).

According to an early research on 3D-city modelling by European Spatial Data Research (EuroSDR) [1] (Fuchs et al., 1998), of all objects, buildings have obtained the most attention. The building model is also one of the most detailed thematic concepts of CityGML, which consists of five levels of detail (LoD).

In order to provide building information for applications such as map updating, 3D modelling and monitoring of informal settlements, etc., buildings need to be first detected. Reviews of the early researches on building detection can be found in Grün, et al.(1997), Mayer (1999), Baltsavias (2004) and Kaartinen, et al.(2005). In those times, the buildings were mainly detected from images. With the introduction of new data sources such as Light detection and ranging (Lidar), the focus of building detection shifted from one type of source data to another, or to the fusion of them. Data sources that have been used for building detection are single intensity images (Lin et al., 1994; 1995), single colour images (Shorter and Kasparis, 2009), multi-view stereo images (Fradkin et al., 1999), Digital Elevation Model data (Lafarge et al., 2008), Lidar data (Verma et al., 2006), or the fusion of them (Awrangjeb et al., 2010; Vu et al., 2009).

[1] EuroSDR was called European Organization for Experimental Photogrammetric Research (OEEPE) at the time of this research

The data used for building detection are typically recorded by nadir viewing sensors. Only the rooftops of buildings can be detected whilst the façade information is not available or only to a quite limited extent. Hence the problems of roof overhang or other façade-related problems are difficult to handle by using these data. As a result, terrestrial data is usually required during the 3D modelling of buildings to provide complementary façade information (Tian et al., 2010).

Around early 2007, airborne oblique imagery became another new data source in photogrammetry (Karbo and Simmons, 2007). The images are taken from aeroplane from tilted angles. Thus they can present the information of both building roofs and façades, combining some properties from both nadir and terrestrial view images. This new characteristic introduces a challenge on how to use the additional façade information to assist building detection or other applications.

Multi-view is another important feature of oblique images. They are usually captured from multiple directions with forward and sideward overlap. With this design, a certain building should be seen in multiple images from different perspectives. As stated in (Mayer, 1999), the 3D geometry of an object can be better recovered if it is viewed in more images since multiple images are able to provide more evidence, less occlusion, better camera configurations and mutual-verification.

Due to their advantages on presenting extra information on target objects, oblique images nowadays are widely captured around the world. Companies providing services of capturing and processing oblique images spread from North America (Pictometry, 2012) to Europe (COWI, 2012; Getmapping, 2012) and Asia (Ofek, 2012). However the number of applications of oblique images is still quite limited. Apart from data visualisation, other applications include simple measurements from images (Höhle, 2008), validating existing vector data (Mishra et al., 2008; Nyaruhuma et al., 2012) and texture mapping (Wang et al., 2008). A detailed review of oblique images can be found in Chapter 2. Yet so far, little research has been carried out on object recognition and modelling from oblique images.

In this thesis, multi-view oblique airborne images will be used for building detection. Attention will be paid to exploiting the usage of façades, which differentiates oblique views from nadir views. Considerations should be on combining façades from multiple perspectives to reduce the occlusion problem. Therefore, the initial building detection will be based on the detection of façades from multiple directions. In case not all façades of a building are detected, roof information will be added to complete its outline. Besides façade information, oblique imagery also introduces other new

characteristics compared to nadir imagery such as inconsistent scale and occlusion, which will also be studied in this thesis as well as their effects on building detection.

1.2 Research objectives

According to the domain of the research problems, the general objective and some specific objectives in detail are formed. The whole research is targeted to achieve the formulized objectives.

1.2.1 General objective

The main objective of the work is to develop methods for detecting urban buildings from multi-view airborne oblique imagery.

1.2.2 Specific objectives

The general objective can be further divided into the following objectives:

- To develop a method to locate buildings by using the façade information. Image features need to be selected to distinguish building façades from other objects such as building roofs or ground.

- To generate the 2D outline of buildings whose one or more façades have been detected, integrating all attributes from building façades and roof extracted from multiple images.

- To assess the effects of image properties, such as image resolution, forward and sideward overlap, on building façade detection and the accuracy of the outlining.

1.3 State-of-the-art in building detection

The robust detection of buildings from remotely sensed data is an important part of the automated interpretation of them. The detection result is the fundamental of building reconstruction and 3D modelling. Researches in this domain started in the late 1980s, but due to the complexity and irregular nature of the urban scenes, building detection remains a difficult task. Since the initial research, building detection has been propelled by the development of new data sources and methods in the fields of photogrammetry and computer vision.

Many researches on building detection use airborne images, based on the content in the images, such as intensities, textures and colour. Another group of researches is based on the 3D altimetry information from DEM or laser scanner data. Notice that the matched point cloud from stereo pair or multiple views is also considered as height data here. The literature is

reviewed in the sequence of detection from images only, from height data only and from the fusion of them.

1.3.1 Using image information

Airborne images present rich information about the real world, from which useful features related to buildings are extracted for the purpose of building detection. A large number of algorithms have been developed to extract various features and combine them.

The first group of research is the detection from grey level ortho images or only using image intensities. Stassopoulou and Caelli (2000) explored the use of Bayesian Networks to combine diverse information obtained from orthophotos, ranging from corners to solar angles to create and update the building layer in maps. Peng et al. (2005) detected buildings from high resolution images using an improved snake model. It initialised the seeds close to the building edges, which have higher energies in the model. Then it integrated illumination direction to improve the final outline. Katartzis and Sahli (2008) identified building rooftops from images based on an interpretation model which combines both 2D and 3D contextual information of the imaged scene.

Another group of work detected buildings based on the classification using colour values. They performed the segmentation on the original images employing various strategies. Other information, such as vegetation and shadows, was used to assist the final detection. For example, Müller and Zaum (2005) combined geometric, photometric and structural features from image segments to operate the classification. In the work of Sirmacek and Unsalan, (2008), invariant colour features were used for the images segmentation. Then the extracted edges and shadows were also used for the building detection. A box model was adopted to determine the building shape. Similarly, Shorter and Kasparis (2009) proposed a method to classify a single colour image into buildings, vegetation, shadow, and other features. It firstly located the candidate buildings by employing watershed segmentation on the entropy filtered image. The vegetation and shadow were removed afterwards by the combination of RGB bands.

Because the lack of assistance of 3D information makes the detection of buildings from a single image more complex and less reliable, researchers usually try to look for other supplementary information. Lin (1994; 1995) extracted linear features from images and used perceptual grouping process to generate building hypotheses, which were further verified them using shadow and wall evidence. Croitoru and Doytsher (2004) included the prior knowledge in the process that most of the buildings have right-angle corners in regularized urban areas. Karantzalos and Paragios (2009) combined

image-based costs with prior constraints, related with the geometry of the objects of interest.

Detection from monocular images would be easily misled by the context of one single image, whilst detection from multiple images employ more evidences for mutual-validation. When using images recorded from almost the same perspective, some works firstly generated building hypotheses in one single image and then verified them in other images (Collins et al., 1998; Nevatia et al., 1997; Noronha and Nevatia, 2001). These approaches are not to perform complete stereo analysis but to merge the higher level structures in different images only. More recently, height evidences from stereo or multi-view matching were integrated in the process of detection, which will be discussed below in section 1.3.3.

One early research with oblique airborne was presented in a paper by McGlone and Shufelt (1994). In that paper, an approach was presented where vertical and horizontal lines were extracted in monocular oblique images after the vanishing points had been determined. The lines were then grouped to form building hypotheses. The hypotheses were verified using evidence from other images. This work was at an early stage, so the detection rate was at a quite a low level, also probably due to the fact that only cues from lines were employed.

Another similar work was done by Shufelt (1996) using a single view. The vanishing points were determined from parallel lines, and primitive building models were then composed from 3D lines with vertical and orthogonal horizontal orientations in object space. The verification of the models was done by computing the consistency of inter-surface intensities and the shadow-ground transition. This work utilized the geometry in the detection, but it was limited to simple gable roof buildings.

Overall, progress has been made on building detection from images, employing evidences of corners, walls, shadows etc.. However these works mainly aim at rural areas with lower building density and complexity. Still the detection rate is at a rather low level. This low level is probably due to limitation of only using 2D clues for the detection of 3D objects.

1.3.2 Using height data

Height data is another data source for building detection. In the early days, height information was provided by DEM. Later laser scanning data or Lidar data appeared as a new form of height data which is very precise in height. Boosted by the highly overlapping large image blocks, dense matched point clouds currently have become another source of height data (Leberl et al.,

2010). Here, height data means all kinds of forms which can provide altimetry information.

Weidner and Förstner (1995) investigated building detection from a DEM, generated from stereo pairs of aerial images. Their research focused on simple buildings, which were modelled by parametric and prismatic models. The boundaries of the prismatic models were formed by the minimum description length (MDL). Ma (2005) generated DEM from Lidar at first by converting it into grids. Building boundaries were detected from the DEM and regularized by MDL.

Rather than generating DEM from original data, Maas (1999) tested the usability of laser scanning data directly on object detection. He calculated height texture (variety of height) using different algorithms and used the results of them as different bands for the classification of buildings, trees and ground vegetation.

Another more popular way of using laser scanning data is to use its geometric features. Dorninger and Pfeifer (2008) detected buildings directly from Lidar data by firstly segmenting the point cloud into planar faces. The planar segments were subsequently grouped into building roofs. The roof outlines were then generated by the 2D α-shape and regularized afterwards.

New methods nowadays have been implemented in building detection from height data. Lafarge (2008) extracted building outlines using marked point process from DEM generated from 3-view images. First, a rough approximation of the building footprints was realized by a method based on marked point processes: the building footprints were modelled by rectangle layouts. Then, these rectangular footprints were regularized by improving the connection between the neighbouring rectangles and detecting the roof height discontinuities. A similar work was done by Tournaire et al. (2010) where buildings were also initially presented as rectangles. Then a function was defined to determine a global energy of the footprint configurations. A Reversible Jump Monte Carlo Markov Chain (RJMCMC) sampler coupled with a simulated annealing algorithm was used to minimize the energy so as to achieve the best building outlines.

Lafarge and Mallet (2011) introduced a label propagation procedure with geometric constraints to detect buildings. It firstly segmented the point cloud involving discriminative features. Then shapes and lines were extracted from the segments. Buildings, trees, and ground were classified under geometric constraints in a 2D grid space arrangement, whose initial labels came from the labels from points projected onto it. The labels were propagated using a Markov Random Field (MRF) with pairwise interactions, whose adjacency set

was given by a breakline-dependent neighbourhood. The break lines came from the segments extracted from the point cloud, and were used as hard constraints during the propagation.

Point clouds from multi-view stereo matching are a new source of altimetry data. Yet their use for object recognition is not fully exploited. Lafarge and Mallet (2011) made a comparison between Lidar and 3D clouds of photogrammetrically matched points using the proposed process. Compared to Lidar, photogrammetrically matched points have less altimetric accuracy but do not have strongly marked discontinuities on building façades. Because of this difference, Lafarge applied a softer parameter setting for matched points than the setting for Lidar in the procedure, such as using larger threshold for the distance between points and extracted planes in the segmentation procedure. The proposed algorithm globally provided better results from Lidar than from image point cloud.

General speaking, when merely using height information, vegetation is inevitably the major error source for over detection while missed detections are mainly caused by small flat buildings.

1.3.3 Integrating image info and height data

Some researches integrate images and height information to compensate for disadvantages from using them separately. The obvious advantage of the integration is that the data has both the 3D geometric properties and the radiometric values.

One typical combination of height and images is that vegetation and buildings can be initially extracted from ground, and then the image information can be used to further distinguish them. The combination was initially carried out by using the 3D scene from stereo image pair and the radiometric properties from the source image (Baillard and Maître, 1999), in which DEM were generated from stereo matching. Gerke also published an approach on automatic extraction of buildings using aerial orthoimages image and nDSM (Gerke et al., 2001). In that paper, an nDSM was used to separate ground from non-ground objects, while vegetation was separated from sealed areas using the NDVI.

3D features such as 3D lines can be extracted to assist building detection as long as overlapped multiple images are available. Kim and Nevatia (2004) detect buildings by using both the extracted 3D features (e.g. 3D lines) of the images and DEM derived from them. First, 3D features were generated by grouping image features over multiple images, and hypotheses were generated by neighbourhood searches on those features. Hypotheses in 3D were then verified with evidence collected from the images and DEM.

Expandable Bayesian networks are used to combine evidences from multiple images.

A simple combination of colour image and laser altimetry data was to take the height as another band added to the image for the object classification (Haala and Brenner, 1999).

The general combination of Lidar and images is to use the Lidar points for the shape recognition and then distinguish trees from buildings by their radiometric difference in the images (Hu et al., 2004; Sohn and Dowman, 2003). However, Vögtle and Steinle's research (2000) presented another way around. They used spectral and shape and size parameters to remove non-building objects, and then implemented the DEM segmentation for building hypothesis.

Vu et al. (2009) presented a multi-scale solution based on mathematical morphology for extracting building features from elevation data and spectral image data. The structural information from elevation data were firstly represented on a morphological scale-space. A complex structure could be extracted as a multi-part object in which each part was represented on a scale depending on its size. The building footprints were represented by the boundary of the largest part. The spectral data was used as an additional source to remove vegetation and possibly classify the building roof material.

An implementation of integrating all Lidar, DEM, colour images was proposed by (Awrangjeb et al., 2010). The three sources needed to be registered to each other before the processing. The primary and secondary building masks were first derived from the Lidar and DEM along with NDVI values from the ortho-imagery. The initial building positions were derived from the primary building mask. The final buildings were obtained by extending their initial positions using the two masks and colour information.

A common problem in data fusion is how to integrate extracted features from all data sources. There are some popular methods implemented in the recent research. Five of them were compared in the work of Khoshelham et al., (2010) to classify the space into buildings and non-buildings: thresholding nDSM, Bayesian method (maximum likelihood, minimum distance), Dempster-Shafer method and Adaboost algorithm. The source data were aerial ortho-rectified images and laser scanner data, and the results were analysed at pixel- and object-level. The results showed a better performance of the Dempster-Shafer method followed by the AdaBoost, but both of them yielded high rates of unclassified pixels.

1.3.4 Summary

Detailed results from some reviewed papers are listed in Table 1.1. As the researches were implemented over different study areas with quite different building type configurations, the statistics are not directly comparable. However the averaged statistics over data types can be used for the rough comparison of detection quality.

Additionally, an ISPRS benchmark in 2012 provided pan-sharpened colour infrared (CIR) images with a GSD of 8 cm and airborne laser scanning (ALS) data of 4-7 points/m2 for object classification and 3D building reconstruction. The report (Rottensteiner et al., 2012) presented the research results from several research groups on building detection and modelling. Some participating groups used ALS only or the fusion of ALS and images, but no one purely used image data. The results over all tested areas are also listed in Table 1.1.

Through these results review and comparison, we attempt to statistically find out the results that can generally be achieved with one data source. This can also be used to guide the evaluation of our detection results from oblique images.

The completeness and correctness are the most commonly used indices for evaluating the result of detection, defined as completeness: $\text{True Positives} / (\text{True Positives} + \text{False Negatives})$ and correctness: $\text{True Positives} / (\text{True Positives} + \text{False Positives})$. X in the table means that data is not available. During the analysis, we exclude the best and the worst result in the analysis, since they were probably caused by some special conditions, such as very uniform building types or very poor quality data sources.

Overall speaking, the detections from single images result in relatively low detection rates. Many of them are only about 50% to 60%, and the better ones are only 70% to 80%. However, when the height information is included, the results are significantly improved. It is quite reasonable since the approaches with height are then done in the object space. This also is consistent with the statement in (Mayer, 1999) that the extraction should be done in object-space since much of the knowledge of the real world can be used.

The combination of imagery and height data inherits both the advantage of building detection in merely height data of 3D clues and in single imagery of colour information, and it therefore should be considered as the most suitable data sources for building detection.

After reviewing research on building detection, it is obvious that image information without the use of auxiliary 3D information from stereo matching or DEMs is not sufficient to generate higher quality results. More attention is given to Lidar or the fusion of Lidar and images, since the 3D information plays an important role in extracting buildings clues. However there has been little research only using multiple images for detection. Oblique images implicitly enable the exploitation of 3D dimension through the visibility of façades. Because of the accessibility of façades we can also theoretically collect more evidence per building compared to only using roofs. In addition we can produce a point cloud out of multi-view oblique images. Therefore detections from oblique images using 3D information from multi-view matching as well as the radiometric information should be able to achieve a similar detection results from the fusion of Lidar and images.

Table 1.1 Results of building detection from literatures

Work	Pix Comp (%)	Pix Corr (%)	Blding Comp (%)	Blding Corr (%)	Blding No.	Data	Resolution
(Lin and Nevatia, 1995)	68.3-96.5	X	54.2-87.5	73.7-97.2	X	Single intensity image	X
(Nevatia et al., 1997)	70	92	70	94	X	Single intensity image	X
(Müller and Zaum, 2005)	77.3	79.4	X	X	240	Single colour image	30 cm
(Sirmacek and Unsalan, 2008)	X	X	86.6	X	177	Single colour image	30 cm
(Shorter and Kasparis, 2009)	78.7	51.6	55.4	48.2	2643	Single colour image	15 cm
(Nevatia et al., 1997)	70	92	70	94	X	Multiple images	X
(Noronha and Nevatia, 2001)	97.1	95.7	93.5	88.6	218	Stereo pair	30-130 cm
	X	X	94.2	100	65	Three images	
(Maas, 1999)	X	X	96.6	99.8	5425	Laser scanner data	> 1 p/m^2
(Lafarge et al., 2008)	X	X	84.7	90.3	X	DEM from satellite images	X
(Tóvári and Vögtle, 2004)	X	X	95	96	X	nDSM from laser scanning data	X
(Tournaire et al., 2010)	X	X	83.8	85.7	X	DEM	50 cm GSD
(Vu et al., 2009)	X	X	67.4-83.0	73.3-79.6	87	Lidar Images	1 p/m^2 20 cm
(Awrangjeb et al., 2010)	78.3	88.9	97.1	97.9	X	Last pulse Lidar DEM Images	0.5m spacing 1 m 15 cm
ISPRS benchmark (Rottensteiner et al., 2012)							
J. Niemeyer (HAN)	87.0-93.8	86.9-91.4	78.6-83.8	72.1-100	107	Lidar	Lidar: 4-7 p/m^2 Images: 8 cm
P. Dorninger (VSK)	85.4-86.3	98.1-98.7	75.0-78.6	100			
C. Liu (TON)	67.8-88.5	95.7-98.9	55.4-75.7	93.5-100			
D. Grigillo (LJU)	93.2-95.1	94.1-95.4	81.1-85.7	100		Lidar Images	
Q. Zhan (WHU)	76.9-84.4	83.9-92.6	57.1-78.4	43.5-79.2			
W. Yao (TUM)	86.8-89.8	90.1-93.9	75.0-89.2	91.7-100			
A. Moussa (CAL)	87.0-93.2	94.7-95.4	66.1-83.8	100			

1.4 Scope and limitations

This research aims at detecting all buildings in the urban area, without limitations on building types. However, we would like to distinguish buildings from sheds or other man-made platforms, so that only the constructions larger than 5 meters × 5 meters with height no lower than 3 meters are defined as buildings. Buildings smaller than this size are not considered.

The detected buildings are finally presented using 2D outlines, consistent with the form of object detection (Rottensteiner et al., 2012). 3D information

is sometimes available, but they are not quite accurate with all details. Thus we do not include the 3D information in the detection result. As the targeted applications are generally focus on smaller scales, tiny structures attached to the main building are not extracted.

1.5 Outline of the thesis

An introduction of oblique imagery is given in the next chapter. It contains the general characteristics of the imagery and its modern applications. Then the available datasets are described and compared in detail. Some common processes on the oblique datasets are also included in that chapter.

After Chapter 2, the main scientific contributions of the thesis are described in two major parts. Each part corresponds to one stage of the developed detection process, which consists of two separate chapters on method and results respectively.

Part I focus on the building façade detection from imagery. Chapter 3 in this part elaborates the selected image features and the approaches how to combine these feature to generate 3D façades. The façade detection results are presented and analysed in Chapter 4. By assessing the results, the image attributes for façade detection are assessed.

Part II covers the aspect of outlining the detected buildings. Chapter 5 and Chapter 6 are organized in the same way as the two chapters in Part I. The developed method specifies the features extracted from 3D space and 2D images for building outlining. Then a graph-cut method is used to integrate these features. In Chapter 6, the results are shown and evaluated. Some simulated image configurations are also designed in Chapter 6 to assess the image attributes on the accuracy of the final outline. The thesis is finally concluded by a chapter that contains the general conclusions and recommendations.

Chapter 2 Oblique Imagery

Oblique airborne images are photographs taken at tilt angles to the earth's surface. Unlike nadir images, which mainly present top views of objects, for instance roofs of buildings, oblique images display more sides of objects. As they are closer to the nature of human view, oblique images are intuitively easier to interpret for untrained users. Figure 2.1 shows an example of conventional nadir image and oblique image over the same area. The oblique image clearly present the height variation and façade structure of buildings.

(a) (b)

Figure 2.1 Comparison between (a) nadir image and (b) oblique image (from Bing Maps)

2.1 Brief review on developments of oblique images

The invention and usage of oblique images dates back to decades ago, in World War I or even earlier. Since 1940s, Trimetrogon photos, one type of oblique photography, have been used to map Antarctica (TMA, 2012). Then in the 80's, oblique photographs were used in military and archaeology for reconnaissance purposes (Grenzdőrffer et al., 2008). However, due to their inconsistent scale and occlusion problem, they were of little use in Photogrammetry.

Pioneer work on building detection from aerial oblique images appeared in the 90's. In that period, experiments were mainly implemented on single oblique image. Principle photogrammetric theory on vanishing points was incorporated to detect building structures (McGlone and Shufelt, 1994; Shufelt, 1996). Shadow and wall evidences from oblique views were also explored for building description (Lin and Nevatia, 1995; Nevatia et al., 1997).

Multiple photographing from all directions is a possible solution for the occlusion problem. In the time of analogue approaches, it was restricted by high expense and intensive labour requirements. However, triggered by developments in digital photographing techniques, as well as by the demands by applications of bird's view in the digital presentations of the earth, the development of technologies for handling oblique images has progressed very quickly in the recent years.

Petrie made an extensive review of recently developed technologies for acquiring multiple digital aerial oblique images (Petrie, 2009). According to Petrie's review, the oblique imaging configurations are classified into three classes: fan configurations of digital cameras, groups of digital cameras set in block configuration and combination of near-vertical and oblique points cameras. The "Maltese Cross" is one distinctive configuration in the third class (Figure 2.2), composed by one vertical camera and four cameras with large oblique views.

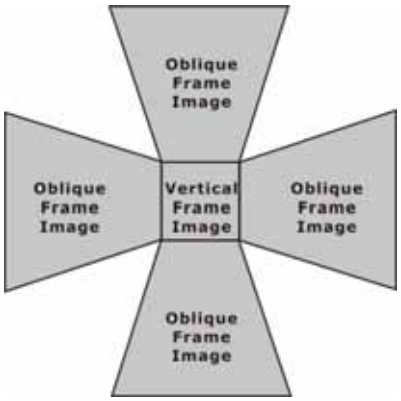


Figure 2.2 "Maltese Cross" ground coverage of a five-camera system producing a single vertical image and four oblique images (Petrie, 2009)

In the meantime, companies over the world have also started capturing and processing oblique images. As the most famous company, Pictometry (2012), started on oblique images since 2005. It is stated that their images have covered more than 85% of the populated area of the United States. The company also brought out their desktop application on viewing and measuring oblique images – Pictometry® Electronic Field Study™(EFS). EFS enables direct measurements of coordinates, elevations, heights, distances and area. The principles of the measurements and accuracy analysis are described in (Höhle, 2008).

Next to Pictometry, several other companies are competing in the market, some also with their own software packages. Ofek from Israel (Ofek, 2012)

developed the MultiVision technology to display and analyze spatially anchored oblique aerial photographs based on vertical aerial photographs. MultiVision also provides image data and analysis on it (see details in Grenzdőrffer et al., 2008). The quality of the measurements is restricted by the quality of the underlying DEM. PixoView technology at GEODIS BRNO has also started acquiring and processing oblique photographs in Czech Republic since 2006 (GEODIS, 2012). Between 2010 and 2011, Slagboom en Peeters acquired oblique images of a few cities in the Netherlands (Slagboom en Peeters, 2012).

2.2 Characterization

Oblique images show some special characteristics concerning objects' appearance compared to nadir images. Some characteristics closely related to Photogrammetric processes will be discussed below.

2.2.1 Tilt angle

Oblique images are traditionally distinguished by their tilt angle, while "The image tilt angle is defined as the angle formed by the optical axis with respect to the Z-axis of the world coordinate system" (Shufelt, 1999). Based on the definition of tilt angle, aerial images are classified into three classes (Shufelt, 1999):

- *vertical* photograph: tilt angle less than five degrees;
- *low oblique* photograph: tilt angle between five to thirty degrees;
- *high oblique* photograph: tilt angle larger than thirty degrees[†].

In practice, tilt angles are selected in such a way that the horizon is not included in the field of view. This is because it would lead to unbalanced lightning conditions, and for practical applications the horizon area is mostly not of interest.

2.2.2 Varying scale

Due to the tilt shooting angle, oblique images have varying image scale, given that the area is flat and horizontal. Refer to Figure 2.3, bound of the captured area is approximately a trapezoid. Pixels in the foreground of an oblique image depict a smaller object area compared to pixels in the background. In other words, the resolution in the foreground is finer than in the background.

[†] Without specification, oblique images mentioned in this thesis afterwards are by default high oblique photographs, but they normally do not contain the horizon.

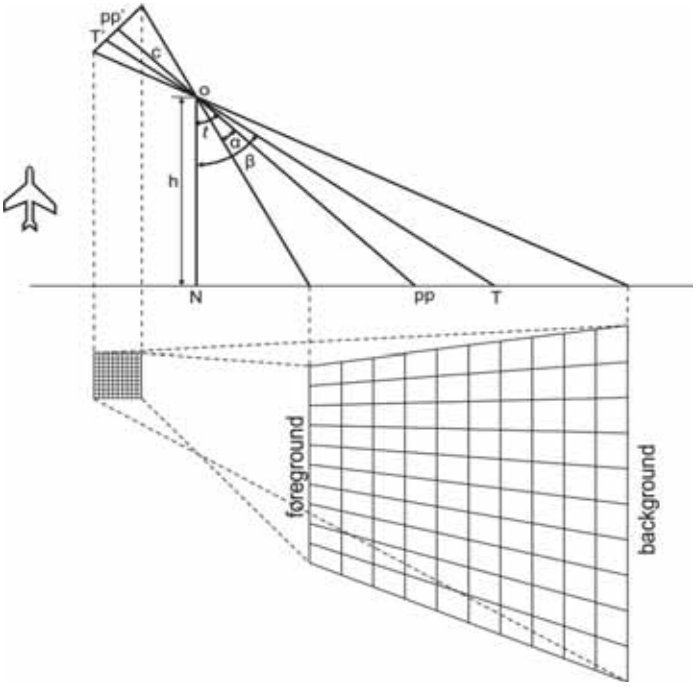


Figure 2.3 Geometry of oblique image (modified according to the Fig. 1 in (Höhle, 2008))

Symbols used in Figure 2.3 are listed below:

- N – the nadir point;
- pp – the intersection of the camera axis with the terrain;
- pp' – the principle point of the images;
- c – the camera constant;
- h – the flying height above ground;
- t – the tilt of the camera axis;
- T – a target point on the terrain;
- T' – the imaging of the target point T in the image;
- α – the half field of view;
- β – the angle between a direct line from the lens to T and the vertical direction;

Then the image scale factor m_T at T is defined as:

$$m_T = \frac{h \cdot \cos(\beta - t)}{c \cdot \cos\beta} \qquad \textbf{(2-1)}$$

The minimum scale m_{min} of an oblique image is at the foreground and the maximum scale m_{max} is at the background, thus:

$$m_{min} = \frac{h \cdot \cos\alpha}{c \cdot \cos(t - \alpha)} \text{ and } m_{max} = \frac{h \cdot \cos\alpha}{c \cdot \cos(t + \alpha)} \quad \textbf{(2-2)}$$

2.2.3 Occlusion

Occlusion is another remarkable characteristic of oblique images. In principle there are two occlusion problems: self-occlusion and occlusion by other objects. The former is common for every object, while the latter is caused by high-rising-objects in front of the target object towards the camera. As shown in Figure 2.4(a), a big building is composed by a higher part (marked by purple circle) and a lower part (marked by blue circle). The red line segment indicates a vertical facade of the higher part, which is also marked by a red parallelogram in Figure 2.4(c). To the west of this building, there is another lower building marked by orange circle. In the oblique view from west (Figure 2.4(b)), the facade and the lower part of the building is self-occluded. The occlusion by other objects is then presented in Figure 2.4(c), where part of the building marked by orange circle is occluded by the big building.

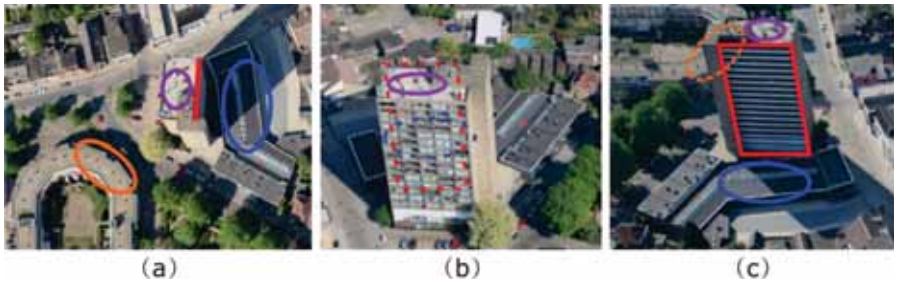


Figure 2.4 Occlusion problem caused by oblique view. (a) nadir view; (b) oblique view from west; (c) oblique view from east. Solid lines – visible objects, dashed lines – occluded objects

Visibility of an object also depends on the flight height and the tilt angle of the camera. At a constant focal length the occlusion problem becomes more serious when the height decreases or the tilt angle increases.

Although occlusion is unavoidable, it may be compensated by images taken from multiple directions. Normally multiple images from all sides have to be acquired and processed to obtain full information (Grenzdőrffer et al., 2008). In practice, the photographic strips are designed along one principle direction, which is north-south or east-west with a camera of four oblique heads. By this means, the collected images are always facing four cardinal directions.

2.3 Modern Applications

Modern applications of oblique images based on new imaging developments can be assumed as just started from the beginning of this century. Other than the previous visualization and interpretation purposes, people have already foreseen the value of their applications in the fields of 3D modeling, cadastral mapping, planning, emergency response, navigation, etc.

By revealing the appearance of building facades, texture mapping is by nature the most direct and common application of oblique images (Frueh et al., 2004; Jurisch and Mountain, 2008; Kim et al., 2010). Commercial software such as MultiVision (Grenzdőrffer et al., 2008) and TerraScan (Terrasolid Ltd., 2012) also embed (semi-)automatic building texturing for 3D models. In general, model texturing consists of the following steps: locate single facade from input model; then for each facade select the best image for texture clipping, including an occlusion test; if necessary, some invisible facade parts need to be completed from nearby structures; the texture is finalized by image enhancement and mosaicking.

Applications in the cadastral domain are another area of interest attracting researchers. A systematic discussion on the potential of using oblique images within the context of cadaster is provided in (Lemmen et al., 2007). Although the paper aimed specifically on Pictometry data, it is also extendable to modern oblique images in general. Work integrating oblique images in the cadastral domain can be categorized as follows.

Measurement from oblique images: measurements from oblique images are the basic issue of making use of them. Therefore various evaluations on measurement from oblique images were firstly tested. For example according to (Höhle, 2008), the root mean square error (RMSE) of measurements from single Pictometry images (monoplotting) were 1.5 m and 0.6 m for planimetric coordinates and height respectively. The value of these two measures were 6.14 m and 0.38 m found by Prandi et al. (2008) and 0.86 m for planimetric coordinates by Lemmen et al. (2007). The measurement accuracy of monoplotting depends highly on the quality of the used DEM, which may explain the relatively large difference on the accuracy. The underlying DEM, is, however, not directly accessible within the used software EFS, and thus further evaluation is not possible. In addition, the quality of image orientation parameters has a large influence on the measurements.

Verification of existing data: an approach to validating road vector data was presented in (Mishra et al., 2008). It used single oblique image, and employed 3D building models for occlusion tests. Nyaruhuma et al. (2010) developed a robust method for building verification from 2D cadastral maps.

The research integrated several evidences from multiple images from different views to reason the existence of buildings.

Recognition of new objects: object detection from modern high oblique images has not been covered yet but (Fradkin et al., 2001) detected buildings from multiple overlapped **low** oblique images. Because facades are narrowly available in those images, the resolution on them is not high. Therefore it could not use detailed facade structures for the detection process. Point clouds were matched from multiple views. Then the facades were extracted through vertical accumulation of points. The building hypotheses generated from extracted facades were lastly verified in multiple images. The available images in this research are **high** oblique images, so that we exploit the usage of the facade structures in the detection.

Disaster management: given that facades are visible in oblique images, they are considered to be much more helpful than traditional nadir images for disaster management. Pictometry images were used in (Gerke and Kerle, 2011) for damage assessment after the Haiti earthquake in January 2010. They defined building damage categories by classifying intact facade, intact roofs and destroyed roofs on each images, and then combined the classification for the per-building assessment from images of all perspectives.

2.3.1 Available datasets

As specified in the research objectives (Chapter 1.3), oblique images will be used for automatic building detection. Two data sources over Enschede, the Netherlands are available: Slagboom en Peeters dataset from 2011 and Pictometry dataset from 2007.

2.3.2 Slagboom en Peeters (SenP) dataset

Slagboom en Peeters B.V. is a company in the Netherlands, providing photography flights for customers in the Netherlands and Belgium. The single engine Cessna 180J airplane is used for the geo-oblique aerial photography. The image capturing system uses the five-digital-camera configuration: one directed nadir and the others viewing forward, backward, left and right. The ground coverage of this configuration forms the "Maltese Cross" shown in Figure 2.2.

The flight over Enschede was made in late April 2011. The flying height was around 350 m, leading to a ground sample distance (GSD) from 5 cm to 10 cm in one image. The overlap between images was about 60%. In this way, one object can be viewed in at least eight images.

The nadir looking camera is mounted with a wide angle lens with a focal length of 21 mm. One nadir image covers around 350 m × 550 m area.

2.3.3 Pictometry dataset

Pictometry is an aerial image acquisition and data processing technology developed and patented by US-based Pictometry International Corp. As patented (Schultz et al., 2011), the Pictometry imaging system contains an image capturing system and geo-locating systems – a global positioning system (GPS) receiver, an inertial measurement unit (IMU), clock, gyroscope, compass and altimeter. The images were captured also with the five-camera configuration.

The dataset use in this research was made available through (Blom, 2012), the licensed company of Pictomety in Europe. They were captured in mid February 2007 at the flying height of 920 m. The nadir images provided by Pictometry are only orthorectified ones. Since neither the original perspective images nor information about the used DTM is available, they are not used in this research.

The oblique images were taken by cameras with focal length of 85 mm. The tilt angle is around 50 degrees, and the GSD for foreground and background are 10 cm and 16 cm respectively. For every point on the ground, at least one image per viewing direction is available. However the along track overlap is not very high, leading that the stereo overlap could not be guaranteed. Therefore, normally four to eight images are available for a target object.

2.3.4 Comparison between the two datasets

Table 2.1 shows the important parameters of the SenP and Pictometry datasets. It can be seen that the resolution of SenP is about two times finer than that of the Pictometry images (Figure 2.5). Correspondingly, the image coverage of SenP in object space is much smaller than that of Pictometry (Figure 2.6).

Table 2.1 Parameters for oblique images from SenP and Pictometry

	SenP	Pictometry
Flight height (m)	350	920
Tilt angle (degree)	45	50
Focal Length (mm)	51	85
GSD (cm)	5 – 8	10 – 16
Image Size	5616 ×3744	4008 × 2672
Image overlap	about 60%	about 20%-30%

Figure 2.5 (a) SenP and (b) Pictometry © Blom images viewing the same building, maintaining resolution differences.

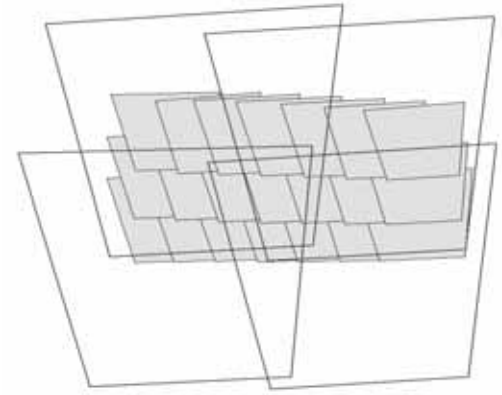


Figure 2.6 Coverage in object space of northward images of SenP and Pictometry. SenP: solid; Pictometry: hollow.

2.4 Image orientation

Approximate image orientations were provided for both SenP and Pictometry data sets. A semi-automatic aerial triangulation as well as camera self-calibration was performed. The modeled interior orientation contained the location of the principle point and the focal length as well as the aspect ratio of the pixels and the radial distortion for each camera.

Several ground control points (GCP) were collected by using differential GPS measurements in the field. The accuracy can reach 0.02 m. In order to relieve the efforts on GCP collection, scene constraints, including horizontal and vertical lines as well as right angles, are incorporated to improve the orientation (Gerke, 2011). Tie points measurements are supported by a scale-invariant feature transform (SIFT) matching algorithm (Lowe, 2004).

Some additional tie points are extracted in Image Modeler. Then they are optimized to a maximum total number and distribution. Least square bundle adjustment is applied integrating tie points, GCPs and the scene constraints for the image orientation and self-calibration.

After image orientation, the residuals in three directions for SenP and Pictometry are shown in Table 2.2. The accuracy of the aerial triangulation is 15 cm (2 to 3 pixels) for SenP and 20 cm (1 to 2 pixels) for Pictometry.

Table 2.2 Residuals (RMSE) at GCPs in three directions

	SenP	Pictometry
x (cm)	12	17
y (cm)	15	20
z (cm)	14	16

The difficulties we were facing in image orientation are:

- The tie points matched by SIFT are not so reliable since the SIFT matching method is not robust on perspective changes. It is difficult to separate inliers and outliers, so that the remaining outliers influence the final bundle adjustment.
- The initial exterior orientation is not accurate, which leads to difficulties in the approximations.
- The uncertainty in the self-calibration is another reason for the uncertainty in the image orientation.

The image residuals in SenP are larger than for Pictometry, thus the orientation result is worse for SenP data compared to Pictometry. Firstly the problem with the separation of in and outliers is much more critical in SenP because of the worse initial EO information from SenP. Secondly, fewer tie points and scene constraints (manually) are measured for SenP.

2.5 Point cloud generation from multi-views

Multi-view stereo matching is an important way for automatic acquisition of surface descriptions from images. Thereby quite a lot of algorithms exist on it in the field of photogrammetry and computer vision.

An approach to dense matching oblique airborne images is presented in (Le Besnerais et al., 2008). Pixel-wise similarity criterion was developed for the special viewing geometry of oblique images. A dense depth map is obtained through global regularization. This approach mainly aims at images with low resolution, which have the GSD of several meters.

Another approach works with the same high resolution airborne oblique images from Pictometry and FLI-MAP4 image (Gerke, 2008). It evaluates the overall accuracy of the 3D point cloud derived using the Semi-Global-Matching technique (Hirschmüller, 2008). Around 70 percent of all matches can be within a range of ±3 pixels to the reference.

We select Patch-based Multi-View Stereo matching (PMVS) (Furukawa and Ponce, 2010) to generate all point clouds used throughout the research. This method does not require any initialization of the object but the (relative) orientation of images, which also fits our aim of developing an unsupervised approach. The density of the generated points depends on the image resolution and the input parameters. We use different settings of the parameters and the control of the input images to meet the characteristics of the datasets and different requirements in the stages of the work. The method is firstly briefly described. Then the qualities of the point clouds from different settings with different datasets are simply evaluated as a guide for the following work with them.

2.5.1 Point cloud generation method

Referring to the work by (Furukawa and Ponce, 2010), the PMVS process consists of three steps:

1) *Matching*: features are founded by Harris and difference-of-Gauss operators in each image, and then matched across multiple images. This step results in a sparse set of oriented patches covering the surface.

2) *Expansion*: starting from the initial patches, the existing patches are repeatedly spread to nearby empty space to obtain a dense set of patches. This generates at least one patch in each image cell, which is a polygonal mesh model.

3) *Filtering*: visibility (and a weak form of regularization) constraints are used to remove erroneous patches, so as to refine the mesh.

Some key parameters control the density and accuracy of the generated point cloud:

- *level*: it specifies the level in the image pyramid that is used. *level* 0 is the full resolution, but due to the Bayer pattern in the digital cameras, *level* 1 is more proper than *level* 0. Increasing the value of level can significantly speed-up the computation but results in sparse points and reduced accuracy.

- *threshold* and *wsize*: these are the parameters for computing the photometric consistency over images. The algorithm samples *wsize* × *wsize*

pixel colours in each image. If the photometric consistency (normalized cross correlation) scores higher than the threshold, the patch is accepted as the initial patch.

- *csize*: the image cell size controls the size of the mesh. The smaller the value, the denser the point cloud. But it may lower the accuracy if the expended mesh is much denser than the initial patches.

- *minImageNum*: Each point must be visible in at least *minImageNum* images for being reconstructed. A minimum of two images is enough for the matching. Increasing the minimum viewing image numbers for each matched point can highly increase the quality of the matching result.

- *maxAngle*: It presents the certain amount of baseline for reconstructed points. The baseline is measured by angles between directions of visible cameras from each point (Figure 2.7), so that a point is not reconstructed if the maximum angle between directions of two visible cameras is below the threshold.

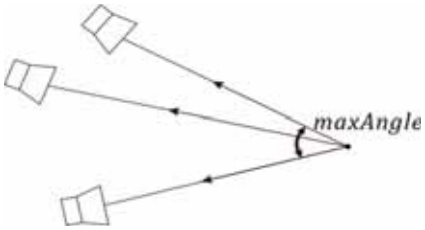


Figure 2.7 Sketch of *maxAngle* defined in PMVS

2.5.2 Different point clouds and quality assessment

Based on data sets and usages, different settings are defined to generate point clouds. A simple assessment of their quality is implemented to have a rough idea on the noisiness of different point clouds and their comparison. Both roof planes and façades of buildings are integrated into the assessment.

A small area with several buildings is selected (Figure 2.8(a)), containing five gable roof buildings (1-5), three flat roof buildings (6-8), one skillion roof building (9) and one flat roof with small structures (10). Building 10 is not used for the assessment because the structures on the roof may cause bias on the plane estimation. Figure 2.8(b) presents the 3D models of the selected buildings. The roof planes are obtained from Lidar, which has an absolute point accuracy of 3-5 cm (Oude Elberink and Vosselman, 2011). The façades are vertical planes reconstructed from the cadastral map which represent the exact positions of the building facades. This overcomes the problem caused

by roof overhang, thus facilitating the assessment of the point cloud on the façade.

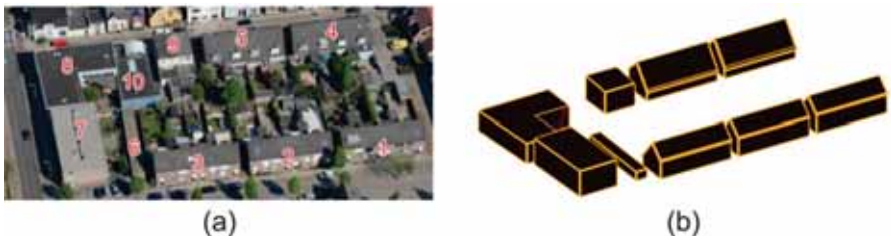


Figure 2.8 Selected area for point cloud assessment. (a) selected area in one SenP image; (b) 3D model of the selected buildings.

Two cases are simulated for the comparison: 1) point cloud generated from images viewing one direction and 2) point clouds from images viewing all directions.

Point cloud from images viewing one direction

As one façade of a building is normally only visible in images taken from one viewing direction, images from one particular direction are used to generate point clouds. This one direction point clouds have one special property relating to the occlusion problem. Since the points come from images from the same direction, those parts of the objects that are subject to (self-) occlusion are left out, which makes the matching between 2D and 3D significantly easier. Nevertheless the short baseline between images from single direction would result in a relatively bad intersection angle in object space. In the experiment, images viewing east are used for the assessment of one direction.

Point cloud from images from complete four directions

Point clouds from all directions are generated to obtain a complete geometric model of the buildings in the scene. It is useful for reconstructing building outlines and roofs in 3D. However, compared to Lidar, it is noisier and less complete at the places with no texture. Therefore approaches which require homogenous point spacing may not work.

The PMVS parameters *threshold*, *csize* and *maxAngle* are set to 0.7, 1 and 5 respectively to obtain accurate and dense point clouds. The *maxAngle* 5 is reasonable in case that all images from one direction have similar perspective. Setting a too high value for it would lead to many skipped matches. The *wsize* is set to 11 for SenP and 9 for Pictometry to guarantee the accuracy while also keeping the computational time at a reasonable level. Different settings are made by giving certain values to *minImageNumber* and *level*. Then the

results from those settings are compared to determine the proper value for them.

The assessment is based on the point density, signed mean distance (d_{mean}) and the RMSE (σ_p) from the points to the plane in the model (Figure 2.8(b)). Positive mean distance indicates that the plane is outer the building whilst negative sign indicates that the plane is inner. The median of the values from all planes is used as the value for one setting.

Table 2.3 lists the specified settings and their corresponding results along with the setting number. The detailed results for each plane for all settings are graphed in Figure 2.9.

Table 2.3 Settings for point clouds from PMVS and quality assessment

Direc-tion(s)	Dataset	*wsize*	*minImage Number*	*level*	Point Density (per m^2)	d_{mean} (m)	σ_p (m)	Setting number
East	SenP	11	2	1	29.1	0.10	0.76	**(1.1)**
		11	2	2	12.8	0.07	0.99	**(1.2)**
		11	3	2	9.2	0.14	0.63	**(1.3)**
	Pictom-etry	9	2	1	6.7	-0.68	1.04	**(1.4)**
All	SenP	11	2	1	45.8	0.05	0.43	**(2.1)**
		11	2	2	15.8	0.01	0.54	**(2.2)**
		11	3	2	12.8	0.03	0.41	**(2.3)**
	Pictom-etry	9	2	1	6.8	-0.09	0.72	**(2.4)**

Comparing the corresponding setting in two cases using only images from east and from all directions (e.g. setting 1.1 and 2.1), results from all directions are better in both point density and accuracy. Increment in viewing directions leads to the increment of images involved, so that the number of matched points is higher. More significantly, it extends the base line for matching, which greatly improves the accuracy of the points.

The mean distances are all positive for SenP and negative for Pictometry. The absolute deviations are larger when the points cloud is generated from east than from all directions. Checking the mean distance of the point clouds from east in detail (Figure 2.9(a)), SenP point clouds has a positive system error around 0.1 m towards the shooting camera, especially the points on the roof. The error of Pictometry is 0.68 m away from the camera. The $\mathrm{d_{mean}}$ values from all directions (Figure 2.9(b)) are more evenly distributed around zero. The absolute deviations for SenP are all less than 0.5 m whilst the ones for Pictometry reached as large as 1 m.

The $\boldsymbol{\sigma_p}$ values decrease dramatically when the viewing direction increase from one to four. For example, $\boldsymbol{\sigma_p}$ for setting 1.1 is 0.76 m, but decreasing to 0.43

m for setting 1.4. Since the façades and roof planes tested in east are visible in the same number of images, their RMSE are mixed up in Figure 2.9(c). However, the RMSE values for roof planes are smaller than for façades in the case of all directions (Figure 2.9(d)), because the most roofs are visible in more images than façades. But when calculating the RMSE for a certain setting in Table 2.3, we do not distinguish roofs from façades.

After comparing between the point clouds from one and four directions, statistics within one case are examined. When the settings for *minImageNumber* and *level* are the same (such as setting 1.1 and 1.4), SenP dataset provides better results due to its higher overlap and resolution. However the computation time is quite high for SenP to be processed with *level* 1. With SenP data, the *level* 2 setting generates one third of the point density than *level* 1 rather than a quarter, because the rate of the success matching increased by using a larger sample size. In order to make the process within a tolerable time, we select *level* 2 for SenP.

The generated point cloud with *minImageNumber* equal to 3 is better than 2 with a sacrifice on point density. Therefore for SenP we take the point cloud with *minImageNumber* equals to 3 and add the point cloud with *minImageNumber* equals to 2 in the unmatched areas.

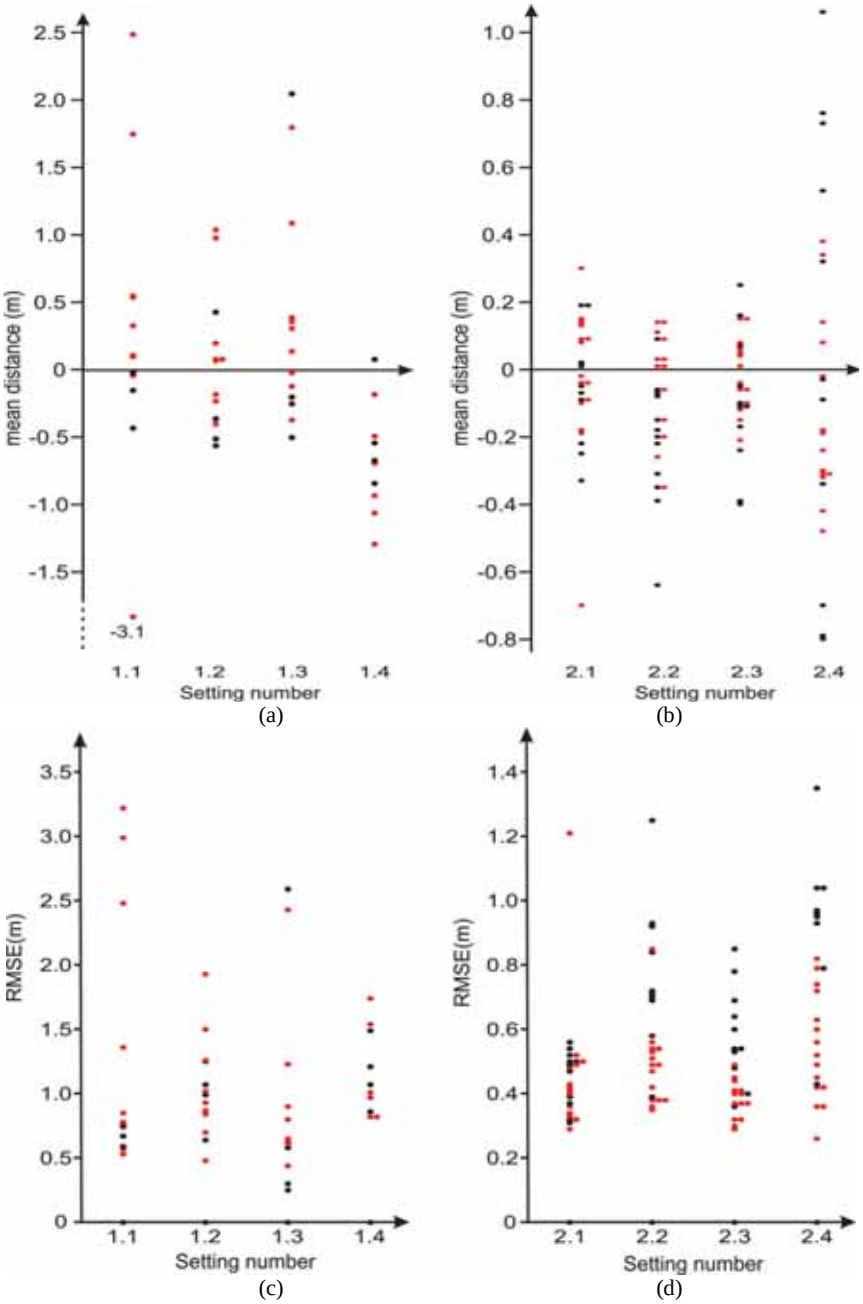


Figure 2.9 Mean distance and RMSE of point cloud to each visible plane for each test setting listed in Table 2.3. Red – roof plane, black – facade.

PART I: BUILDING FAÇADE DETECTION

Façades as dominant components of buildings bear several characteristics that distinguish buildings from other objects. They can be used as an indication of buildings, yet the usage of them was restricted because they cannot be clearly viewed in nadir images. Façades of buildings usually show relatively large regular structures, such as windows and doors, which result in characteristic image texture, and are normally vertical, whereas roofs can be horizontal, but also slanted within a certain range. For these reasons, façades will be initially detected in oblique airborne images for the purpose of building detection.

This part of the thesis aims at developing a method to detect building façades from oblique images. Texture features in images and stereo features from overlapped images are exploited. Experiments are carried out with Pictometry and SenP datasets. Based on the results, the impact from images resolutions and overlap rate are compared. In principle the method is an unsupervised approach, and therefore some parameters are set according to basic assumptions, or primary test over small sample sets. Sensitivities of crucial parameters are analyzed to test the robustness of the detection. Developed approaches are described in Chapter 3 while the results and analyses are presented in Chapter 4.

Chapter 3 Method and approaches

3.1 Major features used for detecting façades

In most research work, building detection from airborne images is based on the appearance of roofs in nadir images. However, clearly visible façades from oblique views contain more unique structures compared to roofs, such as windows, doors and balconies.

Considering the objects in the environment, selected features should enable a successful separation of building façades from natural objects such as trees, or other man-made object such as roads, or buildings parts like roofs. Given that consideration, the features are defined in both image space (2D) and object space (3D). Explanations are given with the help of Figure 3.1.

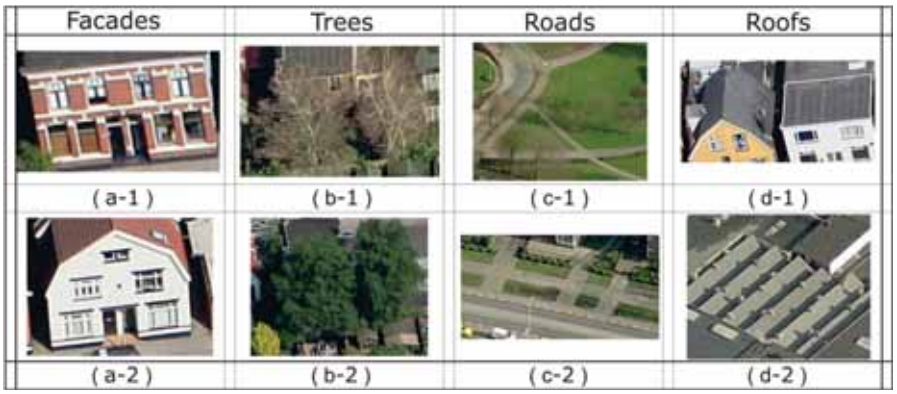


Figure 3.1 Comparison of building façades with trees, roads and roofs

Except for some rare cases of extraordinary architecture, edges of building structures such as balconies, windows and doors are usually aligned along horizontal and vertical directions, which appear as parallel lines in images (a-1 & a-2)‡. In comparison, the lines from the natural objects, mainly trees in urban area, usually have arbitrary directions (b-1 & b-2). Therefore one feature for distinguishing a façade would be parallel lines in two dominant directions in images.

Kurb lines as shown in (c-1) may sometimes be curved whereas as in (c-2) that they may still have two dominant directions. Given that the size of building structures are much smaller than the pattern formed by roads, the density of parallel lines in the images should be another feature.

‡ Strictly speaking the parallel lines from object space are not parallel in perspective images, but pointing to a common vanishing point. This is addressed later on.

However roofs as shown in (d-2) or other objects which also have dense parallel lines cannot be easily distinguished from façades by merely working with 2D features from a monocular image. Thus employing 3D features from multiple image views is helpful. In general, façades are in vertical planes, but other objects like roads or roofs (d-1) are in either horizontal or tilt planes. Hence height gradient is another helpful feature in distinguishing them.

To summarize the above mentioned features, we have 1) image textures on building façades should have dense lines in two dominant directions; and 2) height gradient on building façades should be larger than on other objects.

3.2 Workflow

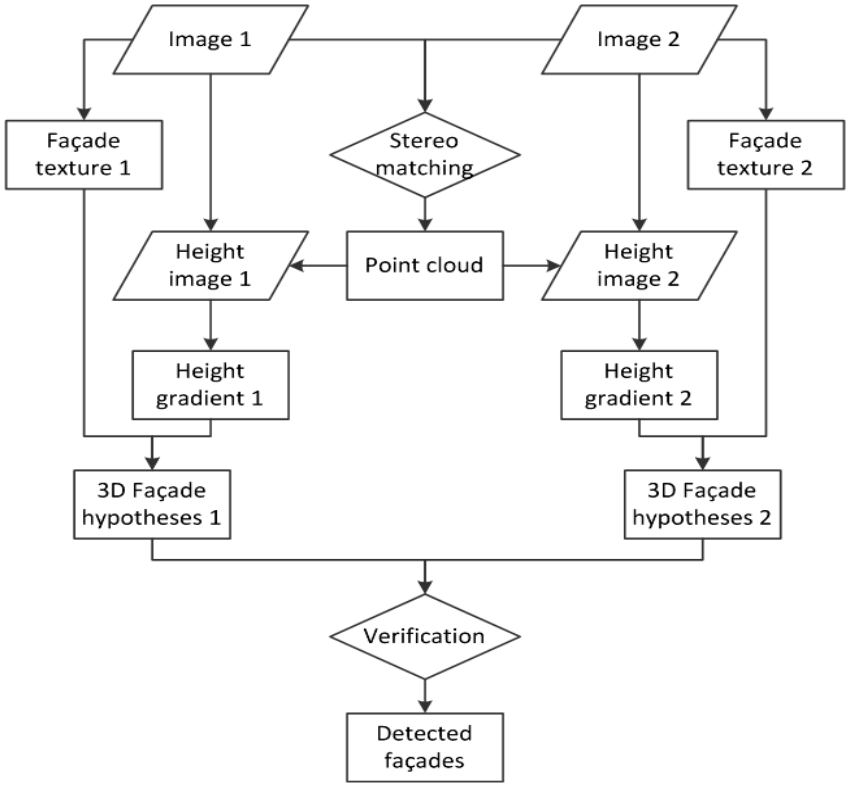


Figure 3.2 Flowchart for façade detection from a stereo pair

Evidences from façade texture and height gradient are calculated separately. Since the texture feature is already processed from images, the height feature is therefore converted into image space to facilitate their combination. The point cloud is first generated from image matching (cf.

section 2.5) and then projected onto images. Each image has its corresponding height image, in this way the height gradient can be calculated in the image space.

Façade detection is done in every image separately. In each image, the evidences from the original image and the projected height information are combined and façade hypotheses are generated. Since pixels implicitly bear height information, the hypotheses can be projected back into object space. Façade verification is then done by testing the 3D hypotheses from different images.

To simplify the façade detection process, the flowchart of it is drawn based on a stereo pair of images with similar perspectives (Figure 3.2). The minimum number of visible images is two to detect one façade. If there are more than two images available, the point cloud is generated from more images, and the verification is done by testing hypotheses from all of them.

The above approach addresses façade detection from one direction. To complete the detection of all visible façades, the façade detection process should be implemented from four directions separately.

3.3 Evidence from façade texture

Evidence from façade textures mainly relies on vertical and horizontal structures. For a certain image patch, the processes of looking for this evidence includes identifying the directions of vertical and horizontal in the image patch as well as testing their density.

3.3.1 Characterization of vertical and horizontal structures in image

The evidence from façade structures is retrieved from the assumption about vertical and horizontal structures presented on building façades. From projective geometry it is known that parallel image lines meet in a common point – the vanishing point. If an airborne image is taken under a significant angle with respect to the horizon, there is one vanishing point referring to the vertical lines available in the scene. Further there are multiple horizontal vanishing points. Given the exterior orientation of the images, one can compute the position of the vertical vanishing point in the image, as well as the plane defined by all horizontal vanishing points. For further details, refer e.g. to (McGlone and Shufelt, 1994).

The vertical vanishing point is mainly used here. Horizontality is exploited implicitly. To simplify the explanation of the plumb lines over the whole image, it is assumed that the aircraft was flying horizontally during image

acquisition and that the oblique images are significantly tilted in one direction only. Based on this simplification, plumb lines in object space in the vertical plane passing through the principle point are still vertical in the image space. The other lines deviate from the vertical.

The degree of the deviation can be defined through the calculation of the vertical vanishing point, thus the vertical lines in the image should be along that direction. Figure 3.3 shows the deviation angle over the images from both Pictometry and SenP, denoted deviation image. Lines through the points with the same deviation all intersect at the vanishing point. The maximum deviation angles are located at fore-ground corners of the images, which are about ±25 degrees in SenP images. Pictometry images were taken at twice the height of SenP data so that the angles are only within ±10 degrees.

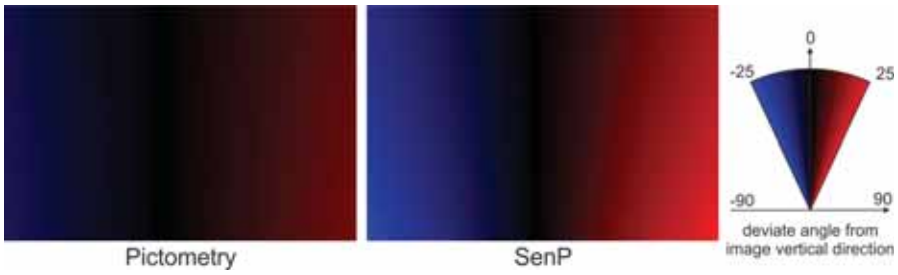


Figure 3.3 Examples of *deviation images* from Pictometry and SenP images.

The vanishing points are far out of the images for the reason that the tilt angles are about 45 to 50 degrees and the images were taken high above ground. Accordingly, parallel lines in object space are locally parallel in image space within the area of a building façade. Frames of windows and doors are assumed to be either vertical or horizontal, and thus their projections on the oblique images appear as about-vertical lines (corrected by the deviation angle) and lines of the same direction separately (Figure 3.4). Our hypothesis, consequently, is that a patch in the image which shows both about-vertical and parallel textures at a different direction indicates a façade candidate.

Figure 3.4 vertical and horizontal structures on building façade. Black – vertical edges, white – horizontal edges. Oblique images © Blom

3.3.2 Identifying vertical and horizontal structures

The process of looking for façade candidate patches is solely executed in one image. It consists of three steps:

I. extract line pixels and compute the direction for each of them;

II. on each pixel, a moving window is employed to test whether the patch inside the window could be part of a façade by checking the number of pixels aligned in vertical and horizontal directions;

III. accumulate the evidence of all moving windows to generate an evidence map.

Step I: Extracting line pixels. The line pixels are extracted by using the line extraction method proposed by Förstner (1994). This method is based on the average squared gradient of each pixel. Gradients in column and row directions on each pixel are firstly calculated from the original image, denoted g_c and g_r respectively. Then, the average squared gradient is described by the matrix:

$$\boldsymbol{N} = \boldsymbol{\nabla g \nabla g^T} = \begin{bmatrix} \sum \boldsymbol{g_c^2} & \sum \boldsymbol{g_c g_r} \\ \sum \boldsymbol{g_c g_r} & \sum \boldsymbol{g_r^2} \end{bmatrix} \qquad \textbf{(3-1)}$$

where $\nabla g = (g_c,\ g_r)$. Pixels on line features are located based on two eigenvalues λ_1, λ_2 of matrix N. There are two indicators derived:

1) If the sum of the two eigenvalues is large, there is large change of the grey value in the image. So that the trace of the matrix is used to select the pixels corresponding to image features:

$$\boldsymbol{tr(N) = \lambda_1 + \lambda_2} \qquad \textbf{(3-2)}$$

If $tr(N)$ is higher than a threshold T_{tr} then the pixel is considered as a feature pixel;

2) Line features need to be distinguished from corner pixels by checking the ratio between two eigenvalues λ_2/λ_1, which yields the degree of orientation or an isotropy. $\lambda_2 = 0$ indicates a straight edge while $\lambda_2 = \lambda_1$ indicates a corner. The roundness function is given by:

$$\boldsymbol{q = \frac{4 \times \det(N)}{tr^2(N)} = \frac{4\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)^2}} \qquad \textbf{(3-3)}$$

The value should be smaller than a threshold T_q for a line pixel.

The values of the two indicator thresholds are defined by some empirical value. We set consistent values over two datasets: $T_{tr} = 5000$ and $T_q = 0.3$.

Only pixels on line features are kept and augmented with the respective line direction they belong to. The resulting image is a *direction image*. Figure 3.5 shows part of the SenP image and its corresponding *direction image*. The pixels of line features are coloured by pixel direction. The other pixels are left blank.

(a) (b)

Figure 3.5 Direction image of part of a SenP image, edge pixels are coloured according to their directions (a) original image; (b) *direction image*

Step II: Façade patches testing. An *n*×*n* moving window is employed to count the façade evidence over the *direction image*. *n* depends on the image scale and the size of façade. The moving window size should be capable of holding a single façade structure like a door or a window, but still within the façade size in order to focus on the structures.

Recalling the discussion in section 3.1, if an image patch is on a building façade, the structure in the patch should show two dominant directions (plumb and horizontal), and the density of the structures in these dominant directions should reach a certain value. Therefore, two dominant directions the image patch in a moving window are calculated previously.

- Local **vertical direction** is retrieved from the *deviation image.* Vertical direction at the centre of the window is used as the vertical direction of the whole patch, because locally parallel image lines are assumed.

- Local **horizontal direction** is assumed to be the dominant direction of the edges pixels except the vertical ones. The dominant direction is found through binning directions apart from the vertical ones in the moving window into overlapping direction bins. Directions from the bin with the maximum count are averaged to obtain the final horizontal direction. In case no peak from the bins can be discovered, the window is treated as a non-façade patch.

Edge pixels from the direction map within certain degrees of vertical or horizontal direction are assigned as pixels of that direction respectively. Based on the total number of edge pixels (N_{total}), the numbers of vertical edge pixels ($N_{vertical}$) and the number of horizontal edge pixels ($N_{horizontal}$), three criteria are defined to test whether an image patch can be on the façade.

1) Since structures of windows and doors are commonly expected on the façades, a façade candidate window should have dense vertical and horizontal lines. In other words, the number of pixels of these directions over the window size should be higher than a threshold T_1:

$$\boldsymbol{N_{vertical}/(n \times n) > T_1} \qquad \textbf{(3-4)}$$

$$\boldsymbol{N_{horizontal}/(n \times n) > T_1} \qquad \textbf{(3-5)}$$

The above two formula equals to:

$$\boldsymbol{MIN(N_{vertical}, N_{horizontal})/(n \times n) > T_1} \qquad \textbf{(3-6)}$$

In some buildings, the contrast of façade structures is not very high, leading to missed edge pixels. Therefore the specified T_1 should not be

high to miss the façade. The number of false positives will be reduced by the other constraints.

2) To compensate for the relatively tolerant T_1 value, a minimum T_2 is defined for the total number of vertical and horizontal pixels:

$$(\boldsymbol{N_{vertical}} + \boldsymbol{N_{horizontal}})/(\boldsymbol{n} \times \boldsymbol{n}) > \boldsymbol{T_2} \quad \textbf{(3-7)}$$

3) Unlike objects such as trees with all pixels having arbitrary direction, pixels of vertical and horizontal should be the majority of detected edge pixels. So that the ratio between the summery of vertical and horizontal pixels and the total edge number should be larger than a threshold (T_3):

$$(\boldsymbol{N_{vertical}} + \boldsymbol{N_{horizontal}})/\boldsymbol{N_{total}} > \boldsymbol{T_3} \quad \textbf{(3-8)}$$

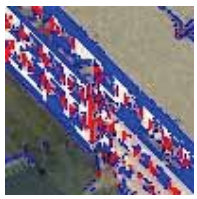
Original image			
Image with vertical and horizontal pixels (red – vertical; blue - horizontal)			
Window size *n*	100 × 100	100 × 100	100 × 100
$N_{vertical}$	653	735	187
$N_{horizontal}$	1692	730	278
N_{total}	3031	2406	2036

Figure 3.6 Detected vertical (red) and horizontal (blue) pixels over selected image patches. Images @ Blom.

The constants T_1, T_2 and T_3 are defined by a test over a small sample size. Details are presented in section 4.2.1. Figure 3.6 shows an example of the detected vertical and horizontal pixels overlaid on the original image patch. The window size is enlarged for visualization purposes, therefore the calculated values for equations (3-1) to (3-4) may not meet the specified thresholds. According to the example with buildings and trees, it is clear that the criteria are suitable for distinguishing buildings from trees: building façades present clear horizontal and vertical structures, but the set of vertical and horizontal edges detected for trees is quite sparse.

Step III: Façade evidence accumulation. An evidence counter is used on each pixel to record the judgement from the last step. If a window fulfils the conditions of equations (3-1) to (3-4), it will contribute an increment to the *evidence counters* of all the pixels in the window. The windows move at one pixel steps, such that one pixel will be visited $n \times n$ times by a window. After finishing the moving windows over the whole image, all pixels have a value from the evidence counter, resulted in a map of façade texture evidence, denoted as *Texture_Evidence_Map*. The value assigned to each pixel is correlated to the likelihood that the particular pixel is part of a façade patch. The *Texture_Evidence_Map* will then be combined with the evidence from height.

3.4 Evidence from height gradient

The second important evidence which supports the façade detection is related to the height gradient, in the form of a map with the vertical gradient as its pixel value, denoted as *Height_Gradient_Map*. Since height gradients at the façade plane are supposed to be larger than those on the ground or on the roof plane, we assume that values in the *Height_Gradient_Map* are (again) positively correlated with the likelihood that the respective image patch represents a façade.

Only images with approximately the same direction are selected in order to generate points on the planes facing the viewing direction. Images are normally taken pointing four cardinal directions (north, south, east, west), hence they are also clustered into four direction groups.

To achieve the integration with the texture evidence and to facilitate later steps on façade patches using image processing methods, heights of the 3D points are projected into the source images and interpolated to a *height image*. Since the points are projected into only the images that have been used for the dense matching, those parts of the objects that are subject to (self-) occlusion are left out. For this reason, no separate occlusion detection step is necessary. Figure 3.8(b) is the cropped point cloud of the area in Figure 3.8(a). It is generated from image viewing from similar directions with (a) so that the occluded regions in (a) are left blank in the point cloud (b).

It may happen that small holes and noise exist on the projected map. Therefore, we apply a moving average filter to interpolate it. Figure 3.8(c) is the interpolated *height image* of one of the matching source images. It presents the general height but still contains noise, which may be due to the unstable geometry for stereo matching from images with short baselines.

The height gradient (hg) is then calculated from the *height image*, sketched in (Figure 3.7). Gradients along image columns (g_y) and rows (g_x) are firstly derived. Then the gradient g and the angle δ between the real slope direction and image vertical can be calculated:

$$\boldsymbol{g = \sqrt{g_x^2 + g_y^2}} \qquad \textbf{(3-9)}$$

$$\boldsymbol{\delta = \arctan(g_x / g_y)} \qquad \textbf{(3-10)}$$

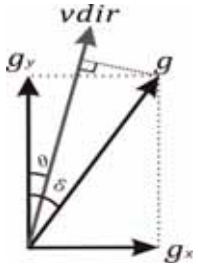


Figure 3.7 Calculation of height gradient alone vertical direction

As g does not always exactly align the vertical direction, so the real height gradient is the projection g onto the plumb line:

$$\boldsymbol{hg = g \,/\, cos(\delta - \theta)} \qquad \textbf{(3-11)}$$

where θ is the angle value from *deviation image* (cf. Figure 3.3).

Comparing Figure 3.8(d) with (a), the dark regions in (d) coincide with the façades and tree crowns in (a). But the gradient map is still too noisy to be solely used for façade detection and not feasible to distinguish between façades and trees. Thus it will be combined with the *Texture_Evidence_Map.*

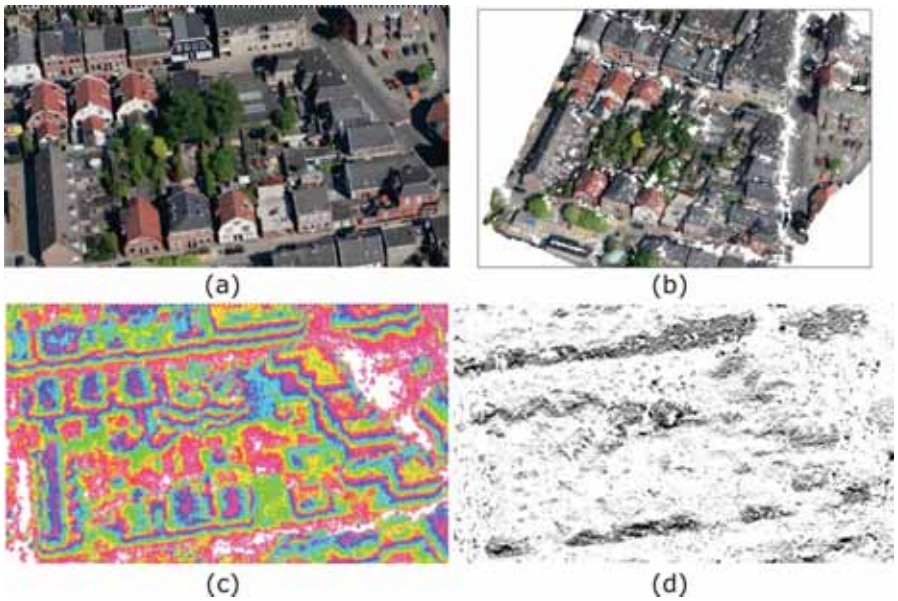


Figure 3.8 Height gradient evidence generation. (a) Crop of one original SenP image; (b) point cloud generated by PMVS from images of similar perspective with (a); (c) projected *height image* of the cropped area, colour present height levels; (d) *Height_Gradient_Map* of the cropped area, the darker the larger the gradient.

3.5 Façade hypotheses

Façades hypotheses are generated per image by combining pre-calculated evidence maps. With the height information from projected height maps, hypotheses can be reconstructed into object space.

3.5.1 Façade regions in 2D

As already mentioned in section 3.3 and 3.4, the pixel value of both evidence maps are positively correlated to the likelihood of the pixel being on façades. To avoid unifying the values of different units from two evidence maps, they are multiplied to generate the final façade possibility map *Façade_Possibility_Map*, in which each pixel value indicates its chance of being on the façade. that is:

$$Texture\_Evidence\_Map \times Height\_Gradient\_Map = Facade\_Possibility\_Map$$

Based on the range of each evidence map, a threshold T_{facade} is set to select pixels which probably belong to façade patches. Although we set this threshold empirically for the whole test site, we found out through experiments that the final result is quite insensitive to the selected value. An

example is given in Figure 3.9, but details are discussed in section 4.2.2. The value of *Façade_Possibility_Map* of the sample image varies from 1 to 8×10^5. Comparing (b) through to (g), the selected number of pixels decreased while T_{facade} increases. Except for (g), the others all contain almost the complete target façade, and are able to distinguish the façade from its surroundings, proofing that the façade pixel selection is not very sensitive to T_{facade}.

The selected roof and ground pixels adjacent to the façades are caused by the moving window. The pixels around the façade area may also obtain a high score of the texture evidence because the moving windows around it also contain part of the façade area. When the whole window is recognized as façade window, the texture evidence counter of the pixel will be incremented. However, the highest score still appears at façade pixels, so that when T_{facade} increases, the roof and ground pixels close to façades are generally not selected as façade pixels.

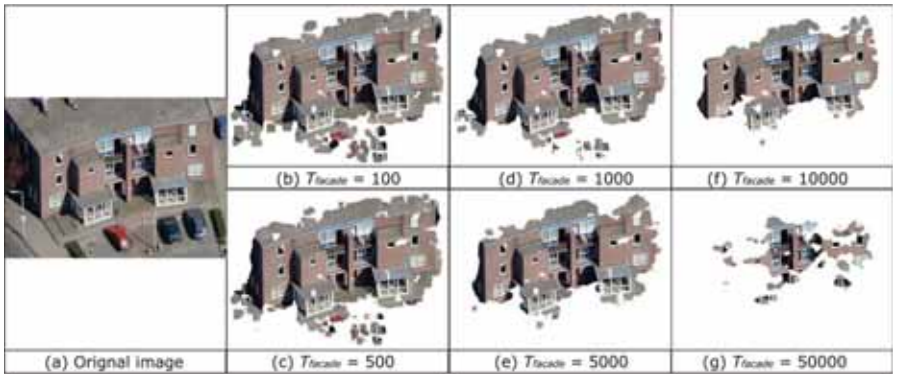


Figure 3.9 Selected façade pixels by a range of $T_{façade.}$ Example from SenP image.

Independent façade hypothesis regions are then extracted through a search for connected components within the selected façade pixels. Considering that some narrow connections between large components will be close to each other and that small fragments with few pixels will appear as well, we firstly implement morphological erosion to better separate close-by components and to remove small fragments. Afterwards, morphological dilation using the kernel of the same size with the kernel of erosion is applied to approximately recover the pixels around the main components which have been removed in the erosion step.

3.5.2 Façade region partition

Each refined connected component from detected façade pixels is considered as a façade region. A façade region may contain several vertical façades, for instance two vertical façades in the façade region in Figure 3.10(a). Consequently, three scenarios may occur: a) one region contains different façades of a building, b) curved façades, or c) planes of different depth (Figure 3.10(a-c)). Directly fitting a vertical plane to one of these scenarios will result in large errors. Hence the façades should be firstly separated.

When many buildings are quite close and aligned in the same direction, they could be detected within one region (Figure 3.10(d)). Although they may not be separated after the final detection, they should still be partitioned into smaller sub-regions to facilitate the verification process.

Three successive procedures are used to make reasonable partitions on the original façade region.

- **Region segmentation by horizontal direction**: taking the x, y coordinate from image and coding the horizontal direction (cf. section 3.3.2 step II) as grey value of the region pixels, the region pixels would be segmented by the horizontal direction.

- **Region partition by depth**: taking one segment from the previous step, a vertical plane parallel to the horizontal direction of that segment is used to calculate the depth value of each pixel. Then the pixels are clustered again by the depth value.

- **Partition using fixed size**: fixed maximum region height and width is lastly used to divide the too large regions that remain after the previous two steps. The size is not crucial because the partitions will later be connected again if they belong to the same façade, see Section 3.6.

Original façade regions are now partitioned into sub-regions by means of the following procedures. Figure 3.10(e-h) shows the partitions of the original façade regions of (a-d) respectively. Planes with different directions and parallel planes are successfully separated. In (h), partitions do not exactly fit the border between houses. It is because the houses are too close to each other and show no difference in depth: by merely looking at the façade textures, they could not be successfully separated. In those cases, however, the subsequent procedures will not be negatively influenced since geometrically all building façades are represented by one vertical plane.

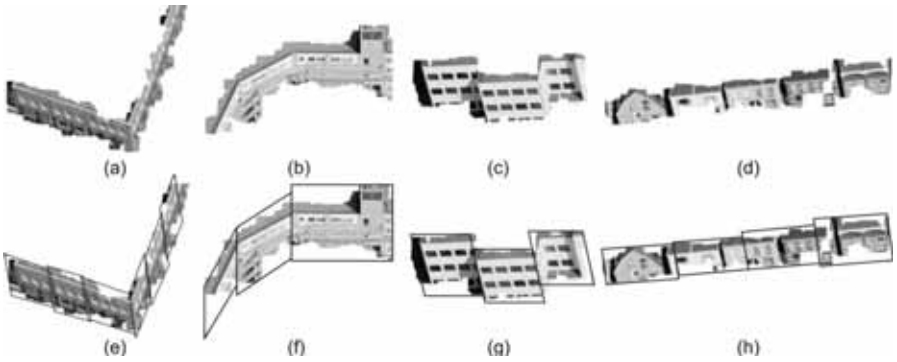


Figure 3.10 Façade partitions from original region

3.5.3 Façade hypotheses in 3D

3D façade hypotheses are built in the form of vertical rectangles. The key issue in this step is to define the vertical façade plane and then define the boundary of the rectangle.

The façade points from each façade region in one image can be generated by back projecting the image pixel with the height value into the object space. In the ideal case, all the points should fall into the same vertical plane, namely the real façade plane. However, due to the errors of image calibration, dense matching and interpolation of the *height map,* projections result as a volume of points. Hence we need to fit a plane to the points.

In order to reduce the error, the normal vector on the XOY plane (N_x, N_y) of the fitted façade is pre-defined for each partitioned sub-region by averaging the horizontal direction of each pixel. Then after fitting the plane, the rectangle bound of all the façade point projections on the plane is taken to form the façade rectangle.

3.6 Façade verification and connection

The generation of façade hypotheses is carried out simultaneously on each image in the image set. Therefore, the sets of hypotheses can be used to verify each other. If one façade hypothesis can be verified from at least one image other than its source image, it is marked as being verified.

The verification between two 3D façade hypotheses is implemented by applying the criteria of parallelism, overlap and proximity (Figure 3.11). Because façade hypotheses are vertical planes, their projections on the X-Y plane are taken to simplify the comparison.

- The parallelism of the two façades is checked by limiting the intersection angle δ to a threshold. We set quite strict 5 degrees given the relatively correct façade orientation in the façade fitting process.

- The length of the overlap of the two projections along the adjusted direction should not be less than a certain proportion of each façade, such as 75%.

- The proximity is judged by the distances from the middle point of each line to the other, both of which should be less than a certain length. The selection of the threshold values generally depends on the image resolution, as well as the uncertainty from the dense matching.

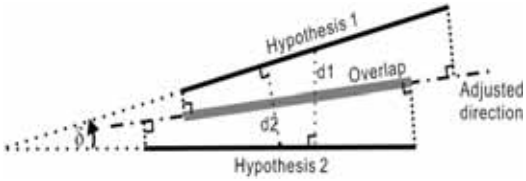


Figure 3.11. Façade hypothesis verification

Façade hypotheses from different images that confirm each other are grouped to generate one refined façade. The orientations of the hypotheses in a group are averaged by using the façade length as weight to calculate the orientation of the new façade. The maximum length of all hypotheses is used for the final façade, resulting in a façade that is possibly longer than the actual one. This is useful in the connection step.

The aim of the façade connection is to join the façades which have been fragmented during the façade region detection and partitioning. Figure 3.12 shows the results from the façade connection of the samples in Figure 3.10. Partitions of buildings (b) and (c) have different horizontal directions or depth, thus the connection process was not applied. Partitions in (a) and (d) are connected. Together with the previous refinement, the façade plane can be assumed to be quite accurate.

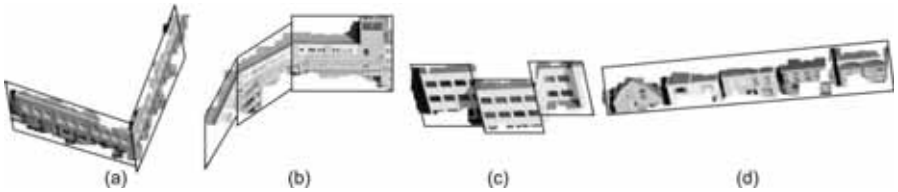


Figure 3.12 Result of façade connection

3.7 Summary

In this chapter we proposed an automatic façade detection approach exploiting the façade texture from single images and height gradients derived from multiple images. Multiple overlapping images viewing to the same direction are required to detect façades facing that direction. Therefore, images viewing from all directions (normally four cardinal directions) are required to detect all façades.

The output of this approach is a set of vertical rectangles. The 2D projections of these rectangles represent the location of detected building façades, while the height of the rectangle shows the maximum height of the detected façade pixels.

Chapter 4 Results and evaluation

4.1 Experimental design

A study area is chosen to test the proposed approach for façade detection on different building types. Images are from both the Pictometry and SenP datasets in order to check the impacts of resolution and overlap on the results. Accuracy of the detected façades is assessed using reference data.

4.1.1 Study area

The study area is located south to the centre of Enschede, the Netherlands, covering about 1 km2 (Figure 4.1). The study area is selected to cover various building types in order to test the versatility of the developed methods. Buildings inside the area include gable roof houses, regular and irregular roof buildings, squat factory buildings, round towers, etc. Many buildings are located close to each other, less than 2 m distance. In this sense, it is a quite challenging scene, different from many researches that only deal with limited number of building types.

Figure 4.1 Study area. Image © Microsoft Corporation © 2010 Terraltaly

4.1.2 Image data

The study area is covered by both the SenP and Pictometry datasets. However in some cases it is only partially covered by images from a specified direction. Detailed coverage reference to the boundary of the study area is presented in (Figure 4.2). In total 34 images from Pictometry (6 from north, 10 from south, 8 from east and 10 from west) and 157 image from SenP (34

from north, 38 from south, 54 from east and 31 from west) are used in the experiment.

Although the study area is completely covered by Pictometry images in all directions, the point cloud as derived from image matching will be incomplete due to its low overlapping ratio. On the other hand, most of the regions covered by SenP images are suitable for façade detection, but only the East viewing images almost fully cover the study area.

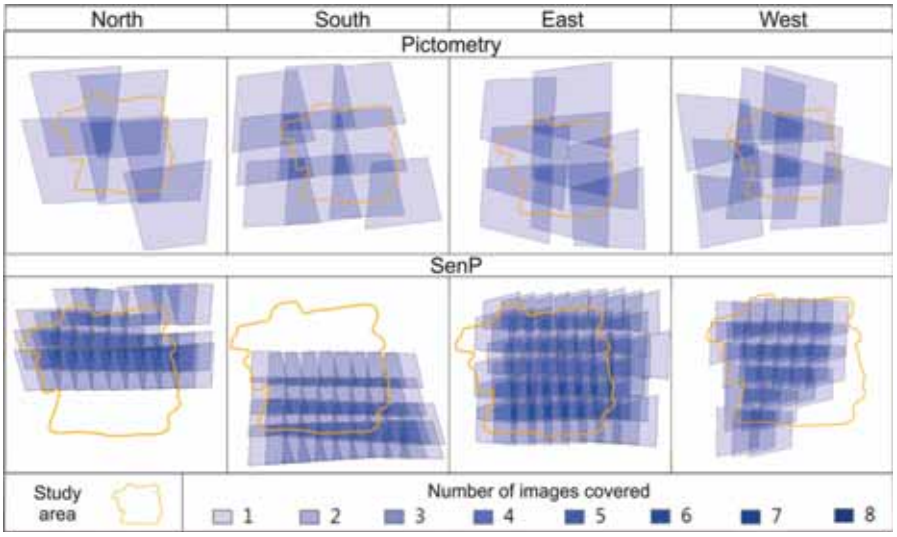


Figure 4.2 Image coverage from Pictometry and SenP datasets

4.1.3 Reference data

As the façades are assumed to be vertical, it is their 2D projections that really matter in terms of accuracy assessment. Therefore the map "GBKN (Grootschalige BasisKaart Nederland = Large Scale Base Map Netherlands)" (GBKN, 2012) is used for the result assessment. The map has a nominal accuracy of 10-20 cm. The 2D map of the study area is shown in Figure 4.3.

The reference data shows the footprints of buildings, more specifically only the outer building facades. For example, there is only one big block at the bottom-left corner of the study area. As visible in Figure 4.1, it contains many facade structures inside the polygon on various height levels. In that case there are many façades inside the polygon that will be detected by the algorithm, but not assessed. Basically this means we are interested in facades which are intersecting with the ground since these are the components of the 2D building map outline.

Figure 4.3 Reference cadastral map of study area and the boundary

Although the façades are generated in 3D, their heights are not yet checked. The reasons for this are 1) The final target is to draw 2D outlines of buildings; 2) A whole façade may consist of many nearby aligned buildings, and the connection step takes the maximum height for the new façades, leading to a potentially larger height of detected façades.

4.1.4 Assessment design

Detected facades are assessed through their 2D projections, namely one line segment for one facade. The line segments are compared with the polygons from the reference map. The assessment consists of detection quality and geometric accuracy.

Detection quality is assessed through detection completeness and correctness, both of which are measured in the whole study area level and object level.

Before the description of assessment measurements, some terms need to be clarified first. Taking one façade from one set (façade A), the facades in another set (façade B) close to (<2m) it are compared to it. The methodology for the comparison is the same as for the façade verification described in section 3.6, but with 10°, 2 m and 50% as the thresholds for angle difference, distance and overlap respectively. If the differences between façade B to façade A are below the thresholds, then they are matched. It is possible that façade A can find multiple matched façades.

Completeness measures how much of the reference is detected. Taking one reference building as a whole, its façades are the checking units (Figure 4.4). One building is marked as detected when at least for one façade of it we can find a matching detected façade. Then *building_completeness* presents which percentage of the buildings is detected. However, this measure does not distinguish between buildings with only one detected façade and those with all façades detected. *façade_completeness* is therefore used to assess this difference, checking the ratio between the matched reference façades and all detected façades in one building. The ratio is calculated by both façade length and façade number.

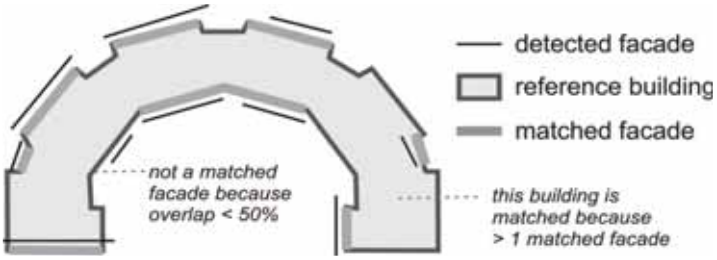


Figure 4.4 Sketch of completeness with a detected building, showing the matched façades and their corresponding detected façades.

Correctness is to express how well the detected results fit with the reference. A detected façade is considered correctly detected if one reference façade matches it; *overall_correctness* detects the ratio between correctly detected façades and the total number of extracted facades. Referring to Figure 4.5, for one correctly detected façade, we can project all its matched reference façades onto it, thus the detected façades are separated as *overlaps* and *unmatched parts*. The *overall_correctness* is unable to present the detection quality of every detected façade. For this, the *single_correctness* is employed. It computes the portion of the overlaps of the whole façade.

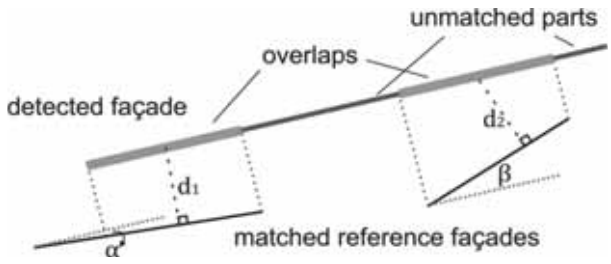


Figure 4.5 Matching from reference façades to one detected façade.

The ***geometric accuracy*** measures the orientation difference and distance from the detected façades to their references. As it is more common that one detected façade matches more than one reference façades, the geometric accuracy is tested based on detected façades.
The *orient_accuracy* measures the angle difference between a detected façade with its matched reference(s) in 2D (α and β in Figure 4.5) while the *distance_accuracy* measures the distance from the middle point of the intersection to the reference (d_1 and d_2 in Figure 4.5). If there are multiple matches, we take the weighted average with length as weight.

All assessing measurements are summarised in Table 4.1, including their meaning as well as their functions.

Table 4.1 Assessment measurements and meaning

Completeness: how much of the reference is correctly detected.	
building_ completeness *(%)*	The ratio between the number of matched buildings ($N_{building\_match}$) and the total number of reference buildings ($N_{building\_reference}$). $building\_completeness = \frac{N_{building\_match}}{N_{building\_reference}} \times 100$ **(4-2)**
facade_ completeness *(%)*	The ratio between matched façades and all façades in a reference building. It is calculated by length l and façade count N, as defined in equation (4-6) and (4-7). Denote matched reference façades as *reference_match*: $facade\_completeness\_l = \frac{\sum l_{reference\_match}}{l_{building\_outline}} \times 100$ **(4-3)** $facade\_completeness\_n = \frac{N_{reference\_match}}{N_{building\_facades}} \times 100$ **(4-4)**
Correctness: the difference from the detected façade to the reference.	
overall_ correctness *(%)*	Ratio between the number of detected façades that matched with real façades (N_{detect_match}) and the total number of detected façades (N_{detect_total}) $overall\_correctness = \frac{N_{detect\_match}}{N_{detect\_total}} \times 100$ **(4-5)**
single_ correctness *(%)*	Consider one of the detected façades for which match(es) in the reference are found. Then calculate the ratio between total length of the overlap(s) ($l_{overlap}$) and the length of the detected façade (l_{detect_facade}). $single\_correctness = \frac{\sum l_{overlap}}{l_{detect\_facade}} \times 100$ **(4-6)**
Geometric accuracy: compare the matched pair of detected façades and reference in orientation and distance.	
orient_ accuracy *(degree)*	Measure angle difference ($angle\_dif_i$) between a detected façade and its matched reference(s) in 2D, taking the length of corresponding overlap ($l_{overlap\_i}$) as weight. $orient\_accuracy = \frac{\sum(angle\_dif_i \times l_{overlap\_i})}{\sum l_{overlap\_i}}$ **(4-7)**
distance_ accuracy *(m)*	Measure the distance from the middle point of the intersection to the reference (dis_dif_i), taking the length of corresponding overlap ($l_{overlap\_i}$) as weight. $distance\_accuracy = \frac{\sum(dis\_dif_i \times l_{overlap\_i})}{\sum l_{overlap\_i}}$ **(4-8)**

4.2 Parameters setting and sensitivity analysis

As mentioned at the beginning of this part, the developed method is an unsupervised approach. Therefore, we set the parameters used during the process empirically or by the nature of the dataset. The values we take for all the parameters are given in the Appendix I. The settings of some key parameters are explained or discussed using a small training set.

4.2.1 Parameters for façade patch testing

During the façade patch testing process in section 3.3.2, step II, some parameters are set to decide whether the tested window can be a façade patch or not. They are:

- $\boldsymbol{n}$ for window size;
- $\boldsymbol{T_1}$ for $\frac{MIN(N_{vertical},\ N_{horizontal})}{n \times n}$: the ratio between the minimum number of pixels on vertical or horizontal edges and the number of pixels in the window;
- $\boldsymbol{T_2}$ for $\frac{(N_{vertical} + N_{horizontal})}{n \times n}$: the ratio between the sum of the numbers of pixels on horizontal or vertical edges and the number of pixels in the window;
- $\boldsymbol{T_3}$ for $\frac{(N_{vertical} + N_{horizontal})}{N_{total}}$: the ratio between the sum of pixels on horizontal or vertical edges and the number of edge pixels.

Window size *n*:
One floor of a standard building is around 3 m in height, and may consist of structures such as windows, doors or balconies. The window size is intended to be similar to the height of a floor. Taking Pictometry for example, given that the image resolution is 10-16 cm, n should be 20-30. We take 30 to make sure to include one entire floor. Similarly for SenP, the window size should be 40-60 (selected 50).

Thresholds T_1, T_2, T_3:
In order to define the value for the thresholds and the robustness of the results from the thresholds, a small sample set is selected from four categories: building façade, tree, building roof and ground. As the thresholds are all ratios, which are independent from the actual window size *n*, only samples from the SenP images are selected.

The scenes containing the selected samples are shown in Figure 4.6. f_i, t_i, g_i and r_i mean that the sample is part of the building façade, the tree, the ground and the roof, respectively. The cutouts are larger than the samples in order to give clearer ideas on the context of the samples.

For each sample, $N_{horizontal}$, $N_{vertical}$, N_{total} are collected for all the moving windows within the sample. Then for each window, values that relate to three thresholds are derived and plotted (lower part of Figure 4.6). The plot shows that the points related to façade windows are clearly distinguishable from the other categories. Therefore, for each threshold, we take the value which selects 95% façade points to obtain all façade windows but still omit the noise, namely 0.04, 0.12 and 0.5 for $\boldsymbol{T_1, T_2, T_3}$ respectively.

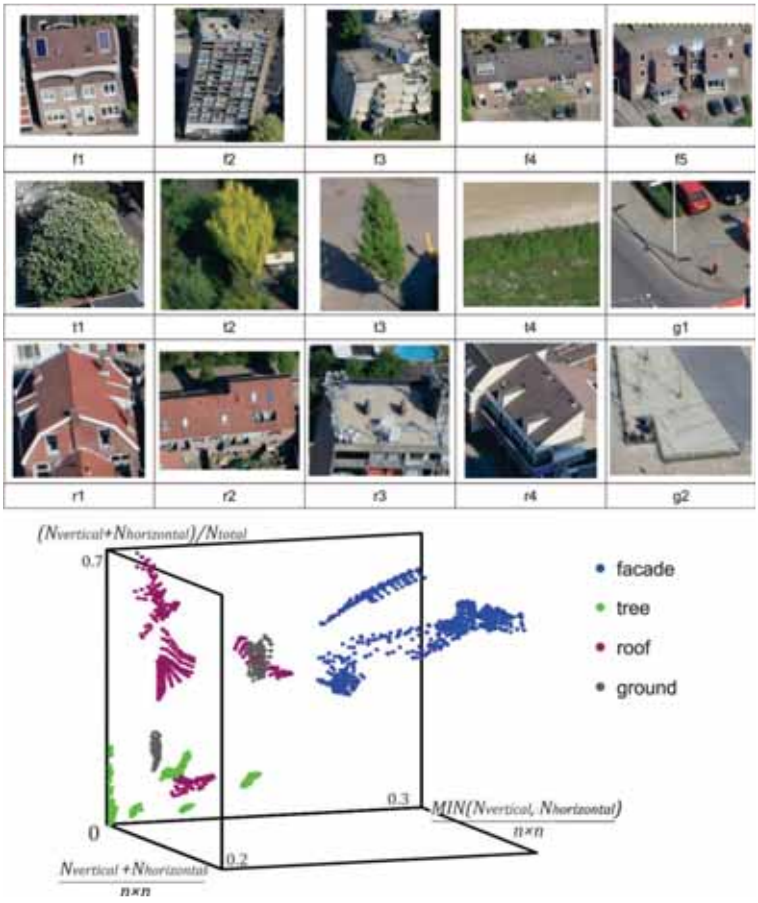


Figure 4.6 Samples and statistics to define thresholds for facade patch testing. Images are cut from one for SenP image.

4.2.2 Threshold to select façade pixels in images

As described in section 3.5.1 the 2D façade pixels are selected by setting a threshold T_{facade} on *Façade_Possibility_Map.* It is also briefly discussed with Figure 3.9 that the result is not very sensitive to this threshold, so we empirically set the value of 300 for Pictometry and 1000 for SenP.

The sensitivity of this threshold is further examined for both the Pictometry and SenP data sets. We start by fixing T_{facade} as the values mentioned above (T_{facade_basic}), resulting in a number of selected façade pixels (N_{facade_basic}) per image. The threshold T_{facade} is then varied by an exponential change (2^x) to shorten the experiment and to better show the effect of different threshold values. If x is in [-3, 5], then:

$$\mathbf{T_{facade} = T_{facade\_basic} \times 2^x} \qquad \textbf{(4-9)}$$

For each T_{facade}, a new number of façade pixels N_{facade_x} is computed and compared to N_{facade_basic}, resulting as a comparison function:

$$\boldsymbol{f(x) = \frac{N_{facade\_x} - N_{facade\_basic}}{N_{facade\_x}}} \qquad \textbf{(4-10)}$$

N_{facade_x} is used as the denominator rather than N_{facade_basic} due to the fact that as the number of pixels drops with the increase of the threshold, the impact from the decreased pixels on façade pixels becomes more dramatic.

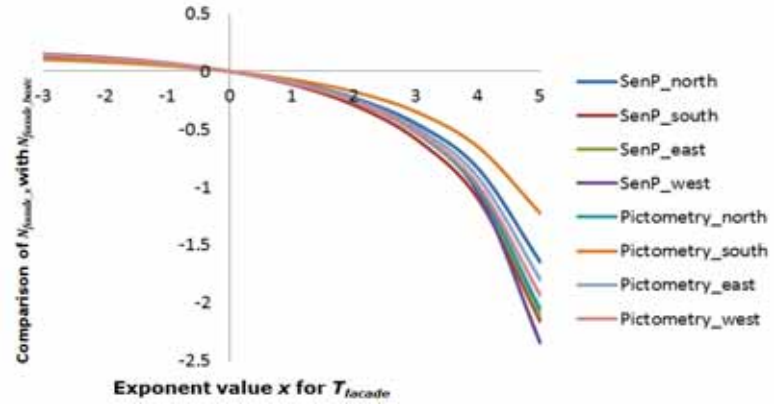


Figure 4.7 Percentage change of the facade pixel number with the change of thresholds.

In total eight images are tested (one from each direction per dataset). The results are charted in Figure 4.7. Almost the same trend is shown for all tested images. Moreover, it is only when the power of two exceeds the range of [-3, 3] that the change becomes significant. On this account, we conclude that the final number of façade pixels is not sensitive at all to the change of

T_{facade}, which also coincides with the result shown in Figure 3.9. Therefore we keep the previous settings.

4.3 Results and evaluation

Following the proposed method, facade hypotheses set are generated from each single image. In each cardinal direction, the hypotheses sets verify each other.

The *detection count* of a façade is defined as the number of images in which the façade has been detected. The detection quality and geometric quality are compared for various values of *detection count*. In this way, the impact of image resolution and overlap are also evaluated.

All the results are firstly visualised in section 4.3.1. Figure 4.8 shows the *detection count* of each detected façade, Figure 4.9 shows the *building_completeness* and Figure 4.10 shows the *façade_completeness*. The detailed analysis on the results is discussed in following sections.

4.3.1 Results visualisation

Figure 4.8 *Detection count* of detected façades from SenP (up) and Pictometry (down), coloured by count value


Not detected
0 - 10
10 - 20
20 - 30
30 - 40
40 - 50
50 - 60
60 - 70
70 - 80
80 - 90
90 - 100

Figure 4.9 *completeness* from SenP (up) and Pictometry (down), coloured by *façade_completeness* computed by façade count.


False detection
50- 60
60 - 70
70 - 80
80- 90
90 - 100

Figure 4.10 *Single_correctness* from SenP (up) and Pictometry (down). Grey polygons are from the reference.

4.3.2 Detection count

The *detection count* is visualised for each detected façade (Figure 4.8). It is obvious that the *detection counts* of the facades from SenP are commonly higher than from Pictometry. The maximum *detection count* is 10 from in the SenP data whilst it only less than 5 in the Pictometry data.

The *detection count* is mostly determined by the number of images, in another words by the overlap of image sets. There can be more available images at one spot when the overlap is high, thus the *detection count* is probably larger as well. This can be shown by the observation that the *detection counts* from SenP are much higher than from Pictometry, which have an overlap of 60% and less than 20% respectively. Figure 4.11 shows the exact number of detected façades corresponding to each value of *detection count*. Naturally fewer images are available at the border of the study area. Thus, in the centre part more façades with higher *detection count* are obtained, refer to Figure 4.8.

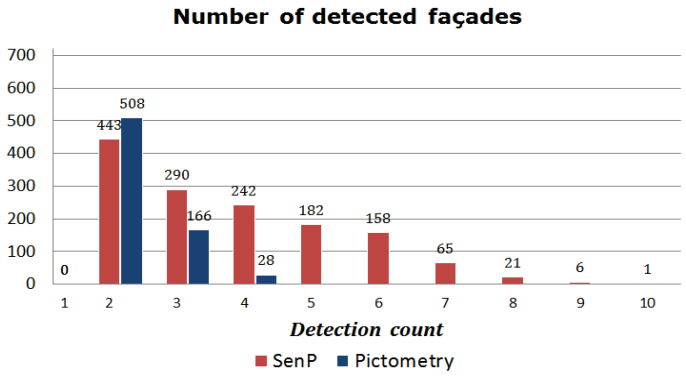


Figure 4.11 Number of detected façades over *detection count*.

4.3.3 Results on completeness

Completeness is evaluated over the entire level of whole study and on the basis of single buildings. Figure 4.9 presented the undetected buildings in blue and the detected buildings are coloured by completeness based on the façade count (Function 4-4). Detailed statistics are shown in Figure 4.12.

In total more than 90% of the buildings are detected from SenP images, leaving only 26 out of a total of 335 buildings undetected. The detection rate is about 70% for the Pictometry images.

The undetected buildings in SenP are distributed evenly. However, a group of undetected buildings in the Pictometry dataset is located in the east of the study area. This area is hardly covered by more than one image, which may lead to undetected façades and as a consequence, to missed buildings.

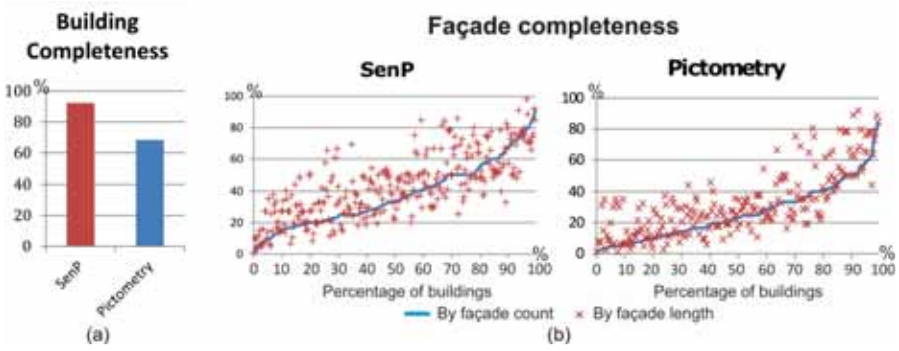


Figure 4.12 Results of (a) *building_completeness*; (b) *façade_completenss*. In (b), Buildings are sorted by the *façade_completenss* computed by façade count, shown in blue line. For each building, its *façade_completenss* computed by façade length is presented by the cross on the same vertical line.

Building colours in Figure 4.9 are commonly in orange, indicating that the façade completeness values are only around 40% to 50%. The colour in Pictometry is even lighter than SenP. This is also exhibited by the solid lines in Figure 4.12. Taking 50% completeness as a threshold, 70% and 90% of the buildings in SenP and Pictometry respectively are below this threshold. In other words, for most of the buildings less than half of their façades are detected.

In both data sets, crosses presenting the completeness by length reveal a general distribution higher than completeness by façade count. The probable reason is that the long façades have the same weight in the computing the completeness by count, but a higher weight in the completeness by length. However, we are more interested in knowing the completeness by count, since in the consequential step one façade regardless of length takes equal efforts to recover

Possible reasons for missing the facades are:

- Lack of images viewing in a particular direction: for example almost one third of the study are is not covered by SenP images viewing either north or south;
- Occlusion: the side facades of many buildings are usually occluded by the buildings close by its side. Furthermore, the back façade may be partially occluded by irregular structures, e.g. shrubs, sheds or fences
- Lack of texture or low contrast due to shading: there are some cases that few or no window or door on the sides of some house, then this algorithm could not recognize those facades.

4.3.4 Results on correctness

Figure 4.10 shows the detected façades overlaid on the reference building polygons. Façades for which no match in the reference was found are shown in blue. The *single_correctness* of each matched façade, ranging from 50% to 100% (<50% is recognized as unmatched), is shown from light orange to red.

In general, the *single_correctness* values of the detected façades are lower in the Pictometry than in the SenP image set. More small façades are detected from the latter image set.

Most of the detected façades are correctly detected. It is noticeable that the façades of lower correctness are usually along the building sides of "teeth-like" shape, for instance the buildings at the northeast corner of the study area. Correctness of the façades detected along these sides is lowered by the façades on the "teeth".

Note that the majority of facades of the semi-circular building close to the centre of the study area, are detected in both the Pictometry and the SenP data, proving that the developed method is capable of dealing with curved facades by fitting small line segments. Meanwhile, the parallel facades of the two buildings to the right of the semi-circular building are successfully located. These are all due to the strategy of facade partitioning.

For both the Pictometry and the SenP datasets, it is not feasible to completely separate the facades of nearby buildings. When the gap between a buildings aligning in a row is small (<3m), their facades are usually detected as a whole.

In general, four types of false detected façades can be discovered in the figure:

I. Wrongly detected façades: where there is no façade at all, but other objects like tall fences or sheds (Figure 4.13(a)). The number of these errors is very small. In total they are less than 5% of the total façade number.
II. Detection of a sloped façade: where the building façade is not vertical. The location is in the middle of the sloped plane, not matching to the intersection of the real façade with the ground (Figure 4.13(b)).
III. Detection of façades of inner structures. There are structures on top of a building block, but their façades are not included in the boundary of the cadastral map (Figure 4.13(c)).
IV. Not so accurate detection: the difference between the detection and the real façades exceed the matching thresholds (Figure 4.13(d)).

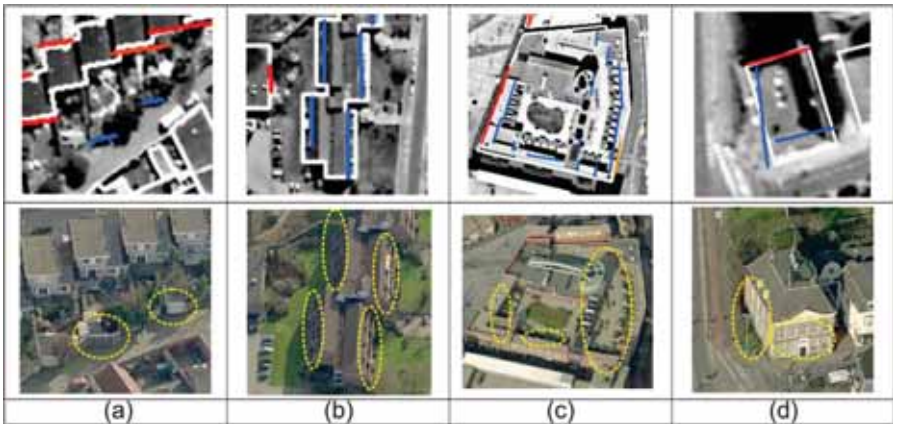


Figure 4.13 Examples for unmatched façades. First row: 2D data overlaid on mosaic image: Blue line segments – unmatched façades; reddish line segments – matched façades; white polygons – reference outline. Second row: part from images © Blom: yellow circles, place where unmatched façades occur.

Table 4.2 Percentage of false detections types (%).

Data set	SenP				Pictometry			
False Detection type	I	II	III	IV	I	II	III	IV
Percentage in false detections	18.5	1.1	54.5	25.9	16.8	4.3	34.1	44.7
Percentage in all detections	4.6	0.3	13.4	6.4	5.0	1.3	10.1	13.2

Percentages of all false detection types are calculated in Table 4.2. Detection of façades of inner structures (type III) contributes to a large portion of the false detection. Strictly speaking, they are not wrongly detected façades, but counted as being wrong due to the absence of inner structures in the reference data. Another main false detection type is the inaccurate detection (type IV), which is more serious in the Pictometry than the SenP image set. Pictometry has worse image resolution and less overlap, so that the geometric accuracy of the detected façades from it is less than from SenP. Details are discussed in section 4.3.5. Façades are assumed to be vertical in this method, thus sloped façades cannot be detected correctly (type II). Sloped façades only exist in three buildings, so the portion of type II false detection is quite low. Detection of fences or shed façades is also one type of false detection (type I), about 16% to 18% of the false detection in SenP and Pictometry respectively.

If merely seen from the perspective of façade detection, only type II and type IV can be considered as error because they result from poor geometric accuracy, so the correctness in terms of faces are 93.3% for SenP and 85.5% for Pictometry. However, if building detection is purely based on the detection of façades, false detections of type I will result in false building detections, whilst other types can still be used as indication of the location of buildings. As the thesis aims at detecting buildings, the correctness of the façade detection can be considered as about 95% for both datasets (100% - type I error).

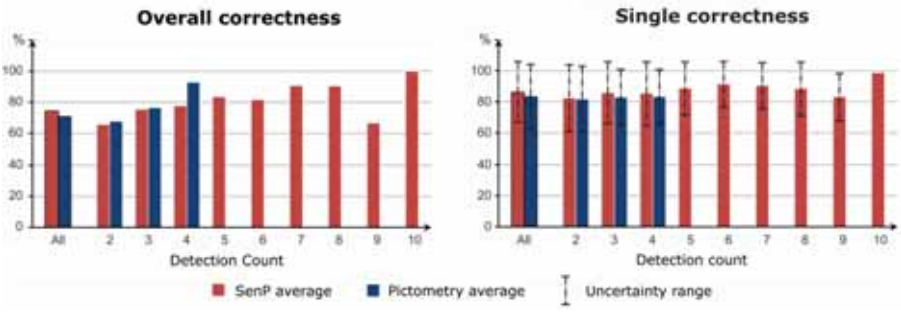


Figure 4.14 Results of correctness

The following statistical analysis is based on the grouping of the detected façades according to their *detection count*. Then the *overall_correctness* and the averaged *single_correctness* are computed for each group as well as for the group containing all the facades (Figure 4.14). In order to present the variance of the values within a group, the RMSE values are also computed. The RMSE is presented by uncertainty range in the figure, which is $\pm$ RMSE value from the average. The last group within *detection count* equals to ten only contains one object: it was excluded from analysis.

The *overall_correctness* defined by equation 4-5 is shown in the left bar chart in Figure 4.14. In general the *overall_correctness* is about 70% for both data sets in spite of different resolution and overlap of image sets. Based on the discussion on four types of false detection above, the first three types cannot be eliminated by better image resolution and overlap.

Comparing groups of different *detection count*, a weak trend of increasing of *overall_correctness* with the increment of the number for both datasets can be observed. This observation can be explained by the fact that for an increasing number of *detection count* (i.e. involved images), the chances for the false detections to remain are low. The values are higher for Pictometry than SenP in the groups with low *detection count* (2-4). It cannot be simply concluded that Pictometry generate better results than SenP when the *detection count* is the same, since the façade set of the same *detection count* from both data sets are not the same. For instance, the façades viewed by four images in Pictometry can be viewed by more images in SenP.

The *single_correctness* is quite consistent over groups and datasets. It reaches as high as around 85% for all and the RMSE in each group does not vary a lot. As stated, the "teeth" shape buildings have an impact on the *single_correctness*, which mainly based on the reference rather than image sources.

Comparing each corresponding group between SenP and Pictometry, the *single_correctness* is lower in the latter image set. This is cause by the incorrect positions of the façade ends. The end can be improved by the better resolution from SenP, which resulted in a slightly higher correctness value. Under the assumption that the extension of the facade is decided by the 3D facade points from image facade pixels, it gives a hint that the outlining of the facade regions in 2D is done reliably.

4.3.5 Results on geometric accuracy

The geometric accuracy in terms of angle and distance differences arealso assessed in groups by *detection count* (Figure 4.15). Façades detected from SenP have a much higher accuracy than the ones from Pictometry. Similar patterns of changes on averaged and RMSE values with *detection count* are presented in the charts for them.

The angle difference from all façades from SenP is about 1.5 degrees with the RMSE of 1 degree, whilst the difference increases to 2.5 degrees with doubled RMSE from Pictometry. The distance difference is also about 0.2 m worse from Pictometry than SenP. These accuracies are influenced together by image resolution and overlap. Better resolution increase the location of image features and more hypotheses from more images also enhances the positional accuracy after the adjustment.

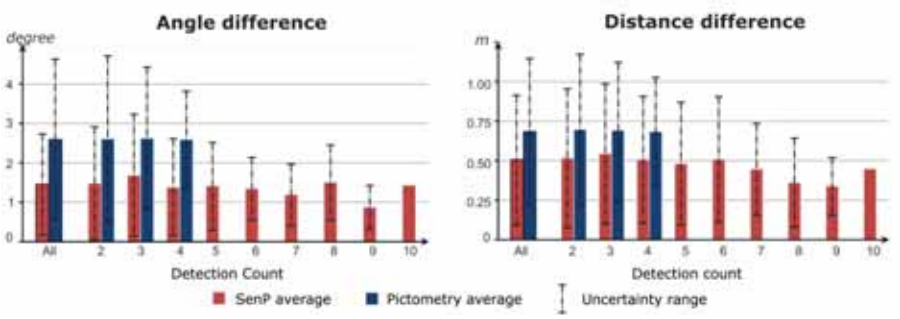


Figure 4.15 Results on *geometric accuracy*.

The *detection count* has positive impacts on the geometric accuracy, evidenced by the reduced value of differences and residuals. Taking SenP image set for example (the same trend in Pictometry), when *detection count* changed from 2 to 9, the averaged angle and distance differences decrease by about 50% and 40% respectively. The RMSE of the angle difference also reduces from 1.5 degrees to less than 0.5 degree. For the distance. it reduces from 0.4m to less than 0.2 m.

4.4 Discussion

The previous section presents the façade detection results by implementing the algorithm to the same study area using two datasets. In this section, the impacts of resolution and overlap on the result will be discussed.

4.4.1 Impact of resolution

The impact of resolution is induced by comparing the corresponding results from detection quality and geometric accuracy.

Positively, higher image resolution helps in the detection of building façades. Although the correctness is not significantly improved by higher resolution, the detection completeness is better based on the total number of detected façades. The correctness has not much to do with that, because the false detection of the façades of inner structures, the faces of fences and sheds as well as incorrect detection of slope façades (Figure 4.13 and Table 4.2) cannot be avoided by increasing the image resolution. The correctness is indirectly improved by better geometric accuracy.

Recalling that the overall angle difference improved from 2.5 degrees to 1.5 degrees while the distances difference reduced from 0.7 m to 0.5 m from Pictometry to SenP, it can be concluded that the geometric accuracy of the correctly detected façades are significantly increased by the better resolution. In particular this conclusion can be proven by the bars with *detection count* from two to four, i.e. the accuracy increased independently from the number of images.

From the test data, it appears that the higher resolution does not help much to improve the already high correctness (more than 85%) of a single detected façade. The errors are mainly due to the generalization of the tiny structures on the façades, which is determined by the defined level of generalization rather than by the image resolution.

4.4.2 Impact of overlap

Considering that the *detection count* highly depends on the number of images in which the target is visible, high overlap is one of the essential factors for multiple verifications. Thus the difference between the results of different verification groups of one specific dataset implicitly reflects the impact of overlap on façade detection.

Firstly, high overlap contributes to the overall correctness. False detections could be successfully removed through verification in overlapped images.

Secondly, overlap adequately increases the geometric accuracy of the fitted façades in the sense of reducing the absolute error as well as the standard deviation. The effect of overlap on the correctness of single façade (*single_correctness*) is not yet observed.

4.5 Conclusions

With the experiment on 335 individual buildings of various types, the proposed façade detection method proved its performance using both the Pictometry and SenP datasets.

The completeness on building level are 90% and 70% from SenP and Pictometry images respectively, even in the case that the study area is merely fully covered by images from one direction out of four. The façades are detected correctly by up to 70%. However, considering the detection on building level, the correctness can achieve 95%. Compared to the recent result from ISPRS bench mark (Rottensteiner et al., 2012), (also cf. Table 1.1), the completeness values from both data sets are within the range of the values from other researches, and the result from SenP is as good as the best methods in the benchmark test. The correctness is also in a high level, whereas the correctness from other works mainly ranges from 80% to 100%.

The geometric accuracy of detected façades is 1.5 degrees in orientation and 0.5 m in distance for images of high resolution and high overlap (SenP). Yet it can still reach around 2.5 degrees deviation in orientation and 0.7 m in distance with half of the resolution and very low overlap (Pictometry).

Verticality is the dominant assumption in the proposed approach. If the structures on non-vertical façades are still vertical and horizontal, they could be still detected, but the geometry of the fitted façades cannot be correct.

The proposed method is also sufficient for detecting curved facades and distinguishing parallel facades. But it fails when the texture on the facade is weak.

SUMMARY OF PART I

This part mainly contains the description of the developed method for detecting building façades from multi-view oblique images, and the experiments with two datasets over one study area. The results from both datasets are satisfactory compared to the recent research on detection. The comparison between the two image sets shows the impact of image resolution and overlap on the detection results.

The work in this part is innovative in the sense that building detection via façade detection is not covered in previous studies. The whole process is completely independent from any other data source. The whole approach requires only little training efforts to set some parameters, which can be considered as another advantage of this method.

This part is the first stage of the whole process of building detection. It is able to detect of 70% (Pictometry) to 90% (SenP) of buildings. The result from the work will be used as the input for the following stage of building outlining.

PART II: BUILDING OUTLINING

Extracting building outlines is one step further in building mapping after façade detection. The extracted outlines can be useful for updating 2D cadastral maps or 3D modelling. The sole use of façade information as retrieved from the method developed in Part I might not be sufficient to outline entire buildings since façades often do not get completely detected. Therefore, in Part II the outlines are going to be extracted using roof information in addition.

Each detected building is considered separately for outlining, by integrating geometric and radiometric information of the images. Features that are important for determining the outlines are discussed first. A classification approach is subsequently implemented to classify the pixels into on-roof and off-roof ones. The contour of the on-roof pixels is finally generalized and regularized by the 2D projection of the detected façades and extracted 3D lines from multiple views.

The outlining approach will also be tested experimentally on both the SenP and Pictometry datasets. The results are assessed by using the cadastral map, followed by the comparison between results from the SenP and Pictometry datasets. Chapter 5 presents the developed outlining approach, while the results and discussion are in chapter 6. In order to compare the results from two data sets, only buildings commonly detected are used for the evaluation of the result. Given that detected buildings can also be missed in the outlining step, final detection quality will be assessed in the beginning of chapter 7.

Chapter 5 Method and approach

5.1 General concept for building outlining

Although the outlining is aimed in 2D, the point cloud still functions as the starting point of the whole process. The generated point cloud combines the colour and geometry information from multiple images, so that it is considered to be more reliable compared to starting with only a single image or a pair of images.

Methods for outlining from Lidar data were developed several years ago (Dorninger and Pfeifer, 2008; Niemeyer et al., 2011; Vosselman, 1999), however it is not feasible to directly apply them on the dense matching point cloud. Firstly, the point density varies a lot and it highly relies on the texture on the plane. Secondly, as listed in Table 2.3, the RMSE from the points to the corresponding planes are about 0.4 m to 1 m. It is therefore difficult to extract accurate planes from it, especially on sloping roofs with tiny structures, such as dormer windows or chimneys. Moreover, it is impossible to define the edge when the contour of the roof plane is so noisy and the wall planes are not always available.

Due to the problems arising from the matched point cloud, other complementary information should be extracted from the source images. Considering that one plane usually has coherent colour and texture, using image segments, defined by colour, would be helpful to define the complete plane range. Lines extracted from the original images sometimes indicate the edges of a certain plane, so they will also be used to find the plane contour. The sharp edge is one of the advantages of the images over Lidar, given an appropriate image resolution (as it is the case here).

5.2 Workflow

Buildings are detected and outlined from the detected façades input from the previous part. The overall workflow is shown in Figure 5.1 and described in the following sections in detail.

As the façade completeness for one building can be quite low (cf. Figure 4.9), the point cloud around an input façade needs to be generated first to define the ROI of a target building. An initial ROI is expanded with the aim to include all off-ground points. Façades located in that area are clustered and regarded as façades of the target building. In contrast to the point cloud used for façade detection, which is generated from images viewing in one direction, all images containing the initialised region are employed. Since the target region is quite focused and the input images are cropped, the parameter *level* of PMVS procedure is set smaller than the one used for façades detection to obtain higher point cloud density and accuracy. As a side

effect the cropping of images results in very short processing times for the PMVS method.

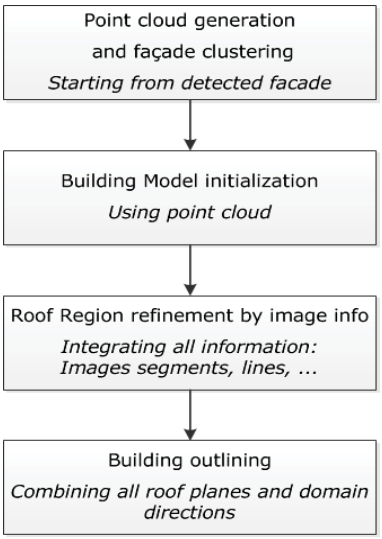


Figure 5.1 Frame work for building outlining from detected façades

Next, a coarse building model is built from the point cloud. Planar patches are extracted from the original point cloud. According to the location and normal of the patches, points belong to them are classified into ground, wall and roof. We take the wall and roof points as the initial building model, eliminating the ground points.

However, the initial point cloud is not good enough for direct outlining. The accuracy of the outlining is affected by the accuracy of the point cloud, the mismatching and the unmatched parts. Therefore image segmentation and lines are integrated into the progress. The ROI is rasterized for the fusion of information from images and object space. We take one roof plane as the processing unit because 1) this allows to obtain more accurate point positions in the images by fitting a plane to the points; 2) the plane bound can be improved by the bound of one or more image segments; 3) it is also likely that the plane bound corresponds to extracted lines. By employing an integration strategy over all information, the coverage of each plane in the grid area is defined.

The dominant directions of the buildings are collected from the input façades, the wall patches, the extracted lines as well as the normal of the sloped plane. Then we stack the regions of all roof planes and do the outlining considering the mean directions.

5.3 Point cloud generation and façade clustering

Taking the detected façades as the input for building detection and outlining, they should be previously clustered in the unit of a building. However, buildings have various sizes, and buildings might not have all detected façades due to some missing ones, thus it is difficult to cluster façades by their spatial distribution. However they could be connected by the off-ground point blob. The assumption is that each separated point blob presents one building block.

A point blob is generated from one façade. Other façades belonging to this blob are added later. To go through all the detected façades, they are firstly sorted by length. For any new building block, we start with the longest façade. The façades used for one building block will not be used for other buildings.

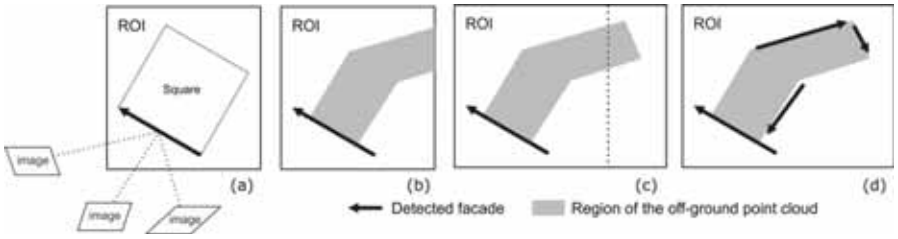


Figure 5.2 Point cloud generation from detected façade and selection of relate façades.

The direction of a façade is defined by its source images, obeying the rule that the viewing images are on its left. Under this definition, the target building should be on the right side of the façade. To generate the point cloud for the input façade, a square taking the input façade as one of its edges is put on the right side of the façade (Figure 5.2(a)). This square is defined as the initial ROI to select viewing images and generate the point cloud.

It is probable that the initial ROI is not correct for the target building, so that it is iteratively adjusted by extracting the off-ground blob (Figure 5.2(b)(c)). When the blob reaches one side of the ROI, it extends to that side. After the iteration, the ROI is redefined by the blob. Then other detected façades close to the blob are selected (Figure 5.2(d)).

5.4 Building model initialisation

The final point cloud inside the ROI is used for the initialisation of the building model. All the points are classified into vegetation, ground, wall and roof.

Then the flat and slope roof components are recognized from the roof points. In the last step, the roof planes are extracted.

5.4.1 General strategy

A building model consists of walls and roof planes. The extraction of model planes however cannot be easily done through point segmentation due to the noisiness from tree crown edges, mismatching or other causes. To cope with the noise, a three-step approach is proposed: 1) wall and roof point extraction; 2) roof component recognition; 3) roof plane detection.

The first step will be done by the classification of points into ground, vegetation, wall and roof points. Wall points can be used directly, yet the roof points need to be further processed to extract roof planes.

Roof points are segmented into planar patches first. The extracted flat planes are relative reliable since their plane normals can be restricted to vertical. The directions of sloped roof planes can be easily biased by the noise inherent in the point cloud. Sloped roof components consisting of multiple sloped planes are therefore extracted as a whole before the single plane extraction. Then the directions of the sloped planes in one component is defined under the constraint that their plane normals should be the same or follow some symmetries.

Flat components are extracted by selecting the flat patches with near-vertical normals. Next, sloped components (e.g. on a gable roof) are recognized by a slicing process to cope with the noise and separate disconnected sloped components. It is possible that one sloped component contains several connected roof planes. They are put in the same component to make their slopes consistent in the outlining process.

5.4.2 Building points extraction

Trees close to buildings need to be considered first. During the leaf-off season, the points on the tree branches are not significant compared to the noise in the generated point cloud, which is also considered as noise, but points on the trees during the leaf-on season should be removed. One advantage of the point cloud from dense matching over traditional Lidar points is that it still carries the RGB colour information from the images. This is used to remove the vegetation points. A vegetation index from red and green bands is adopted (Shorter and Kasparis, 2009), which was said to be more efficient than using the red band proposed in (Sirmacek and Unsalan, 2008):

$$I_v = \frac{4}{\pi}\arctan\left(\frac{R-G}{R+G}\right) \qquad \textbf{(5-1)}$$

A global threshold is set for I_v to remove the major part of the tree crown but preserve all building points. The remaining tree points will be removed in the later steps.

After removing the tree points, points are segmented into patches by surface growing. Then the patches are classified into roof, wall and ground based on the plane normal and plane height.

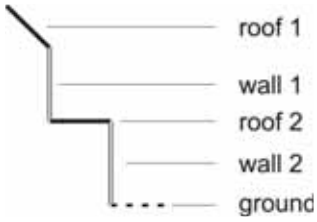


Figure 5.3 Classification of patches in to roof, wall and ground according to plane normals and positions.

Wall patches are firstly distinguished because they have unique vertical planes. Wall patches exist in most cases but not all, depending on their visibility in all images.

After extracting wall patches, points on the ground need to be removed from roof points. An initial local ground height is estimated from the lowest 5% of the original points. The ground height does not have to be very accurate, because we only need to separate roof and ground, between which there is a clear gap.

When no wall patch exists, the estimated height is taken as the local ground height. When wall patches are available, the height of lower wall points is compared with the estimated ground height. If the difference between the two is not much (e.g. 3m), the local ground height is adjusted according to the bottom of the wall. Otherwise the estimated height is kept as local ground height. As illustrated in Figure 5.3, it might happen that only wall 1 is detected but not wall 2. This implies that the height defined by the wall will be much higher than the ground, even higher than roof 2. The points lower than 3 m above the ground height are removed as ground points.

5.4.3 Components recognition from roof points

After removing ground and wall points, roof components should be extracted from the remaining points. Within a building block, components are separated by height, connectivity and roof type.

Roof patches with a near-vertical plane normal are classified as flat roof patches. Due to the noisiness, a roof plane is fragmented into small patches. Then, adjacent patches with similar height are connected to form a flat roof component.

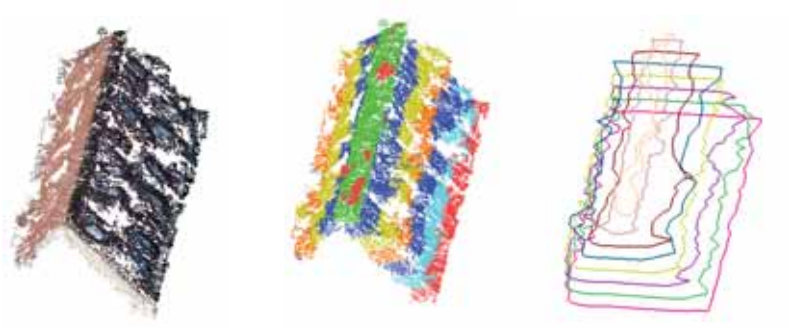
Points, excluding those belonging to flat components, are then separated into connected components. For each connected component, points are sliced into pieces by a certain height interval. From the second slice from top, each slice is compared with the previous slice to examine its validity to be part of a sloped component:

- Area of a slice $Area_i$ is computed from its convex hull. The area of the slice is compared with the area from the previous slice. The increment of the area should be larger than a threshold T_{area_i};
- The points in a slice $Points_i$ are defined as the points within the height interval and inside the convex hull of the slice. The number of its points should be larger than T_{points_i};

T_{area} and T_{points} are initialised as 1 m^2 and 20 respectively. After testing two slices, they are then set to half of the average of the values from the tested slices. Therefore:

$$\begin{cases} \boldsymbol{T_{area\_i} = 1}, & \boldsymbol{T_{points\_i} = 20} \quad , \ \boldsymbol{i = 1, 2} \\ \boldsymbol{T_{area\_i} = 0.5 \times \frac{\sum_{k=1}^{i-1} Area_k}{i-1}}, & \boldsymbol{T_{point\_i} = 0.5 \times \frac{\sum_{k=1}^{i-1} Points_k}{i-1}}, \ \boldsymbol{i > 2} \end{cases} \quad \textbf{(5-2)}$$

A sloped component is then recognised by putting together all slices fulfilling those criteria (Figure 5.4). By doing this, fragments or "tails" from the tree crown edges can be excluded.



(a) (b) (c)

Figure 5.4 Slicing of slope roof points. (a) original slope roof points; (b) Slices of height points in different colours; (c) contours of height slices.

5.4.4 Roof plane detection

Roof planes are extracted from each of the roof components as the processing units for information integration and region definition. For flat components, each separate component at a different height is a single plane.

All points of a sloped component are firstly segmented into small planar patches. Then the dominant plane normals are extracted from all patches. For each dominant normal, planes are fit to the roof points. All points within a certain distance of a plane are assigned to it. This distance is decided according to the residual of the point cloud. There is overlap between the point sets of intersecting planes because the points close to the intersection line are assigned to both planes. After all roof planes are extracted, the assignment of selected points is optimized using the method in (Vosselman, 2012). For one point, the root mean square (RMS) sum of the distance of the neighbouring points to the planes are checked, and the point is assigned to the plane of smaller RMS sum.

Figure 5.5 shows the process of initialising building models from input façade hypothesis (a). The major part of vegetation is removed in (b) but some points at the edges remain. Then wall components (c) and roof components (d) are extracted from all remaining points. The pink and purple flat components are different planes due to different heights. The shown sloped component in green represents two connected gable roofs. Then four planes are extracted from the sloped component shown in (e).

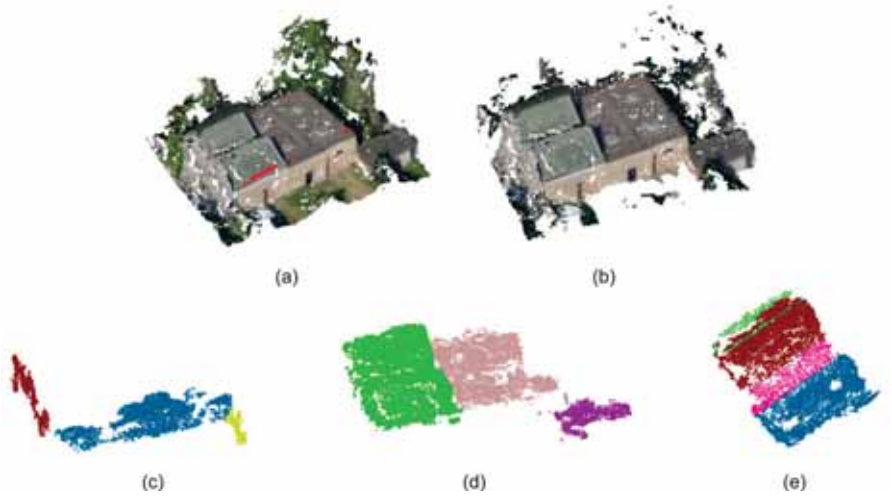


Figure 5.5 Initialisation of a building model from input façade hypothesis. (a) input façade hypothesis in red and generated original point cloud; (b) tree points removed; (c) wall planes; (d) roof components: green – sloped, pink and purple – flat; (e) planes from sloped component in (d).

5.5 Roof region refinement by image information

As shown in Figure 5.5, the initially extracted roof plane points might not be suitable for outlining because the planes are incomplete and the boundaries are quite noisy and irregular. Hence we involve image segments from all source images and extracted lines to better delineate the outline. A 2.5D grid space is created to carry both the information from 3D point cloud and 2D information from images. We mainly focus on the 2D projection of a plane in the grid space due to the 2D outlining purpose. Each pixel contains values from several features. Then a graph-cut classification is used to refine the region of the plane projection.

5.5.1 General strategy

Recalling that all available information for roof region definition is:

- roof planes from point cloud – planar feature in 3D object space;
- projections of wall patches from point cloud – linear features in 2D object space;
- extracted lines from images in 3D – linear features in 3D object space;
- image segmentation – planar feature in 2D image space.

In order to integrate all this information, the ROI is rasterised into a discrete 2.5D grid space, whose grid size is determined according to the point cloud density which is influenced by the original image resolution. The processing is mainly in 2D, but the height information helps to organize the relations between roof planes. This grid space can be presented by an image.

Only a translation and scaling is needed to transfer features in the object space into the grid space. The following sections describe the extractions of lines and segments' information, and the way to integrate all the information to create the region image.

5.5.2 Line extraction

For a target building, we aim to collect the roof edges to assist the refinement of the roof region. In order to use the combined edge information from the images, 3D lines are generated from line features in different images.

Straight line features are extracted from each image using the line extraction method by (Förstner, 1994), which we have also used to extract line pixels in chapter 3. The direction of the gradient is taken from the largest eigenvector of N (cf. formula 3–1).

Line hypotheses are initially defined by stereo intersection of 2D lines from images. Since any intersection results in a 3D line, a minimum of three

matched image lines is required to collect the robust 3D lines. This process not only removes the wrong matches, but also reduces the chance of collecting lines on ground or façades. Lastly the matched lines on the ground or façades are further removed by testing their distance to the roof planes. By this step, we are able to collect highly reliable 3D lines, but their number is quite limited.

5.5.3 Image segment information extraction

Based on the assumption that a plane would mostly have homogenous colour in the image, a colour segment from the image could present an object plane. Therefore, the original images are segmented by colour and texture, and the segments are selected according to the known plane points. Then those segments from all images are combined in the object space to form a possibility map.

Image segmentation

Each source image is firstly segmented. Many image segmentation methods are available, involving features such as brightness, colour, or texture. The segmentation methods of the features vary from mean-shift clustering (Comaniciu and Meer, 2002), region growing to graph-based (Felzenszwalb and Huttenlocher, 2004) or contour based approaches (Arbeláez et al., 2011). Our approach is not restricted to the method used, since it will combine multiple results from images from different views.

We used the graph based method by Felzenszwalb (2004). This method tends to generate the result neither too coarse nor too fine. It builds a graph in which pixels are nodes, and the edge weights are the colour dissimilarity between nodes. This method is able to preserve detail in low-variability image regions but ignore detail in high-variability regions. In this sense it will preserve large roof parts in one segment, excluding tiny structures. For the outlining purpose, the parameter to control the observation scale in the approach is set to give a preference for larger components, so as to put more emphasis to roof segments.

An example of image segmentation results from four viewing directions is shown in Figure 5.6. The segments are quite different from each other. The segmentation in the image viewing south failed to distinguish the roof plane from the façades, probably due to the shadowing. But in the other three directions, two roof wings and the centre rectangle are successfully segmented.

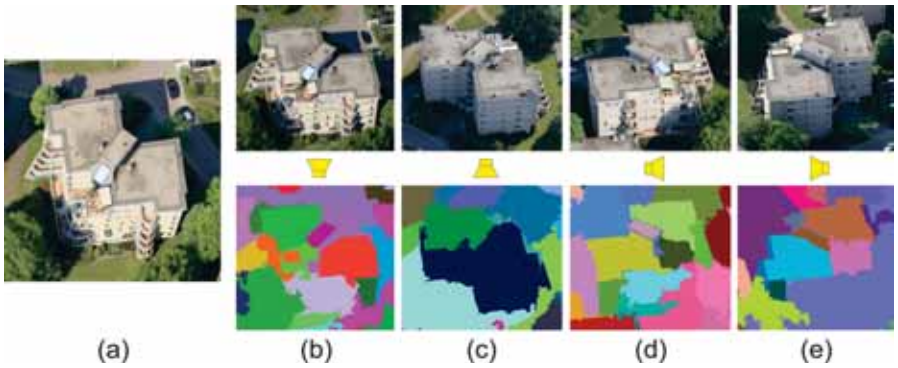


Figure 5.6 Original image and result from image segmentation from four viewing directions ((b)-(e)). (a) is the top view of the building for reference. Images are from SenP.

Image segments selection and weighting

In a single image, one segment is taken as a processing unit under the assumption that the segmentation method is able to successfully separate different objects. The segments are selected and assigned weights by using both roof and wall points. The number of roof points adds positive value to the weight whist the number of wall points is counted as negative.

Roof and wall points are projected into the images for segments selection. Before projection, roof points are restricted onto the fitted plane so as to reduce the inaccuracy of their projections on the image. The normal of a flat plane is set to (0, 0, 1) and the normal of a sloped plane is the one derived through the sloped plane extraction in section 5.4.4.

A simple self-occlusion test for wall points is conducted beforehand. The test is done by intersecting the ray from the centroid of the patch to the camera position with the roof planes. All the points in the patch are set to invisible if there is any intersection. The occlusion from others are not tested because the point cloud is only limited in the ROI of the target building during the process. In most cases, however, the effects from one side occlusion can be balanced out by images from other side.

The segment weight W_{seg} is defined by the number of roof points C_{roof}, and the number of wall points C_{wall} inside the segments. So the segment count C_{seg} is defined as:

$$\boldsymbol{C_{seg} = C_{roof} \cdot \frac{C_{roof}}{C_{roof}+C_{wall}}} \qquad \textbf{(5-3)}$$

Assuming that a roof segment should have many roof points projected onto it, C_{seg} is proportional to C_{roof}. However, there are problems of under-

segmentation where a segment consists of both building roof and wall. The count for this type of segment should be distinguished from the roof segments and given lower count. Therefore the ratio of C_{roof} against all roof and wall points is used as the weight for C_{roof}.

Then two values C_{min} and C_{max} are set to reject the segments with too small number of count and to prefer the segment that has large number of roof points in respectively. Given that the size and point density are different between roofs, C_{min} and C_{max} are related to the total number of roof points N_{roof}

$$W_{seg} = \begin{cases} 0 & , C_{seg} \leq C_{min} \\ (C_{seg} - C_{min})/C_{max} & , C_{min} < C_{seg} < C_{max} \\ 1 & , C_{seg} \geq C_{max} \end{cases} \quad \textbf{(5-4)}$$

In this example, C_{min} and C_{max} are set to 0.1 × N_{roof} and 0.2 × N_{roof} respectively.

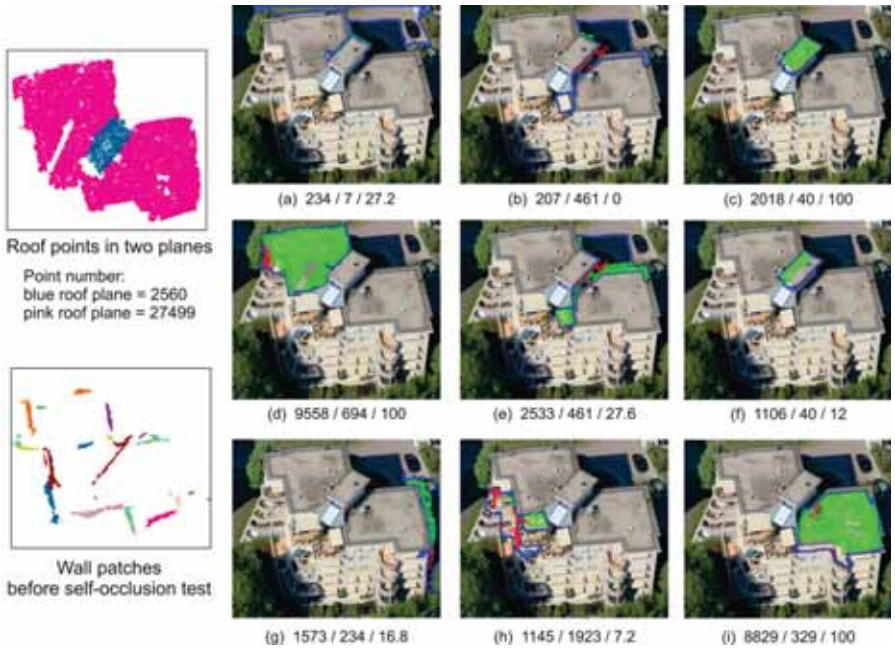


Figure 5.7 Example of image segment selection and weights assignment. Images (a)-(c) are from blue roof plane. Images (d)-(I) are from pink roof plane. Blue line – contour of the segment; green points – projected roof points; red points – projected wall points. The three numbers below the images are C_{roof} / C_{wall} / W_{seg}. Images are from SenP.

Some examples on segment selection and weight assignment taken the north view image in Figure 5.6 are presented in Figure 5.7. The W_{seg} is scaled to 100 to avoid decimals for visualisation. The main part of the roof planes are successfully picked out in (c), (d) and (i) with weight 100. The other segments which contain a small part of the actual roof are assigned low weights.

Comparing between (a) and (b), one sees that C_{roof} is almost identical, but C_{wall} in (b) is larger than C_{roof}, which results in a lower weight. The effect of setting C_{min} and C_{max} relating to total point number of the plane can be checked with (c) and (e). Although the C_{roof} values are similar in them, W_{seg} of the segment in (c) is 100 because C_{roof} exceeds C_{max} of the blue roof plane. W_{seg} in (e) is only 27 because the total number point of the pink roof plane is almost 10 times of the number of the blue plane.

Fusion of segment weight images into grid space

A grid image G_i is created in the ROI for each image i. The value of the image pixels are resampled from the segment weight image SW_i. The segments are selected as the ones presenting the specific roof plane, thus the imaging from SW_i to G_i is:

$$\boldsymbol{G_i(u,w) = SW_i(x,y)} \qquad \textbf{(5-5)}$$

where

$$\boldsymbol{(x,y) = Proj_i(X,Y,Z)} \qquad \textbf{(5-6)}$$

$Proj_i$ is the projection (x, y) on image i from a point (X, Y, Z) in object space.

$$\boldsymbol{Z = \frac{-X \times n_x - Y \times n_y + d}{n_z}},$$

$$\boldsymbol{X = u \times gsd + ROI_{xmin},\ Y = w \times gsd + ROI_{ymin}} \qquad \textbf{(5-7)}$$

X, Y, Z are defined by the roof plane $P\ (n_x, n_y, n_z, d)$, the start point of the ROI (ROI_{xmin}, ROI_{ymin}) and the ground sample distance gsd of the grid space.

Fusion of all these grid images is done by stacking them. The value of each pixel is the average of all values from each image, that is

$$\boldsymbol{G = \sum(G_1, G_2, \dots, G_n)/n} \qquad \textbf{(5-8)}$$

G is the fused segment image, and $\boldsymbol{n}$ is the total number of grid images.

Before using the segment image, the validity of a plane is tested by setting a threshold t. It is necessary to remove the false plane in the model, because the false plane will result in no selected segment in one image or small intersection from the selected segments from different images. If G_{max} is lower than t, it indicates that no efficient segment is selected from the image. In this case, the roof plane should be rejected. Value t is not very sensitive since the segment weights are already controlled during the selection stage.

5.5.4 Roof region from all evidences using graph-cut

Image $\boldsymbol{G}$ is the initial roof possibility image. The aim of this section is to classify pixels in $\boldsymbol{G}$ into *roof* and *background*. The previously collected wall patch projections and the projections of the extracted 3D lines should be taken into consideration during the classification as an external constraint to indicate the region edge. In the resulting roof region image, the roof should be very homogeneous inside but sharp at the boundary.

This problem can be solved by an energy minimisation approach. Energy is defined as the function of the class of each pixel. The classification achieves the optimum when the energy is the minimum.

In our study, every pixel $\boldsymbol{p} \in \boldsymbol{G}$ is assigned a class $\boldsymbol{f}$ in the class set $\boldsymbol{L}$ {*roof, background*}. The energy consists of data terms $\boldsymbol{E_{data}(f)}$ and piecewise smoothing term $\boldsymbol{E_{smooth}(f)}$ from neighbourhood interactions. The two terms are balanced by a constant $\gamma > 0$:

$$\boldsymbol{E(f) = E_{data}(f) + \gamma \cdot E_{smooth}(f)} \qquad \textbf{(5-9)}$$

Here, $\boldsymbol{E_{data}}$ measures the sum of the difference between class $\boldsymbol{f}$ and the observed data on each pixel, denoted $\boldsymbol{D_p(f_p)}$:

$$\boldsymbol{E_{data}(f) = \sum_{p \in G} D_p(f_p)} \qquad \textbf{(5-10)}$$

$\boldsymbol{E_{smooth}}$ is the function makes the object boundary smooth. Take a pixel pair $(\boldsymbol{p}, \boldsymbol{q})$ in the neighbourhood set $\boldsymbol{GN}$, their interaction is denoted $V_{p,q}(f_p, f_q)$, then the smooth term is:

$$\boldsymbol{E_{smooth}(f) = \sum_{(p,q) \in GN} V_{p,q}(f_p, f_q)} \qquad \textbf{(5-11)}$$

Therefore:

$$\boldsymbol{E(f) = \sum_{p \in G} D_p(f_p) + \gamma \cdot \sum_{(p,q) \in N} V_{p,q}(f_p, f_q)} \qquad \textbf{(5-12)}$$

$\boldsymbol{D_p(f_p)}$ and $\boldsymbol{V_{p,q}(f_p, f_q)}$ are then defined in the following paragraphs.

The observed data comes from the initial roof possibility image $\boldsymbol{G}$ where the value for each pixel $G_p \in [0,1]$ measures the possibility for the pixel to be roof. Therefore the difference between the class and the observation is:

$$\boldsymbol{D_p(f_p)} = \begin{cases} \boldsymbol{1 - G_p}, & \boldsymbol{p = roof} \\ \boldsymbol{G_p}, & \boldsymbol{p = background} \end{cases} \quad \textbf{(5-13)}$$

The aim of the smoothness is to measure the interaction between neighbours. Lafarge and Mallet (2012) proposed a label propagation constraint model to define the smoothness term. They use this model to label the pixels in a grid space from point clouds. This model concerns the 3D line segment constraint during the definition of neighbourhood: points that are on two sides of a segment are not defined as neighbours. It also introduced a coherent arrangement of the structures. This arrangement considers the label of the point around the intersection of the planes. A penalty is given when the labels are not coherent with the intersection, for example two different labels on the same side of an intersection

In this work, line segment constraints are also obtained through the previous steps. Therefore a similar smoothness term is defined. In contrast to (Lafarge and Mallet, 2012), the line segments in this work come from three different sources: detected façades, wall patches and matched 3D lines. Semantically these are different objects, especially for many sloped roof buildings due to roof overhangs: the façade projections and line segments from wall patches are the wall, and 2D lines and roof segment weights are roof outlines. Therefore these lines are sometimes in conflict with each other. As we prefer the 2D outline, the projections from detected façades and wall patches are given higher weight. This method cannot, however, ensure that the wall outline is retrieved. When no wall information is extracted, the roof outlines are used instead. This would result in lower geometric accuracy when being evaluated using the cadastral maps.

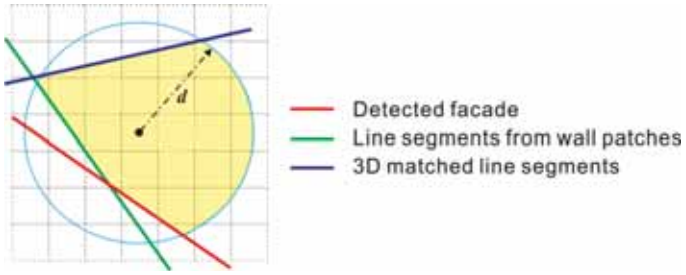


Figure 5.8 Neighbourhood constrained by the line segments. d is the searching radius. All pixels fall inside the yellow scope are the neighbours of p.

The neighbourhood is firstly defined for the definition of the smoothness term:

$$(\boldsymbol{p},\boldsymbol{q})\epsilon \boldsymbol{N}\ \boldsymbol{when}\begin{cases}\boldsymbol{Distance}(\boldsymbol{p},\boldsymbol{q}) \ \leq \ \boldsymbol{d}\\ \boldsymbol{O}(\boldsymbol{p},\ \boldsymbol{L}) = \boldsymbol{O}(\boldsymbol{q},\boldsymbol{L})\end{cases} \qquad \textbf{(5-14)}$$

d is the maximum distance between the locations of two neighbouring pixels. $\boldsymbol{L}$is oneline segment. $O(p,\boldsymbol{L})$ is the oriented side in which the cell $\boldsymbol{p}$ is located with respect to $\boldsymbol{L}$. As illustrated in Figure 5.8, neighbouring pixels should be on the same side of all segments in their search scope.

The pairwise interaction between neighbours is formulated as:

$$\boldsymbol{V}_{p,q}\left(\boldsymbol{f}_p,\boldsymbol{f}_q\right)=\begin{cases}\quad\quad \boldsymbol{f}_p=\boldsymbol{f}_q \,\&\, O(\boldsymbol{p},\boldsymbol{L}) = \boldsymbol{O}(\boldsymbol{q},\boldsymbol{L}) & -(1)\\ \epsilon\ ,\ \boldsymbol{f}_p\neq\boldsymbol{f}_q \,\&\, O(\boldsymbol{p},\boldsymbol{L}) \neq \boldsymbol{O}(\boldsymbol{q},\boldsymbol{L}) & -(2)\\ \quad\quad \boldsymbol{f}_p=\boldsymbol{f}_q \,\&\, L=\emptyset & -(3)\\ \mathbf{1}\ ,\ \boldsymbol{others}\end{cases} \qquad \textbf{(5-15)}$$

$\boldsymbol{\epsilon}$ is the penalty value in [0, 1]. The emphasis on the line segment constraints is higher when $\boldsymbol{\epsilon}$ is given a lower value.

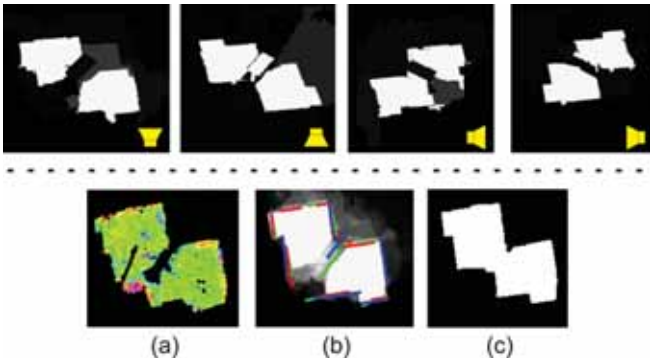


Figure 5.9 Roof region refinement. Upper part is the example of selected regions from images of four viewing directions. Lower part (a) is the original roof plane points coloured on height; (b) is the initial possibility image with line constraints: red – detected façades; green – projections of wall patches; blue – extracted 3D line segments (c) roof region after graph-cut minimization. Light is the roof region, dark is the background.

After the definition of energy, a graph-cut based algorithm (Boykov et al., 2001) is used to minimize it so as to find the optimal classification. Figure 5.9 is an example of the proposed method for roof region refinement. It takes the example of the pink roof plane in Figure 5.8. The result in (b) shows that the roof region is already well distinguished from the background. The small part at the left bottom corner is not recognized as part of the roof due to its

lower height. Three types of line segment constraints are drawn in different colours in (b). The grey value in (b) indicates the data term. Lighter means higher possibility for a pixel to be classified as *roof*. The white region in (c) is the final roof plane region. Due to the effects from the higher small plane (blue plane in Figure 5.7) during the segments selection step, the middle part, which belongs to another part of the roof, is still classified as *roof* for this plane. However it has no impact on the delineation of the final outline of this building. The small part will be distinguished through the further outlining step.

5.6 Outlining

After refining the roof regions of a building, the 2D grid regions are projected onto their original plane to obtain refined roof planes. They are then combined to form the final outline.

In order to generate accurate outline from refined roof regions, a direction set is employed to control the directions of the line segments composing the outline, containing the dominant directions of the outline segments. It is initialised by the directions of the projections from detected façades, wall patches and extracted 3D line segments. If there are sloped components outlined for this building, the slope directions also contribute to the direction set. Given that the roof normal could be biased by noise of the point cloud, the extracted lines are given higher priority. In order to give preference to orthogonality, if one direction is found, it and its perpendicular direction are both added into the direction set.

For a complex building consisting of both sloped and flat components, outlining starts from sloped components since their shapes are supposed to be more regular than that of the flat components. Then all components are stacked and the outer boundary of the building is generated with the direction set updated by the normal of sloped planes.

5.6.1 Outlining a sloped roof component

In general, a rectangle is fitted to each refined sloped plane, and the outline of the whole component is composed by intersecting all rectangles. However, the sloped plane can be easily biased by the noise inherent in the point cloud. The outlines of the sloped components therefore require constraints from additional geometric considerations.

The following constraints are adopted to generate the outline of a sloped component:

- The sloped planes are usually symmetric, thus the edges of the face planes are parallel to the ridge lines;

- The edges of the side planes (if they exist) are usually perpendicular to the edges of the front/back planes, thus the edges of side planes are perpendicular to the ridge lines;

- The sloped plane's normal is perpendicular to the projection of the attached façades and those evidences have higher accuracy than fitted sloped planes. Thus if any projection of the detected façades, wall patches or extracted 3D lines are around, the outline should follow the direction of the projections.

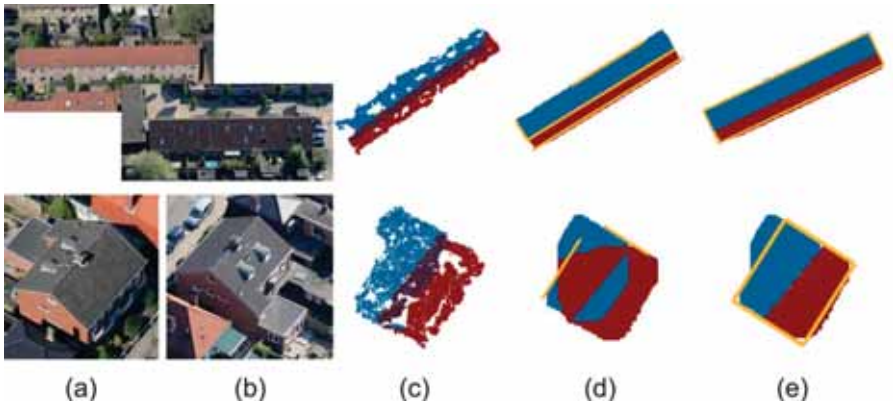


Figure 5.10 Outlining of the sloped component. (a) original image viewing front plane; (b) original image viewing back plane; (c) recognized roof planes from dense matching point cloud; (d)roof planes after region refinement with extracted lines; (e) intersected roof planes with final outline.

Examples of the outlining of two gable roof buildings are shown in Figure 5.10. (a) and (b) are images viewing the front and back roof planes respectively. Both building roof faces are recognized as sloped components and two sloped planes are recognized (c). Refined roof planes and the overlaid projections of all evidences presented in (d).

For the building in the first row, the roof planes after refinement are quite regularized and the projection of the evidences matches the edges of the refined roof planes very well. Two problems need to be addressed for the building in the second row:

1) There is an incorrectly computed overlap of the front and back planes, which is due to the wrong image segmentation used in the plane refinement step. When the slope is gentle, one roof plane side will still be visible from the images viewing from the other side. This leads to a false segmentation covering two roof planes. For roofs with a large slope whose front and back planes can be completely viewed separately in the

images from opposite views, the refined roof regions stop sharply at the ridge (first row, column (d)).

2) There is a difference between the direction of the intersection of the two planes and the direction of the evidence projections.

The first problem can be solved by intersecting roof planes and removing the upper part of the intersection. The second problem is solved by following the geometric constraints mentioned above. In this case, the projections have higher weights and then the roof slopes are adjusted to the directions of the projection. The final outlined results are presented in (e).

5.6.2 Outlining a flat component

Outlining of the flat planes mainly depends on the defined direction set.

For each processed plane, the outer contour of the points is extracted. The outline is the generalization of the contour. The projections of the detected façades and wall patches around the contours are used as the composition of the outline. Then line segments are fitted to the gaps where no façades or wall patches can be found.

The gaps are first fitted by the lines of the dominant directions, and the nodes on the lines are removed from the list. If the gap cannot be fitted by lines using any direction from the direction set, a line with a new direction is fitted by using a Hough transform. Meanwhile, the direction set is updated by the new direction as long as the fitted segment is longer than a threshold.

Finally, the fitted segments are connected according to the following criteria:

$$\begin{cases} angle \leq \theta, dis \leq d, & merge \\ angle \leq \theta, dis > d, & insert\ a\ segment\ perpendiculer\ to\ them \\ angle > \theta \quad , & intersect \end{cases}$$

where *angle* denotes the angle difference between two neighbouring segments. If the angle is smaller than θ (we use 5 degrees here), they are considered as parallel. Then *dis* denotes the distance between two parallel lines, *d* is set to 1 m.

5.6.3 Outlining the entire building

Figure 5.11 shows a sketch of the whole outlining process. All flat planes and the projection(s) of the outlined sloped component(s) at edge height in a building block are stacked for final outlining.

All the original flat planes are outlined separately in a sequence from top to down. Once a plane is finished, the dominant directions are updated by its long edges. Meanwhile, the points of the plane layer one level lower than it are also updated: points inside the outline of the upper plane will be eliminated from the lower plane. If the number of remaining points is too few, the plane will be removed.

After outlining all stacked planes, grid pixels inside all outlines are combined to form the final pixel set in 2D space. Lastly the building outline is extracted from it. The reason for outlining all flat planes separately before the final outline are 1) to reduce the noise separately for each layer and 2) taking the directions of inner structures into consideration for the final outlining.

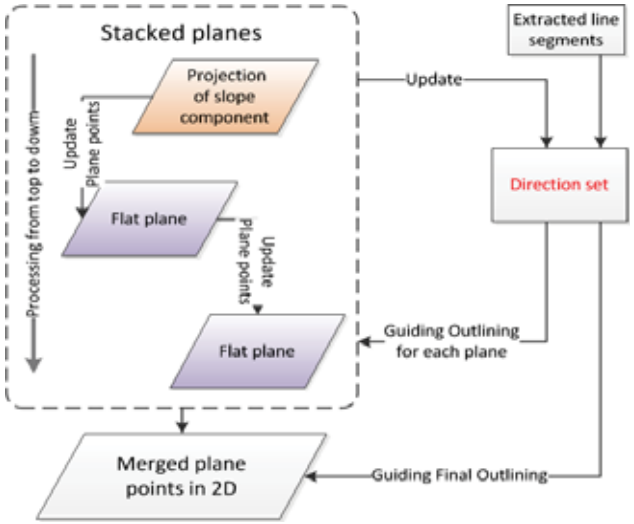


Figure 5.11 Flowchart of stacking plane pixels from multiple plane layers

Chapter 6 Results and evaluation

6.1 Experimental design

This part of the work makes use of the results from facade detection developed in the previous part. The experiment will be carried out over the same area using images from the SenP and Pictometry datasets.

The evaluation methods will only be used to compare the outlining results with detected buildings from the façade detection step rather than all reference buildings. In order to compare the results from SenP and Pictometry, only the buildings detected from both sets are used for the outline quality assessment. The final detection quality over all buildings in the study area will be presented in section 7.1.

In principle, the missed buildings from previous part cannot be detected by the outlining process. False detections, such as fences, can be rejected because no sufficient point cloud can be generated in the ROI around the façades. However, a few buildings may also be missed in the case where no sufficient point cloud can be produced.

6.1.1 Scenario design

Dense matched point clouds are generated using the oblique images from both datasets. Building models are then initialised from them. From each dataset, we obtain a set of building outlines by the proposed outlining method. The comparison between these two sets will be used to evaluate differences caused by different image resolution and overlap.

In order to test the effectiveness of the proposed roof region refining step integrating image evidences, the contours from the original PMVS points of building blocks are also included in the comparison. The convex hull of the raw points of one building block is first extracted using a connection distance of 3 m. The contour is then generated by simplifying the convex hull using a Douglas-Peuker algorithm with a distance setting of 1 m.

Taking into account that roof information from images plays an important role in the proposed method, available nadir images are added in the roof region refining step to test its effectiveness for outlining. However, nadir images are only available in the SenP data set.

In total, five scenarios are involved for the result evaluation. Naming of the scenarios and their meaning are summarised in Table 6.1. Contours and outlines from oblique images generated from one dataset have their corresponding results in the other dataset: *S_c* corresponds to *P_c* and *S_ol_o* corresponds to *P_ol_o*. Only SenP dataset has outlining result from oblique and nadir images: *S_ol_on*.

Table 6.1 Test images sets for outlining

Naming	Data set	Image type	Outline method
S_c	SenP	oblique	Contour from PMVS points
S_ol_o		oblique	Boundary from outlining process
S_ol_on		oblique + nadir	Boundary from outlining process
P_c	Pictometry	oblique	Contour from PMVS points
P_ol_o		oblique	Boundary from outlining process

6.1.2 Evaluation design

The outlining result is assessed based on the evaluation method of building detection from the ISPRS benchmark on object detection and 3D building reconstruction (Rottensteiner et al., 2012), which is based on the assessment method in (Rutzinger et al., 2009).

Besides a pixel-based evaluation, an object-based evaluation is performed to check the correspondence between outlined buildings with the reference. Then the accuracy of the outlines is evaluated by a geometric measure.

Considering that the evaluation is unable to assess the smoothness of the outline, we introduced the quality measure outline *smoothness* to quantify the regularisation of the delineated outlines in addition to the ISPRS benchmark.

Some basic terms need to be introduced first. As the achieved outlining result is in 2D, the outlining process can be considered as classifying the 2D space into *building* and *background*. Comparing the results with reference data it is defined:

- *True Positive (TP):* an entity (pixel/object) classified into the same class (building/background) as the corresponding entity in the reference;
- *False Negative (FN):* an entity (pixel/object) corresponding to a building in the reference that was classified as background;
- *False Positive (FP):* an entity (pixel/object) that was classified as a building, but corresponds to background in the reference.

Then the *completeness* and the *correctness* can be derived, indicating the ratio of actually detected buildings amongst all reference buildings and the ratio of actual buildings amongst all entities being classified as building, respectively:

$$Completeness = \frac{TP}{TP+FN} \times 100\% \quad \textbf{(6-1)}$$

$$Correctness = \frac{TP}{TP+FP} \times 100\% \quad \textbf{(6-2)}$$

I. Pixel-based evaluation

In the pixel-based evaluation, the raster representations of the detection results and the reference are compared. The comparison results are presented as raster images showing the spatial distribution of the *TP*, *FP* and *FN* pixels.

II. Object-based evaluation

The relations between outlined buildings and the reference buildings may be $(x{:}1)$ or $(1{:}x)$, where x could be zero, one or multiple. Therefore, the method described in (Rutzinger et al., 2009) is used to establish correspondences between buildings in the two datasets. It is done by overlaying the reference label image and the label image representing the outlining results. If the overlap between two buildings from outlined building set and reference is larger than 50% of the smaller building, a correspondence is set between them. Then a *topological clarification* (Rutzinger et al., 2009) is used to change the label images to ensure that each building in one set has only zero or one corresponding building in the other set (details in Appendix II). Afterwards the overlap analysis is repeated to find out the *TP*, *FP* and *FN*.

In the evaluation of façade detection all buildings where at least one façade were detected was considered as being detected. However, we did not differentiate between cases when the facades of different buildings were connected to one façade instance (Figure 3.12(d)). Similarly, when one building is wrongly separated into several parts, those parts are still considered to have been correctly detected. In order to address this kind of problem, the number of $m{:}n$ relations is also counted besides *completeness* and *correctness*. m and n can be one or multiple, but 1:1 relation is excluded. This relation indicates the capability of the method or the input data to separate adjacent buildings, which is important for applications in complex urban scene.

III. Geometrical evaluation

The geometrical accuracy of the vector outline is evaluated as well. It aims at testing the accuracy of the extracted outline as defined below.

The accuracy is computed by matching a vertex from the delineated boundaries to its nearest counterpart in reference. For each vertex of an

outline, the nearest point on the boundary of the corresponding object in the reference is searched. This point does not necessarily correspond to a vertex of that polygon. The distance d between the corresponding points is found. If this distance is larger than a threshold, it is discarded, assuming that this part is not semantically belonging to the same building. Finally, the RMSE of the distances $\boldsymbol{RMS_d}$ is computed:

$$\boldsymbol{RMS_d} = \sqrt{\frac{\sum \boldsymbol{d^2}}{\boldsymbol{N}}} \qquad \textbf{(6-3)}$$

where N is the number of points for which a correspondence has been found within a predefined distance.

IV Outline smoothness evaluation

Based on the assumption that building outlines should be regular, we should use as few line segments for the outline as possible. Therefore, we use the number of the line segments in the outline to specify the *smoothness* of the outline, defined as the ratio between $N_{outline\_segments}$ and the number of line segments in the corresponding reference $N_{reference\_segments}$.

$$\boldsymbol{smoothness} = \frac{\boldsymbol{N_{outline\_segments}}}{\boldsymbol{N_{reference\_segments}}} \qquad \textbf{(6-4)}$$

The *smoothness* is measured based on the matched reference buildings after ***topological clarification***. If the smoothness is less than 1, the generated outline is assumed to be more generalized than the original one. If the smoothness is larger than one, the generated outline is assumed to be noisy.

6.2 Results & evaluation

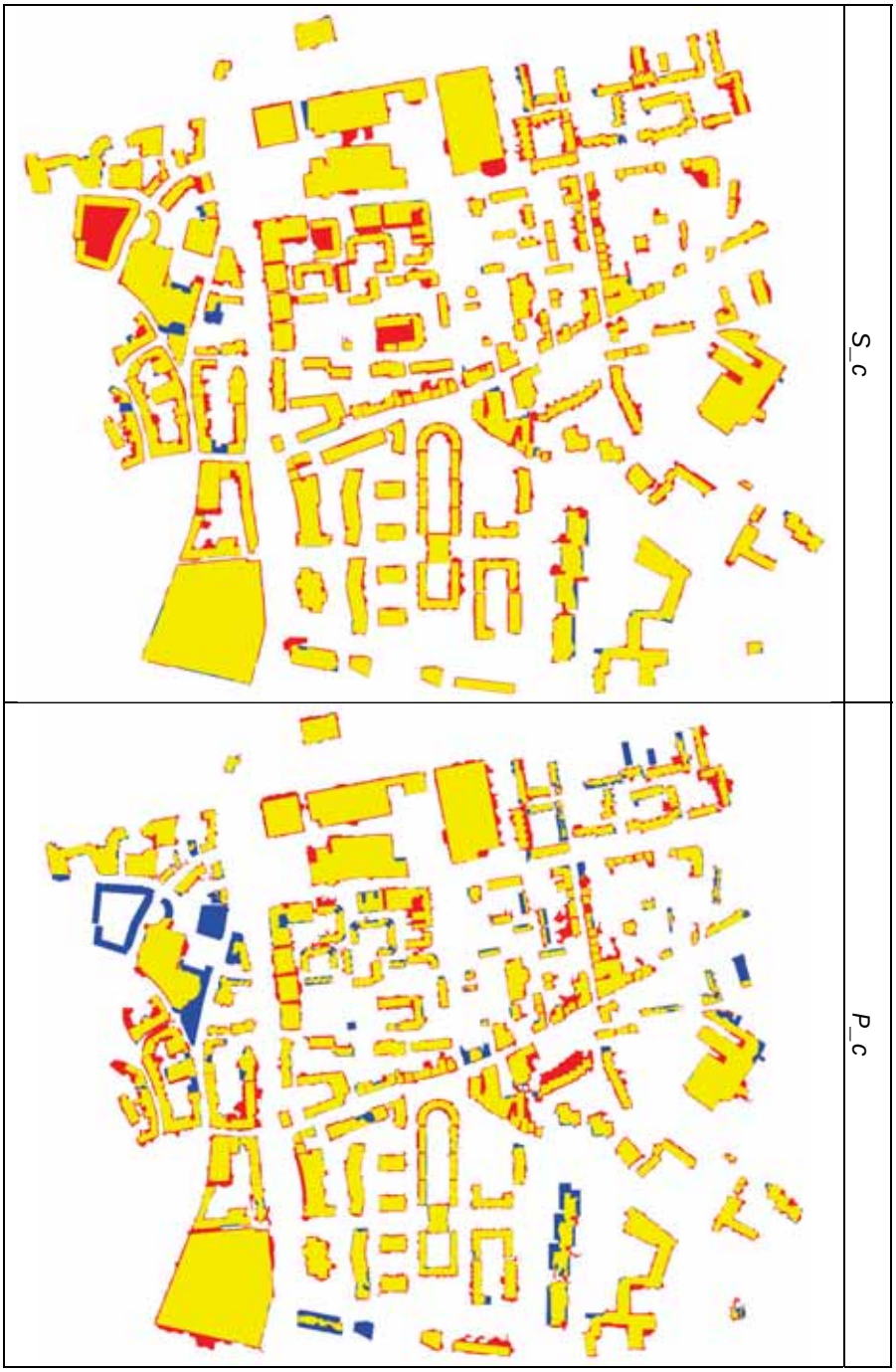
Results from the proposed outlining method on the same set of buildings using the SenP and Pictometry datasets are presented and analysed. Pixel-based results, object-based results and geometrical accuracy are presented in the following sections.

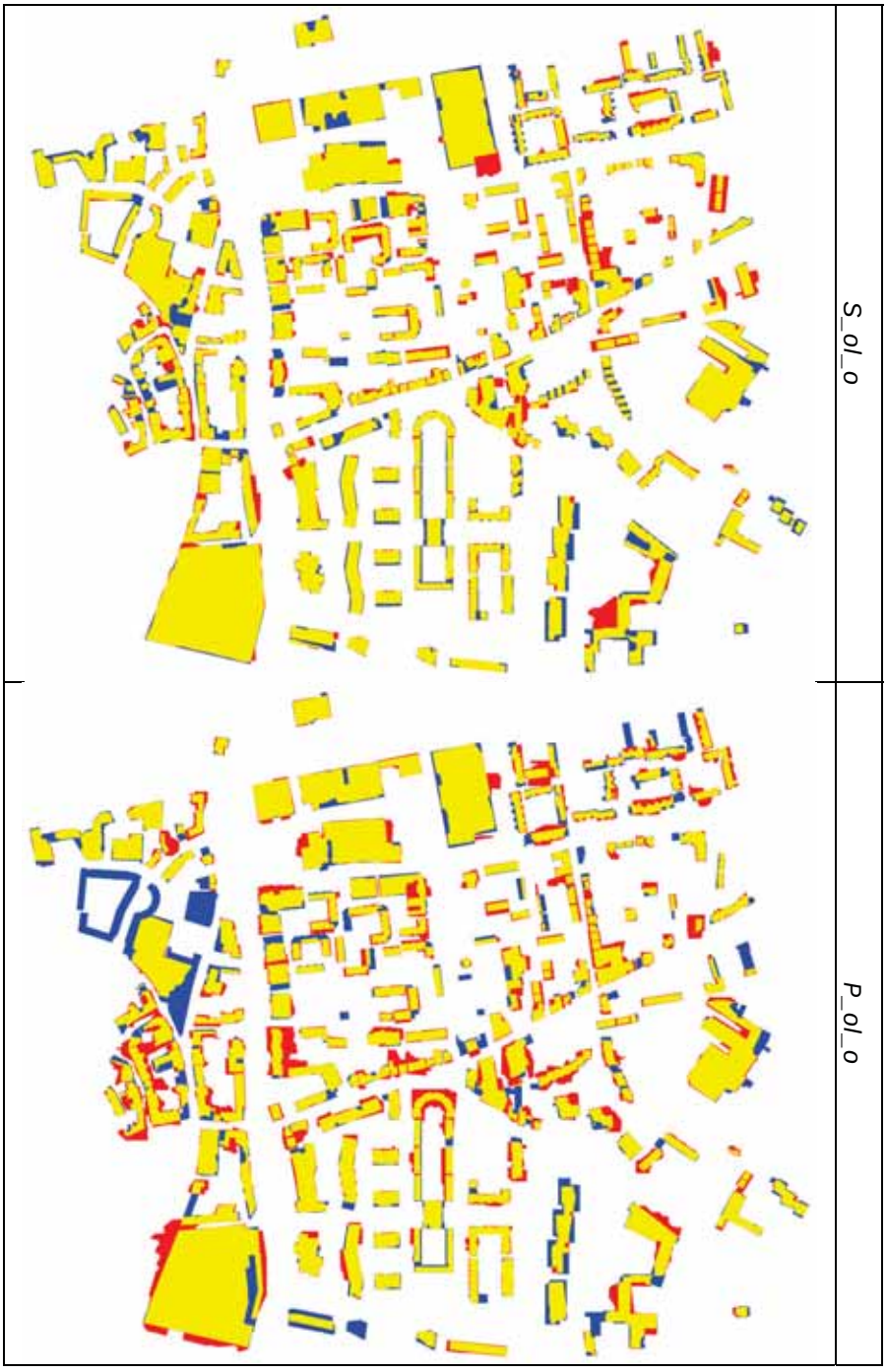
6.2.1 Pixel-based results

Raster images of the pixel-based evaluation corresponding to the five scenarios in Table 6.1 are visualised in Figure 6.1. The first image is the boundary of used reference buildings. In the result images, yellow indicates the true positive pixels, while blue indicates false negative pixels and red indicates false positive pixels.

Generally speaking, SenP is able to produce finer results than Pictometry in the sense that there are less red and blue pixels in the result images from the former data than the latter. By visually counting, there are six missed buildings in the Pictometry results but none in SenP. In some outlined

buildings, there are also large uncovered parts in the result images from Pictometry. Missed buildings and wrong outlining do occur less often in SenP.


S_c
P_c


S_ol_o
P_ol_o

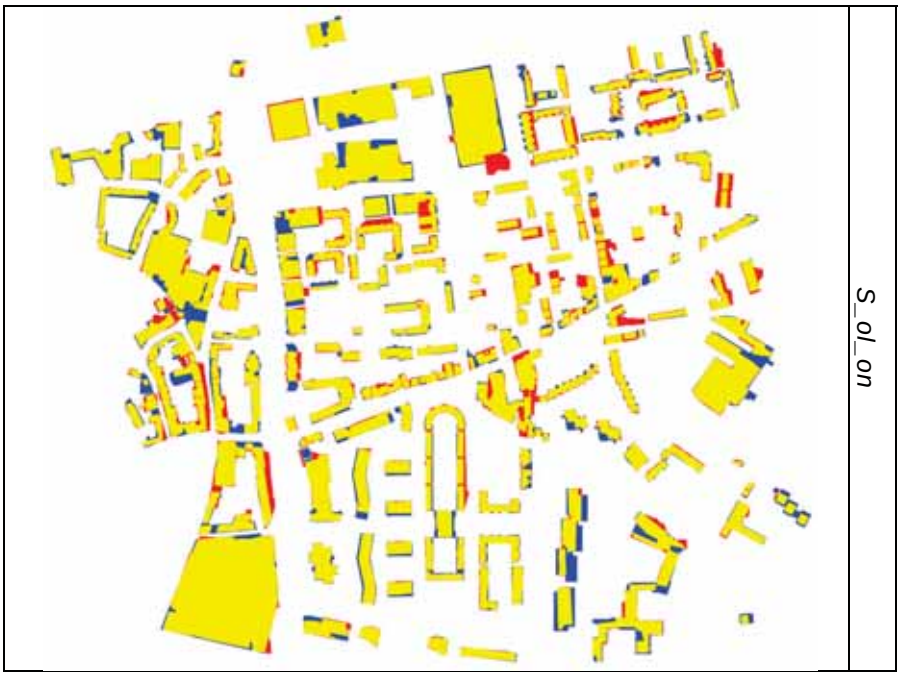


Figure 6.1 Raster images of pixel-based evaluation. Yellow – *TP*, blue – *FN*, red – *FP*.

When comparing between the results from the contour from the point cloud and the outlines by the proposed method, there are less uncovered parts (blue pixels) in the latter than in the contour results. In addition, the contours usually exceed the real building boundaries. The contour either performs poorly with *arc-shaped* buildings, for example the one in the left upper corner. Since it only has a very narrow opening, the contour "jumps" through the opening and takes it as a solid object.

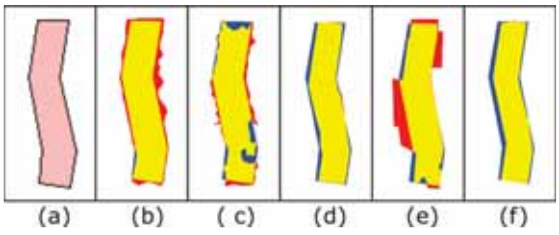


Figure 6.2 Example of the outlines of one building. (a) reference; (b) from *S_c*; (c) from *S_ol_o*; (d) from *S_ol_on*; (e) from *P_c*; (f) from *P_ol_o*.

Moreover, the contours are quite irregular compared with the outlines. A sample building is zoomed in for the comparison (Figure 6.2). The contours

from SenP and Pictometry are quite rough, while the outlines are more regularized.

Adding nadir images to the oblique images for the outlining process results in more accurate outlines. By comparing raster images from *S_ol_o* and *S_ol_on*, it is noticeable that there are less blue or red pixels around the buildings. When oblique images of one direction are absent, the edges in that direction may not be correctly identified, which could be revised by the nadir images. For example, for the large building in the southeast corner oblique images viewing north and west are not available. Consequently, the outline is wrongly defined in the northwest part of this building. The outline is better when involving nadir images.

Statistical comparison of pixel-based assessment between the SenP and Pictometry scenarios is presented in Figure 6.3. Except the *correctness* of contours (*S_c*), the *completeness* and *correctness* of the results from SenP are higher in each corresponding scenarios with Pictometry images. The *correctness* of *S_c* is lowered by some *arc-shaped* buildings, whose insides are filled.

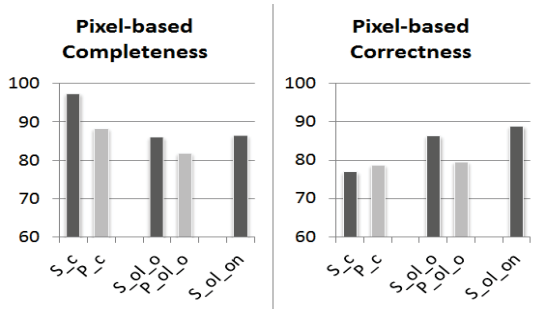


Figure 6.3 Pixel-based evaluation of the results

Comparing the contours and generated outlines within one dataset, the *completeness* is higher from contours but the *correctness* is higher from outlines. The area inside the contour from the original PMVS points for a single building tends to exceed the actual boundary of the reference. Therefore the pixel-based *completeness* of the contours is quite high: 97% and 89% in SenP and Pictometry respectively. The *correctness* of the contours is lower due to the pixels exceeding the building boundaries, which are probably caused by the roof overhang which is not included in the reference as well as the amount of noise in the point cloud around the building.

After roof region refinement, the 2D building regions are restricted by the evidences from images, thus improving the *correctness*. However, the

completeness decreases. On one hand, the roof region depends on the segmentation(s) selected from images. For some buildings especially flat roof buildings, there are narrow borders at the edge which are a bit higher than then roof planes. These borders are sometimes separated in the image segmentation from the main roof plane but segmented as part of walls. In this case, the selected regions from images are actually smaller than the real roof plane. On the other hand, it is negatively affected by the inaccurate input line evidences for outlining, e.g. when the extracted 3D lines are from the inner side of the mentioned border.

By adding nadir images in SenP dataset, both *completeness* and *correctness* are slightly improved. This is reasonable in the sense that nadir images partially solve problems caused by occlusion and increase the evidences for building edges.

6.2.2 Object-based results

The *completeness* and *correctness* of object-based results are based on the topological clarification process. Most of them are above 90% and they do not differentiate much among the different scenarios (Figure 6.4).

SenP still in most cases has better results than Pictometry. By integrating oblique image information in the outlining process, the *completeness* is lowered but the *correctness* is improved. The same trend is revealed after adding nadir images.

There could be a small number of buildings missing due to the failure in obtaining a point cloud in the ROI of the detected facade. Hence the number of outlined buildings is not exactly the same as the number of detected buildings. Failure in finding a valid fusion of image segments (section 5.5.3) is another cause of lower *completeness* after using images rather than contours.

Recall that the topological clarification is based on 50% overlap. This means that unmatched pairs also count for the source of *FP* and *FN*. Since there are many buildings that have incorrect outlines leading to many *FP* or *FN* pixels, the outlined buildings cannot be matched to the reference ones. These buildings are then recognized as *FP* and *FN* buildings in the assessment. As the outlines are more accurate in SenP hence the number of matched pairs increases, leading to better *completeness* and *correctness*. However, the number of mis-matches is smaller in the contours from point cloud than in the generated outlines. This is because the noise from point cloud is usually within 1 m from the actually boundary, which will not cause mis-matches during the topological clarification.

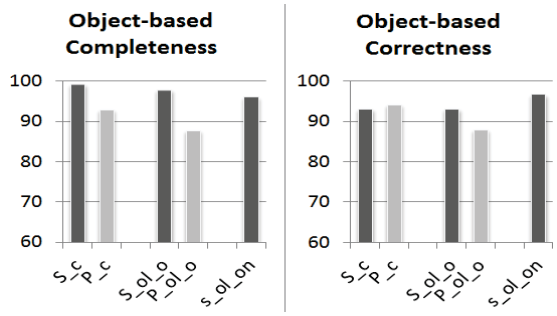


Figure 6.4 Object-based evaluation of the results

Table 6.2 lists the *m*:*n* relations of the outlined buildings for various datasets. The result does not reveal obvious superiority on SenP over Pictometry.

Table 6.2 Relations between outlined building set and reference (outlined : reference, excluding 1:1 relation)

	S_c	***P_c***	***S_ol_o***	***P_ol_o***	***S_ol_on***
m:1	0	4	31	16	14
1:*n*	23	25	19	21	19
m:*n*	0	0	1	0	0
Number of outlines	172	172	184	195	230
Number of reference buildings	219				

The main problem with contours is the over-connection problem ($1{:}n$ relation). The narrow gaps between buildings (<2m) are probably covered by the noise, thus the buildings are wrongly connected in the contours.

After integrating the information from oblique images, the problem of over-connection is a bit relieved, at the cost of introducing quite a number of under-connection problems. The major factor of this problem is the mis-detection of small or low connecting building parts during the building model initialisation step (section 5.5.4). Other causes are wrong evidences from image segmentation due to occlusion, shadow or lack of images from a certain direction.

Nadir images can be used to connect some wrongly separated building parts, for the reason that they provide top views so that the occluded parts become visible again. The $m{:}n$ relation usually does not occur.

Two examples for multiple relations are given in Figure 6.5. For building (a), its small connections between sub-building blocks are missed when only use

oblique image. The small planes of the connections cannot be extracted from the point cloud, because they are not visible from oblique views. The connections are visualised more clearly in the nadir images, which helps to re-connect the sub-blocks in the result from all images.

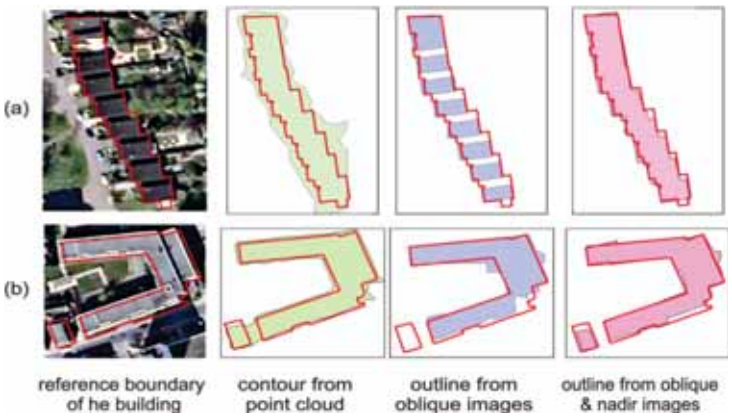


Figure 6.5 Examples from SenP for over-connection and under-connection

Building (b) reveals an opposite example where the images play a role in separating over-connected buildings. The gap between the large and small buildings is filled by the noise of point cloud, resulting in the incorrect connection between them in the contour. The smaller building in (b) is missed from the outline from oblique images. Nadir images then provide a stronger evidence for this building.

In general the 1:*n* problem usually appears on nearby buildings, especially when they align in the same row and have similar roofs. *m*:1 relations normally occur at small and low building structures.

6.2.3 Results from geometrical evaluation

The geometrical evaluation identifies the accuracy of the extracted boundaries. Figure 6.6 shows the histogram of the RMS of distances from the outline vertices to the matched reference. The histogram is computed using a bin size of 10 cm. Figure 6.7 reports the computed RMS of the boundaries by using formula (6-3).

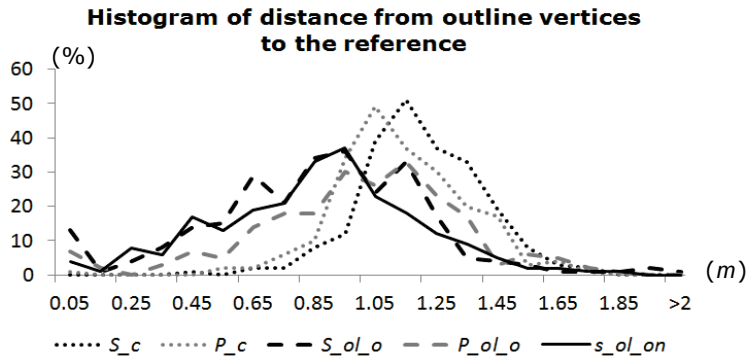


Figure 6.6 Histogram of the RMS of distances from generated outline vertices to the reference (m).

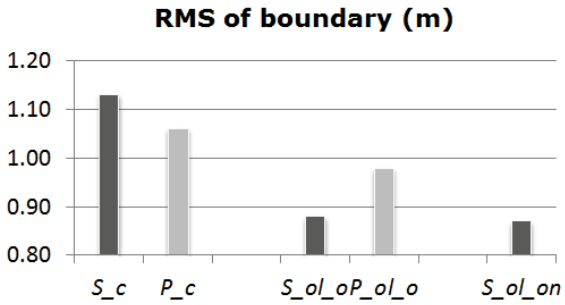


Figure 6.7 Geometrical accuracy of boundaries

The peaks of the histogram from five scenarios are consistent with the computed RMS of the boundary. The RMS of the contours form SenP and Pictometry are approximate 1.12 m and 1.07 m respectively. The point cloud accuracy used for contour generation is about 0.5 m and 0.7 m for SenP and Pictometry respectively. The contour is generated from the convex hull using a 3 m connection distance in order to cope with the parts with sparse matched points. As a result, the concave corners are not presented as corner but connected by the hypotenuse, leading to large errors around those corners. The Douglas-Peuker algorithm maintains the feature points, which could be the point with most deviation in the convex hull. The error from point cloud is therefore enlarged. Moreover, the roof overhang which is not included in the reference data also contributes to the error.

After integrating oblique images, the RMS reduced to 0.88 m and 0.98 m, where SenP has 0.1 m better accuracy than Pictometry. The geometrical accuracy only improves slightly after involving nadir images in SenP data set. The negative effects from the point cloud noise and the contour generation is

reduced, but the error caused by roof overhang remains. The higher border at the edge of flat roofs which is easily missed in the roof refining process is also one of the reasons.

According to the geometrical evaluation of the boundary, it is clear that the images are useful to improve the accuracy of the building boundaries. As SenP has a better result than Pictometry in the scenarios using oblique images, we can conclude that higher resolution and a large number of overlaps have a positive effect on the geometrical accuracy of the outlined buildings. However nadir images from SenP do not provide any significant improvement in the geometrical accuracy.

6.2.4 Results from *smoothness* evaluation

The histogram of the *smoothness* of the outlines from contours and outlines from SenP and Pictometry are presented in Figure 6.8 in different scales: (a) shows the histogram with the smoothness varying from 1/10 to 10 around 1. However there are still a large number of *smoothness* of buildings larger than 10, thus (b) extents the scale to 80 to reveal the data for the contours. Y-axis is the number of buildings.

The *smoothness* values of the contours are almost all above 1, which means that the contours contain more vertices than the reference contours. The peaks of the histograms for contours are both within 5 to 10, and the maxima reaches 80 for the two data sets. It shows that the number of vertices is commonly around five or ten times than the really number. In some special cases, the contours are too complex compared with the reference. Since the contour is simplified by the Douglas-Peuker algorithm with the distance threshold of 1 m, many vertices in the contour means that there are many irregular corners caused by the noise around the building boundary.

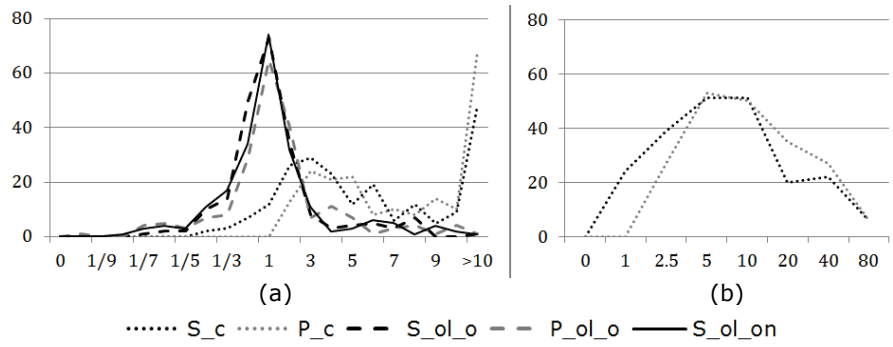


Figure 6.8 Histogram of *smoothness* of five scenarios. (a) and (b) show the *smoothness* in different scales.

In the contrast, all the histograms for the other three scenarios have sharp peaks around 1, namely there are similar number of vertices in the generated outlines with the reference. Given that the geometrical accuracies for those scenarios are less than 1 m, the shapes of the outlines can be considered to be similar with the reference.

6.2.5 Summary

The above result analysis shows the advantage of the proposed methods integrating image information for outline generalization over the extracted contours from the raw point cloud. The oblique images are quite useful to improve the *correctness* of the pixel-based and object-based results with a sacrifice on the *completeness*. The quality is further improved by adding nadir images. The nadir images are also quite helpful in reducing the *m*:1 and 1:*n* relations.

Furthermore, the improvements on geometrical accuracy and outline smoothness after employing oblique images are quite obvious, whilst the particular improvement provided through nadir images is not clear. The geometrical accuracy is affected by the roof overhang, higher border at the roof edges and the generalization effects on the tiny structures of buildings, such as the "teeth like" shape buildings.

Within all the assessments, SenP yields better results than Pictometry. This could be explained by the higher resolution and overlap, which will be further discussed in the next section.

6.3 Discussion on image factors affecting outlining

According to the results analysis, there are three main factors that influence the outlining result: image resolution, image overlap and availability of nadir images. In this section, we will present a detailed discussion based on the designed comparison.

6.3.1 Comparison design

Sixteen buildings are selected from the area where both SenP and Pictometry have full coverage in all four directions to assess the image impacts on the proposed method. The innovation of the proposed outlining method is the integration of image information, thus we will focus on testing the impacts of image factors on the outlining procedure, which is done by changing the input images.

In order to reduce the computation effort for generating point clouds, we start the comparison after the step of building model initialisation (section

5.4). It means that we keep the initial detected façades and the generated point cloud from all oblique images for each data set, but change the input images for roof region refinement (section 5.5) and outlining (section 5.6). Therefore, the results present the image impacts on the procedure of roof region refinement.

The comparison design is described below and summarized in Table 6.3.

I. Impact from resolution (P_r vs. S_r)

The impact of resolution is tested by using the SenP and Pictometry images. As Pictometry only provides oblique images, no nadir images from SenP are used either. The number of input images from SenP (*S_r*) and Pictometry (*P_r*) per direction is made exactly the same. In this setting, the only difference is in the image resolution. In this test, the initial building models are from SenP and Pictometry separately.

II. Impact from overlap (S_o vs. S_hf vs. S_hs vs. S_hfs)

Considering that the Pictometry has so sparse images compared with SenP, only SenP oblique images are used for this experiment. For each direction, 1) fully available oblique images (*S_o*), 2) images of half forward overlap (*S_hf*), 3) images of half side overlap (*S_hs*) and 4) images of half forward and side overlap (*S_hfs*) are used. With this setting we aim to test the impact from forward overlap, side overlap of the image sources. The input images for each design in this group are sketched in Figure 6.9.

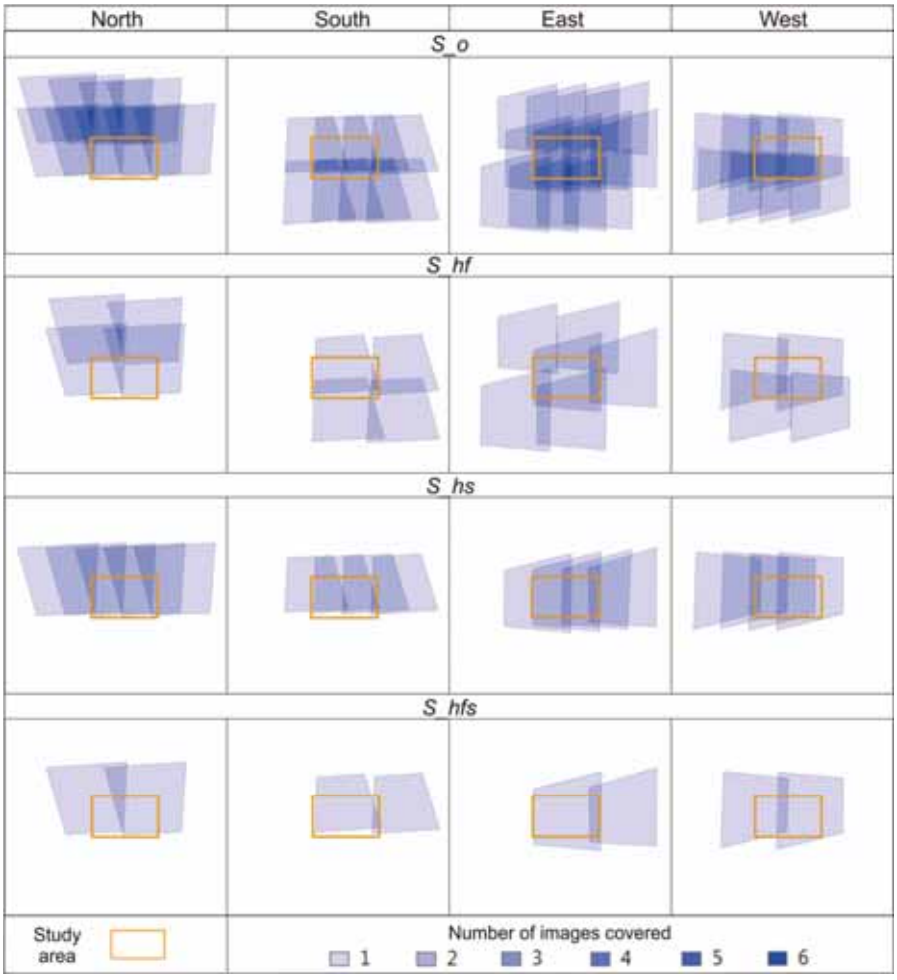


Figure 6.9 Input images for each comparison design

III. Impact from nadir images (S_o vs. S_n)

Results from the former section show that the quality of building outlining improves by adding nadir images. In this sub-section, we try to determine whether the improvement is caused by the increment of image number or the introduction of nadir images. For this reason, the total number of images will be fixed. One third of the total images of *S_o* are replaced by nadir images (*S_n*).

Table 6.3 Naming of testing image sets and their meaning and usage

Naming	Meaning	Impact
P_r	All Pictometry images	Resolution increase
S_r	SenP image set that has consistent number of Pictometry images in each directions	
S_hfs	SenP oblique images with half forward overlap and half sideward overlap	Overlap increase
S_hs	SenP oblique images with half sideward overlap	
S_hf	SenP oblique images with half forward overlap	
S_o	All SenP oblique images	
S_n	1/3 SenP oblique images from *S_o* are randomly replaced by nadir images	Add nadir images

6.3.2 Comparison results

The outlines from all image sets are evaluated to assess the impact of different image configurations. Since the selected buildings have good image coverage, they are all successfully outlined. The *smoothness* is also quite consistent for the generated outlines from SenP and Pictometry, therefore the assessment is only based on the pixel-based result (Figure 6.10) and geometrical accuracy of boundaries (Figure 6.11) evaluation.

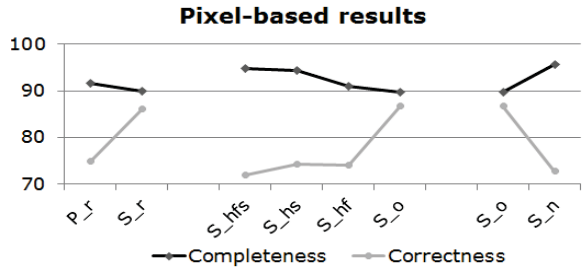


Figure 6.10 Pixel-based result for assessing image factors

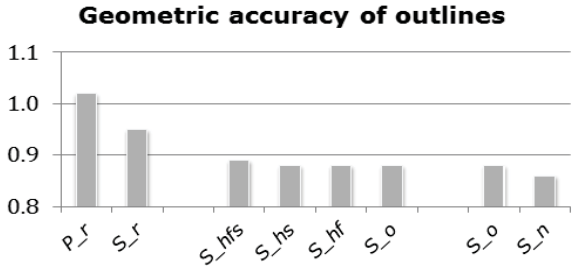


Figure 6.11 Geometrical accuracy of outlines for assessing image factors.

I. Comparison per factor

The impacts from image resolution are assessed by *P_r* vs. *S_r*, where image numbers are the same from all directions but the resolutions are different. The pixel-based *correctness* increases from 75% to 87% yet the *completeness* decreases by 1%. The geometric accuracy is also better in SenP. Based on the comparison, we conclude that the image resolution has positive influence on restricting the building regions and geometrical accuracy.

The impact of the overlap is assessed through *S_hfs* to *S_o*, where the overlap is increasing. We can see that with the increment of the overlap, both pixel-based quality indicators become better, but the geometrical accuracy is quite consistent. According to the comparison between forward overlap (*S_hf*) and sideward overlap (*S_hs*), it seems that the sideward overlap produced better pixel-based quality (similar *correctness* but better *completeness*). Referring to the image configuration in Figure 6.9, which shows that the overlap between images for each direction is better for *S_hs* than *S_hf*, the better completeness should be due to the better image overlap. It further proves that overlap plays an important role in producing higher quality results.

After replacing one third of the oblique images with nadir images, the nadir views do not guarantee a better result. We can conclude that nadir-looking is not the dominant reason for the increment of outlining accuracy. However, the decrease in pixel-based correctness may also because of the changed image configuration.

II. Comparison for all factors

The best result is generated by image setting *S_o* which uses the images higher resolution and has the best overlap. The worst result is from image setting *P_r* with worst resolution and lowest image overlap.

In all the three comparison groups, the pixel-based completeness decreases a bit with the increment of the correctness, but the increase in *correctness* is higher than the decrease in *completeness*. The highest quality is from the image setting of *S_r* and *S_o* when *S_r* has fewer images compared to *S_o*, and even less than *S_hf* and *S_hs*. The reason might be that in both settings of *S_r* and *S_o*, the number of images from four directions is almost the same, whilst in the settings of *S_hf* and *S_hs* the numbers of images are not balanced.

The geometric accuracy of *P_r* and *S_r* is in general lower than the others, which are all round 0.87 m. It could be due to the limitation on the number

of images used for each building. Although *S_r* has similar pixel-level accuracy with *S_o*, its geometric accuracy is worse.

In general all the three factors have positive influence on the outlining quality. But when there is efficient overlap, the geometric accuracy is restricted by other factors, such as resolution, outline generation process etc. Nadir view itself is not sufficient to generate a better result, but the result can be improved as long as nadir images are used as additional information.

6.4 Conclusions

The experiments with the SenP and Pictometry images show that the proposed outlining method is able to generate 2D outlines for almost all buildings in both datasets. Exceptions are the buildings for which no complete point cloud can be generated.

Based on the evaluation method used here, the result is better from SenP than from Pictometry. The *completeness* of the pixel-based evaluation on building area is 82% for Pictometry, and the *correctness* is 79%. They can be increased to 86% and 89% respectively with all SenP images, which is quite comparable with the works in the benchmark. However, we need to keep in mind that the built-up area used in this thesis is overall more complex than the "Vaihingen" dataset from the benchmark test. The geometrical accuracy of boundaries from SenP is within 0.9 m, 0.1 m better than from Pictometry data.

The quality of the results depends on the image resolution and overlap. In order to achieve better results, it is better to obtain images with higher resolution and overlap. Equality of the number of images viewing from all directions is also important, meaning that the difference of forward and sideward overlap should not be too different. Additional nadir images are also proven to play an important role in improving the quality of the outlines.

SUMMARY OF PART II

In this part we explored a way of outlining 2D boundaries of buildings from a complex urban scene by merely using multi-view oblique images. Instead of using the information such as the digital surface model or the point cloud from multi-view matching, we attempted to include original imagery in the process because not all information is included in the DSM or point cloud.

Based on the roof initialisation from the multi-view matched point cloud, an integration strategy using graph-cut is created to refine the initial roof planes. It used a 2.5D representation to combine all information extracted from object space and image space. This method aims to locate boundaries of the roof planes with high accuracy by using the evidences from the not-so-accurate boundaries of the point cloud, boundaries of image segments and the straight line segmented extracted from images. The final outline is the generalisation of the contour of the refined roof region.

This method relies on the initial point cloud and the input images. Buildings are not detected in case of insufficient matched points. For other cases, the approach results in 86% *completeness* and 89% *correctness* for finally outlined buildings. The RMSE from the outlined boundary to the reference is 0.9 m, where we observe an unknown impact from roof overhangs which are not modelled in the reference. High image resolution, high image overlap and availability of nadir images are proven to be important properties to generate best results.

Chapter 7 Conclusion and recommendations

7.1 Final Detection quality assessment

In chapter 6, to facilitate the comparison, only those buildings that are commonly detected in both SenP and Pictometry are selected for the assessment of the quality of the outlining. A final evaluation on the detection quality after façade detection and building outlining steps is assessed here, because falsely detected façades are eliminated and some buildings are missed in the outlining step. As the outlining quality is already assessed in chapter 6, here we only compute the overall *completeness* and *correctness* of the detection on object level.

Recalling that most of the falsely detected façades are roof structures or the inner façades not included in the reference boundary, they will not result in false detection on building level. Only wrongly detected façades outside real buildings, such as from fences may cause problems. However, they can also be successfully removed based on the insufficient point clouds generated around them. Therefore, all the detected buildings are true positive detection, namely 100% *correctness*.

If a reference building has an overlap with the outlined area larger than 50%, the reference building is considered as successfully detected. In total, two buildings and eight buildings are missed for SenP and Pictometry respectively through the outlining step. Consequently, 307 and 226 from SenP and Pictometry respectively out of a total of 335 tested buildings in the study area are successfully detected and outlined, leading to the final *completeness* 92% for SenP and 68% for Pictometry.

7.2 Conclusions

An automatic building mapping method using multi-view oblique imagery has been presented in this thesis. It consists of two separate parts: building façade detection and 2D outlining. Conclusions in detail have been made after each part. A more general overview on the whole approach and an analysis on input images are summarized in this section.

The only input data for the proposed method are the oriented oblique images with multi-views from different perspectives. Both steps integrate point clouds from multi-view image matching and the original information from single images. The advantage is that no registration is needed as the point cloud is generated from the same image set.

This work is one of the first researches dealing with object detection from oblique images. It exploits façade information for the building detection. A few researchers have worked on this topic using nadir images, e.g. the

research by Meixner and Leberl (2010). They pbtained 27.5 degree views on façades from nadir images so that their method requires very high overlap to compensate low resolution on façades from nadir view. However, our proposed method can work with only two images. The proposed façade detection approach is able to handle all buildings with vertical façades.

Instead of merely relying on the matched points for the 2D outlining, we attempt to first refine the roof planes by integrating several forms of low-level image information, such as the colour segmentation and extracted straight lines.

Experiments are done over a complex urban scene with two sets of images. There is no limitation on building type and the input data can be oblique images with about 45° tile angle, different resolutions and overlap. Thanks to the strategy which can handle a certain amount of noise in the point cloud, and which combines clues from multiple image segments, no highly sophisticated dense matching and segmentation algorithms are needed. In principle the methods used in thesis can be replaced, but the selected methods are able to pick up important features but ignore unnecessary details.

Results from the experiments reveal the effects of different imaging configuration: resolution, overlap and available perspectives. An increase in resolution has a positive effect on the completeness of the façade detection, as well as the geometric accuracy of the detected facades. It also helps to increase the accuracy of the outlined boundaries.

Correctness of the façade detection is more controlled by the image overlap. Façades can be validated in more images if the overlap is higher, and the correctness is therefore improved. Higher overlap is also effective on increasing the accuracy of the boundaries by providing more image information.

Missing images from one perspective directly lead to the missing of a set of façades facing that direction, which may lead to undetected buildings. The number of available image perspectives therefore affects the completeness of the façade detection. Settings of equal number of images in all perspectives results in better outlined boundaries. Additional nadir images are also helpful in improving the detection quality.

Since the start of this research, systematic oblique image collection has been boosted by various applications and the relatively low acquisition cost. Currently, oblique images are obtained in the cities all over the world. Furthermore, imaging techniques are developing very quickly for images of

higher resolution and bigger image size, such as the Leica mid-format camera RCD30, mounted in an oblique camera head (Leica, 2012). We suppose that the method described in this thesis will have a wide range of applications and further development with the newly collected data in the future.

Compared with the traditional ground survey, this method provides a quicker way of locating buildings. For the testing area (1 km^2) with about 250 images, it is completely doable on a laptop of 4 cores and 4G memory within one week. Most of the processing time is bound to the generation of point cloud (from each single direction and all directions) and the image texture evidence for façades detection which can be computed in parallel. Hence when processing on servers with dozens of cores and large memory or taking advantages of GPU programming, the processing time will be reduced further.

Processing on the collected oblique images of high resolution and overlap (SenP), more than 90% completeness and 100% correctness can be achieved. The geometrical accuracy is about 0.9 m, including the uncertainty from the unknown roof overhang. Although the accuracy is not high enough for map generation, it can still be used for checking the need for updates, locating newly built objects or checking the change on the planimetric size of existing buildings. Another application of our result is to be used for generating navigation or tourist maps, where the requirement for geometric accuracy is not so stringent. Given its efficiency, this work can be rapidly deployed and used for disaster assessment, or for surveying on illegal buildings for government purposes. Moreover, many existing 3D modelling approaches require building boundaries as input, for which our extracted 2D outlines can also be used.

7.3 Recommendations

Based on the discussion of image factors on the proposed detection and outlining method, we can conclude that in order to obtain better building detection result in the form of 2D outline, the flight plan should consider high resolution and high overlap. Covering the whole area from all directions is quite important. Nadir images should also be included if possible.

In the process of building outlining, the roof planes that are initially extracted from the point cloud are important. When one plane is too biased, the steps following the building model initialisation will be based on a wrong geometry and thus lead to a wrong roof region. Validation of the initial planes in images should be considered further.

The method is unable to recognize small roof planes in the point model initialisation due to the failure to extract sufficient roof plane evidence from the point cloud. However, the details remain clear in the images. How to use the image information to update the initial model is subject for future research.

This method is developed in an area where most of the buildings are less than five floors in height. Only the roof part which is visible from most of the available images is extracted. The occlusion problem will be more serious in the cities with more high-rise buildings. In between such buildings, lower buildings may remain completely occluded in all oblique images and will therefore not be detected. However, buildings which are occluded or partially occluded in one or two directions should be recoverable. In this case, nadir images should be of more help.

The final building outlining is formed by using the generalized contour from the roof region and refining it by the known boundary segments. Outlines from this method are sometimes not so regularised. It is recommended to integrate additional information from extracted image lines into the final outline formation.

Height information from generated point clouds is used during the whole approach. Regarding the CityGML definition (Kolbe et al., 2005), the LoD1 models of the outlined buildings can be generated by defining a general height of one building. Furthermore, building planes employed in the outlining process are extracted in the object space. In this sense, the LoD2 models can also be generated by implementing methods on 3D building modelling, for example the method developed by Oude Elberink and Vosselman (2009).

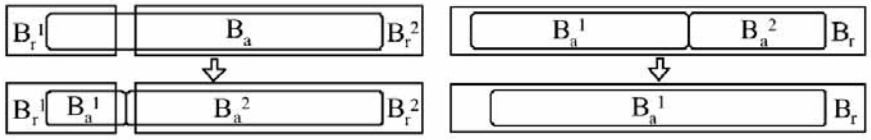
Appendix I

Parameters for façade detection for two image sets.

Meaning (symbol)	Picto-metry	SenP	Section
Height of *deviation images*	100		(3.3.1)
Width of *deviation images*	150		
Window size to calculate the average gradient	3		(3.3.2) Step I
Threshold for selecting image features ($\boldsymbol{T_{tr}}$)	5000		
Threshold for the roundness to select line pixels ($\boldsymbol{T_q}$)	0.3		
Size of moving window to count evidence over *direction image* ($\boldsymbol{n}$)	30	50	(3.3.2) Step II
Angle tolerance to from vertical /horizontal directions to locate vertical/horizontal edge pixels (degree)	5		
Threshold for the ratio between number of vertical/horizontal pixels over all window pixels ($\boldsymbol{T_1}$)	0.04		
Threshold for the ratio between the sum of vertical and horizontal pixels over all window pixels ($\boldsymbol{T_2}$)	0.12		
Threshold for the ratio between the sum of vertical and horizontal pixels over all feature pixels ($\boldsymbol{T_3}$)	0.55		
Window size to interpolate height map	5	9	(3.4)
Threshold to select façade pixels ($\boldsymbol{T_{facade}}$)	300	1000	(3.5.1)
Threshold for the number of pixels in a façade component	500	2000	(3.5.2)
Height of the fixed partition window	80	120	
Width of the fixed partition window	100	150	
Angle tolerance for façade verification (degree) ($\boldsymbol{\delta}$)	5		(3.6)
Overlap threshold for façade verification	75%		
Distance tolerance for façade verification (m)	2	1	

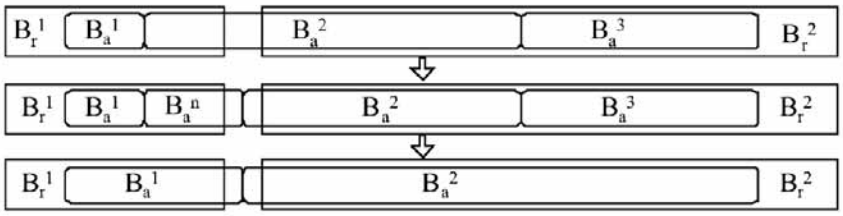
Appendix II

Topological clarification described in (Rutzinger et al., 2009).



(a) Topological clarification for two reference buildings that were merged in the detection process.

(b) Topological clarification for a reference building that was split in the detection process.



(c) Topological clarification in a case where both splitting and merging occurred in the detection process.

Figure Appendix II. Topological clarification. Detected buildings have round corners marked as B_a, and reference buildings have sharp corners marked as B_r. From literature (Rutzinger et al., 2009).

The aim of ***topological clarification*** is to make each building in one data set has only zero or one corresponding buildings in the other data set.

There are three cases:

- First, the cases of merged buildings are resolved. Any detected building that corresponds to buildings in the reference is split into new buildings. The Voronoi diagram of the reference labels is used to decide where the borders between the split buildings are situated [Figure Appendix II (a)].

- Second, the cases of split buildings, i.e., cases where detected buildings correspond to one building in the reference, are resolved by merging the detected buildings [Figure Appendix II (b)], even if there are gaps between them

- Third, by proceeding in this order, the cases where both splitting and merging occurs are also clarified [Figure Appendix II (c)].

Bibliography

Arbeláez, P., Maire, M., Fowlkes, C., Malik, J., 2011. Contour Detection and Hierarchical Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (5), 898-916.

Awrangjeb, M., Ravanbakhsh, M., Fraser, C.S., 2010. Automatic detection of residential buildings using LIDAR data and multispectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing 65 (5), 457-467.

Baillard, C., Maître, H., 1999. 3-D Reconstruction of Urban Scenes from Aerial Stereo Imagery: A Focusing Strategy. Computer Vision and Image Understanding 76 (3), 244-258.

Baltsavias, E.P., 2004. Objects extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. ISPRS Journal of Photogrammetry and Remote Sensing 58 (3-4), 129-151.

Blom, 2012. http://www.blomasa.com/ (Accessed March 9, 2012)

Collins, R.T., Jaynes, C.O., Cheng, Y.Q., Wang, X., Stolle, F., Riseman, E.M., Hanson, A.R., 1998. The ascender system: Automated site modeling from multiple aerial images. Computer Vision and Image Understanding 72 (2), 143-162.

Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5), 603-619.

COWI, 2012. http://www.cowi.com/menu/project/GeographicalInformationandIT/Mappinganddatacapture/Aerialsurveying/Pages/obliquephotosforthecityofgothenburgsweden.aspx (Accessed November 9, 2012)

Croitoru, A., Doytsher, Y., 2004. Right-angle rooftop polygon extraction in regularised urban areas: Cutting the corners. The Photogrammetric Record 19 (108), 311-341.

Dorninger, P., Pfeifer, N., 2008. A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds. Sensors 8 (11), 7323-7343.

Felzenszwalb, P.F., Huttenlocher, D.P., 2004. Efficient graph-based image segmentation. International Journal of Computer Vision 59 (2), 167-181.

Förstner, W., 1994. A framework for low level feature extraction. Computer Vision (ECCV'94), 383-394.

Fradkin, M., Maître, H., Roux, M., 2001. Building detection from multiple aerial Images in dense urban areas. Computer Vision and Image Understanding 82 (3), 181-207.

Fradkin, M., Roux, M., Matre, H., 1999. Building detection from multiple views International Archives of the Photogrammetry and Remote Sensing 32 (Part 3-2W5), 81-86.

Frueh, C., Sammon, R., Zakhor, A., 2004. Automated texture mapping of 3D city models with oblique aerial imagery. Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04), 396-403.

Fuchs, C., Heipke, C., Eder, K., Gülch, E., Förstner, W., 1998. OEEPE survey on 3D-city Models / Performance of the tie point extraction in automatic aerial triangulation. European Organization for Experimental Photogrammetric Research, Frankfurt am Main.

Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multi-view stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (8), 1362-1376.

GBKN, 2012. http://www.gbkn.nl/nieuwesite/basiskaartonline/start.html (Accessed September, 10, 2012)

GEODIS, 2012. http://www.geodis.cz/services/oblique-photography-pixoview (Accessed March 7, 2012)

Gerke, M., 2008. Dense image matching in airborne video sequences. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science 37 (Part B3b), 639-644.

Gerke, M., 2011. Using horizontal and vertical building structure to constrain indirect sensor orientation. ISPRS Journal of Photogrammetry and Remote Sensing 66 (3), 307-316.

Gerke, M., Heipke, C., Straub, B.-M., 2001. Building extraction from aerial imagery using a generic scene model and invariant geometric moments. IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, 85-89.

Gerke, M., Kerle, N., 2011. Automatic Structural Seismic Damage Assessment with Airborne Oblique Pictometry (c) Imagery. Photogrammetric Engineering and Remote Sensing 77 (9), 885-898.

Getmapping, 2012. http://www2.getmapping.com/Products/Oblique-Aerial-Photography (Accessed November 9, 2012)

Grenzdőrffer, G.J., Guretzki, M., Friedlander, I., 2008. Photogrammetric image acquisition and image analysis of oblique imagery. The Photogrammetric Record 23 (124), 372-386.

Gruen, A., Baltsavias, E.P., Henricsson, O., 1997. Automatic extraction of man made objects from aerial and space images (II), Centro Stefano Franscini. Birkhäser-Verlag, Ascona, p. 394.

Grün, A., Baltsavias, E., Henricsson, O., 1997. Automatic Extraction of Man-Made Objects from Aerial and Space Images (II). Birkhäuser, Basel.

Grün, A., Kübler, O., Agouris, P., 1995. Automatic Extraction of Man-Made Objects from Aerial and Space Images. Birkhäuser, Basel.

Haala, N., Brenner, C., 1999. Extraction of buildings and trees in urban environments. ISPRS Journal of Photogrammetry and Remote Sensing 54 (2-3), 130-137.

Hirschmüller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2), 328-341.

Höhle, J., 2008. Photogrammetric measurements in oblique aerial images. Photogrammetrie, Fernerkundung, Geoinformation 2008 (1), 7-14.

Hu, J., You, S., Neumann, U., Park, K.K., 2004. Building modeling from LiDAR and aerial imagery. ASPRS Annual Conference (on CDROM).

Jurisch, A., Mountain, D., 2008. Evaluating the viability of Pictometry® imagery for creating models of the built environment, in: Gervasi, O., Murgante, B., Laganà, A., Taniar, D., Mun, Y., Gavrilova, M. (Eds.), Computational Science and Its Applications – ICCSA 2008. Lecture Notes in Computer Science, 2008. Springer Berlin / Heidelberg, pp. 663-677.

Kaartinen, H., Hyyppä, J., Gülch, E., Vosselman, G., Hyyppä, H., Matikainen, L., Hofmann, Å., Mäder, U., Persson, U., Sőderman, U., Elmqvist, M., Ruiz, A., Dragoja, M., Flamanc, D., Maillet, G., Kersten, T., Carl, J., Hau, R., Wild, E., Frederiksen, L., Holmgaard, J., Vester, K., 2005. Accuracy of 3D city models: EuroSDR comparison International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science 36 (Part 3/W19), 227-232.

Karantzalos, K., Paragios, N., 2009. Recognition-driven two-dimensional competing priors toward automatic and accurate building detection. IEEE Transactions on Geoscience and Remote Sensing 47 (1), 133-144.

Karbo, N., Simmons, G., 2007. Aerial imagery from different angles, Professional Surveyor Magazine, pp. 18-23.

Katartzis, A., Sahli, H., 2008. A stochastic framework for the identification of building rooftops using a single remote sensing image. IEEE Transactions on Geoscience and Remote Sensing 46 (1), 259-271.

Khoshelham, K., Nardinocchi, C., Frontoni, E., Mancini, A., Zingaretti, P., 2010. Performance evaluation of automated approaches to building detection in multi-source aerial data. ISPRS Journal of Photogrammetry and Remote Sensing 65 (1), 123-133.

Kim, H.T., Kim, S.B., Go, J.S., Eo, Y.D., Lee, B.K., 2010. Building 3D geopatial information using airborne multi-looking digital camera system. Journal of Convergence Information Technology 5 (1), 15-22.

Kim, Z., Nevatia, R., 2004. Automatic description of complex buildings from multiple images. Computer Vision and Image Understanding 96 (1), 60-95.

Kolbe, T., Gröger, G., Plümer, L., 2005. CityGML: Interoperable Access to 3D City Models, in: Oosterom, P., Zlatanova, S., Fendel, E. (Eds.), Geo-information for Disaster Management. Springer Berlin Heidelberg, pp. 883-899.

Lafarge, F., Descombes, X., Zerubia, J., Pierrot-Deseilligny, M., 2008. Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling. ISPRS Journal of Photogrammetry and Remote Sensing 63 (3), 365-381.

Lafarge, F., Mallet, C., 2011. Modeling urban landscapes from point clouds: a generic approach, INRIA Technical Report.

Lafarge, F., Mallet, C., 2012. Creating Large-Scale City Models from 3D-Point Clouds: A Robust Approach with Hybrid Representation. International Journal of Computer Vision 99 (1), 69-85.

Le Besnerais, G., Sanfourche, M., Champagnat, F., 2008. Dense height map estimation from oblique aerial image sequences. Computer Vision and Image Understanding 109 (2), 204-225.

Leberl, F., Irschara, A., Pock, T., Meixner, P., Gruber, M., Scholz, S., Wiechert, A., 2010. Point Clouds: Lidar versus 3D Vision. Photogrammetric Engineering and Remote Sensing 76 (10), 1123-1134.

Leica, 2012. http://www.leica-geosystems.com/en/Leica-RCD30-Oblique_99831.htm (Accessed November 16, 2012)

Lemmen, M., Lemmen, C., Wubbe, M., 2007. Pictometry : potentials for land administration. the 6th FIG regional conference, International Federation of Surveyors (FIG), (on CD-ROM).

Lin, C., Huertas, A., Nevatia, R., 1994. Detection of buildings using perceptual grouping and shadows. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94), 62-69.

Lin, C., Nevatia, R., 1995. 3-D descriptions of buildings from an oblique view aerial image. Proceedings of the International Symposium on Computer Vision, 377-382.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60 (2), 91-110.

Ma, R.J., 2005. DEM generation and building detection from Lidar data. Photogrammetric Engineering and Remote Sensing 71 (7), 847-854.

Maas, H.-G., 1999. The potential of height texture measures for the segmentation of airborne laser scanner data. The Fourth International Airborne Remote Sensing Conference and Exibition / 21st Canada Symposium on Remote Sensing.

Mayer, H., 1999. Automatic object extraction from aerial imagery—a survey focusing on buildings. Computer Vision and Image Understanding 74 (2), 138-149.

McGlone, J.C., Shufelt, J.A., 1994. Incorporating vanishing point geometry in building extraction techniques. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94), 54-61.

Meixner, P., Leberl, F., 2010. Characterizing building facades from vertical aerial images. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science 38 (Part 3A), 98-103.

Mishra, P., Ofek, E., Kimchi, G., 2008. Validation of vector data using oblique images. Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems, (on CD-ROM).

Müller, S., Zaum, D.W., 2005. Robust building betection in aerial images. Proceedings of the ISPRS Workshop CMRT 2005 34 (part 3/W24), 143-148.

Nevatia, R., Lin, C., Huertas, A., 1997. A system for building detection from aerial images. Automatic Extraction of Man-Made Objects from Aerial and Space Images 2, 77-86.

Niemeyer, J., Wegner, J.D., Mallet, C., Rottensteiner, F., Soergel, U., 2011. Conditional random fields for urban scene classification with full waveform LiDAR data. Photogrammetric Image Analysis (PIA 2011), 233-244.

Noronha, S., Nevatia, R., 2001. Detection and modeling of buildings from multiple aerial images. IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (5), 501-518.

Nyaruhuma, A.P., Gerke, M., Vosselman, G., 2010. Evidence of walls in oblique images for automatic verification of buildings. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science 38 (Part 3A), 263-268.

Nyaruhuma, A.P., Gerke, M., Vosselman, G., Mtalo, E.G., 2012. Verification of 2D building outlines using oblique airborne images. ISPRS Journal of Photogrammetry and Remote Sensing 71, 62-75.

Ofek, 2012. http://www.ofek-air.com/Aerial-Photography (Accessed March 7, 2012)

Oude Elberink, S.J., Vosselman, G., 2009. Building reconstruction by target based graph matching on incomplete laser data: analysis and limitations. Sensors 9 (8), 6101-6118.

Oude Elberink, S.J., Vosselman, G., 2011. Quality analysis on 3D building models reconstructed from airborne laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing 66 (2), 157-165.

Peng, J., Zhang, D., Liu, Y.C., 2005. An improved snake model for building detection from high resolution urban aerial images. Pattern Recognition Letters 26 (5), 587-595.

Petrie, G., 2009. Systematic oblique aerial photography using multiple digital frame cameras. Photogrammetric Engineering and Remote Sensing 75 (2), 102-108.

Pictometry, 2012. http://www.pictometry.com/index.php?option=com_content&view=article&id=76&Itemid=85 (Accessed March 7, 2012)

Prandi, F., Achille, C., Brunmana, R., Fassi, F., Fregonese, L., 2008. LiDAR and Pictometry iamges integrated use for 3D model generation. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37 (Part B2), 763-770.

Rottensteiner, F., Sohn, G., Jung, J., Gerke, m., Baillard, C., Benitez, S., Breitkopf, U., 2012. The ISPRS benchmark on urban object classification and 3D building reconstruction. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3.

Rutzinger, M., Rottensteiner, F., Pfeifer, N., 2009. A Comparison of Evaluation Techniques for Building Extraction From Airborne Laser Scanning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2 (1), 11-20.

Schultz, S.L., Gulffrida, F.D., Gray, R.L., Mondello, C., 2011. Method and apparatus for capturing geolocating and measuring oblique images. Patent, U.S. US 7,995,799 B2

Shorter, N., Kasparis, T., 2009. Automatic Vegetation Identification and Building Detection from a Single Nadir Aerial Image. Remote Sensing 1 (4), 731-757.

Shufelt, J.A., 1996. Exploiting photogrammetric methods for building extraction in aerial images. International Archives of the Photogrammetry and Remote Sensing 31 (Part B 6/s), 74-79.

Shufelt, J.A., 1999. Performance evaluation and analysis of monocular building extraction from aerial imagery. IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (4), 311-326.

Sirmacek, B., Unsalan, C., 2008. Building detection from aerial images using invariant color features and shadow information. 23rd Internatinal Symposium on Compter and Information Science.

Slagboom en Peeters, 2012. Geo-obliek. http://www.slagboomenpeeters.com/obliek.html (Accessed March 7, 2012)

Sohn, G., Dowman, I., 2003. Building extraction using Lidar DEMs and Ikonos Images. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science 34 (Part 3/W13), 167-173.

Stassopoulou, A., Caelli, T., 2000. Building detection using Bayesian networks. International Journal of Pattern Recognition and Artificial Intelligence 14 (6), 715-733.

Stoter, J.E., Vosselman, G., Goos, J., Zlatanova, S., Verbree, E., Klooster, R., Reuvers, M., 2011. Towards a national 3D spatial data infrastructure : case of the Netherlands. Photogrammetrie - Fernerkundung - Geoinformation (6), 405-420.

Terrasolid Ltd., 2012. Use LiDAR, oblique images to texture 3-D city models, Earth Imaging Journal.

Tian, Y.X., Gerke, M., Vosselman, G., Zhu, Q., 2010. Knowledge-based building reconstruction from terrestrial video sequences. ISPRS Journal of Photogrammetry and Remote Sensing 65 (4), 395-408.

TMA, 2012. http://www.pgc.umn.edu/imagery/aerial/antarctica. (Accessed December 20, 2012)

Tournaire, O., Brédif, M., Boldo, D., Durupt, M., 2010. An efficient stochastic approach for building footprint extraction from digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing 65 (4), 317-327.

Tóvári, D., Vögtle, T., 2004. Classification methods for 3D objects in laserscanning data. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science 35 (Part B3), 408-413.

Verma, V., Kumar, R., Hsu, S., 2006. 3D Building Detection and Modeling from Aerial LIDAR Data, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06), pp. 2213-2220.

Vögtle, T., Steinle, E., 2000. 3D modeling of buildings using laser scanning and spectral information. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science 33 (Part B3), 927-934.

Vosselman, G., 1999. Building reconstruction using planar faces in very high sensity height data. International Archives of the Photogrammetry and Remote Sensing 32 (Part 3-2W5), 87-94.

Vosselman, G., 2012. Automated planimetric quality control in high accuracy airborne laser scanning surveys. ISPRS Journal of Photogrammetry and Remote Sensing 74 (0), 90-100.

Vu, T.T., Yamazaki, F., Matsuoka, M., 2009. Multi-scale solution for building extraction from LiDAR and image data. International Journal of Applied Earth Observation and Geoinformation 11 (4), 281-289.

Wang, M., Bai, H., Hu, F., 2008. Automatic texture acquisition for 3D model using oblique aerial images. The First International Conference on Intelligent Networks and Intelligent Systems (ICINIS '08) 495-498.

Weidner, U., Förstner, W., 1995. Towards automatic building extraction from high resolution digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing 50 (4), 38-49.

List of publications

- Jing Xiao, George Vosselman, Markus Gerke, 2009, "Edge extraction from multi-view oblique images", The 6th International Symposium on Digital Earth (ISDE6), Beijing, September 2009.

- Jing Xiao, George Vosselman, Markus Gerke, 2010, "Automatic Detection of Buildings with Rectangular Flat Roofs from Multi-View Oblique Imagery", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38 (Part 3A), pp. 251-256.

- Jing Xiao, Markus Gerke, George Vosselman, 2012, "Building extraction from oblique airborne imagery based on robust façade detection", ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 68, pp. 56-68.

- Jing Xiao, 2012, "Automatic building outlining from multi-view oblique images". ISPRS Annals of Photogrammetry, Remote Sensing and the Spatial Information Sciences, ISPRS Congress, Melbourne, Sydney.

Summary

Applying 3D information has become popular everywhere in people's daily life, especially the applications of 3D building models. Remotely sensed data has obtained great attention for the purpose of generating building information owning to its fast and relatively cheap data collection and processing. Nadir viewing sensors was typically used for detecting buildings. From those data only rooftops of buildings can be detected whilst the façade information is quite limited.

Airborne oblique imagery became a new data source recently, which is taken from aeroplane from tilted angles. Thus they can present the information of both building roofs and façades. With the development of new techniques on oblique data collection system afterwards, oblique images are widely captured and processed over the world.

This research aims at developing methods for detecting urban buildings from multi-view airborne oblique imagery. Façades are seldom adopted in the previous work with nadir view because they can be hardly seen or the resolution on it is quite low. However they can be clearly visible from oblique images. In our work, methods are first developed to detection building façades from oblique images to locate buildings. For the detected buildings, the 2D outlines are then delineated by integrating the information from façades and roof extracted from multiple images.

Two oblique image sets are available in this research. They were both taken from the view angle of 45° to 50°, but with different image resolution and overlap. By comparing the results from them, the effects of image factors such as image resolution and overlap are assessed on the process quality of façade detection and outlining.

Façade detection

Image features are analysed and selected to distinguishing building façades from other objects such as roofs or ground. In the proposed method, multiple overlapping images viewing the same direction are required to detect façades facing that direction. Façades are detected adopting evidences of façade texture from single image and height gradient from multiple images. Evidence of façade texture mainly relies on vertical and horizontal structures. The process of looking for this evidence include identifying the directions of vertical and horizontal in the image patch as well as testing their density. Evidence of height gradient is extracted from the point cloud generated by a dense image matching technique. The point cloud is projected into original images to assign each pixel the height information. Façade hypothesis are generated combining the two evidences in each single image. Since pixels

implicitly bear height information, the hypotheses are formed in the object space so that façades are verified by testing the 3D hypotheses from multiple images.

The output of this approach is a set of 2D projections of vertical rectangles presenting the location of detected building façades. 90% and 70% buildings are detected from SenP and Pictometry images respectively. The correctness is 95% in terms of building detection for both data sets. The geometric accuracy of detected façades is 1.5° to 2.5° in orientation and 0.5 m to 0.7 m in distance from images of high resolution and high overlap (SenP) to images with half of the resolution and very low overlap (Pictometry). Both image resolution and overlap are proven to be effective in improving façade detection quality in terms of completeness, correctness and geometric accuracy. The result from the work will be used as the input for the later stages of building outlining.

Building outlining

A point cloud is generated around an input façade to define the ROI of a target building hypothesis. The façades around the component of off-ground points are clustered as the detected façades for this hypothesis. Planar planes are extracted from the off-ground point cloud. They are classified into roof, wall and ground patches so as to get the initial building model. Then the planes are refined by integrating other information by a graph-cut approach in 2.5D rasterised space. The information includes roof and wall patches, projections of detected façades, projections of wall patches, extracted 3D lines from images and segmented images. The outline is generalised from the refined roof planes. A direction set is employed to formalise the line segments in the outline to obtain a higher accuracy.

The pixels based evaluation of outlined buildings reveals that the completeness and correctness from both data sets are within 79% and 89%. The result is better from SenP than from Pictometry. The geometric accuracy of the boundaries from SenP is within 0.9 m, 0.1 m better than from Pictometry data. The uncertainty of the geometric accuracy is relating to the uncertainty from unknown roof overhang as well as outlining approach. Image factors of high resolution, high overlap and nadir images are proven to be effective in generating high quality outlines.

After the whole process, in total 92% of the buildings in the study area are successfully detected and outlined using the image data with high resolution and high overlap, whilst only 68% from the image set with low resolution and overlap. No false detection at non-building location.

This work is one of the first researches dealing with object detection exploiting façade information from airborne images. It provides a fast way of locating buildings from merely oblique images. After a small training on parameters, the whole process can be done fully automatically.

Since oblique image are collected over the world recently, the applications of the method can be very wide. The possible applications are checking for map updates, locating informal buildings, navigation and assist 3D modelling.

Samenvatting

Het gebruik van 3D informatie, in het bijzonder van 3D gebouwmodellen, is in het dagelijks leven steeds populairder geworden. Voor de productie van gebouwmodellen wordt veelvuldig gebruik gemaakt van opnamen vanuit vliegtuigen, omdat deze manier van gegevensinwinning en –verwerking relatief snel en goedkoop is. Voor de detectie van gebouwen werd normaal gebruik gemaakt van verticale luchtfoto's (nadiropnamen). In deze foto's zijn de daken van gebouwen goed zichtbaar, maar is de informatie over gevels vrij beperkt.

Oblieke luchtfoto's, die onder een hoek met het vliegtuig worden genomen, zijn recentelijk een nieuwe bron van gegevens. Deze foto's bevatten informatie over zowel daken als gevels. Als gevolg van nieuwe technieken voor de opname van oblieke foto's worden deze foto's nu op grote schaal over de gehele wereld opgenomen en verwerkt.

Dit onderzoek richt zich op de ontwikkeling van methoden voor de detectie van stedelijke bebouwing in oblieke luchtfoto's. Gevels worden zelden beschouwd in eerder werk met nadiropnamen, omdat deze nauwelijks te zien zijn of omdat de resolutie vrij laag is. Gevels zijn echter goed zichtbaar in oblieke foto's. In ons werk zijn eerst methoden ontwikkeld om gebouwen te lokaliseren middels de detectie van gevels. Voor de gevonden gebouwen zijn vervolgens 2D omlijningen bepaald door informatie over de gevels te integreren met informatie over daken, die uit meerdere overlappende foto's is afgeleid.

Voor dit onderzoek waren twee datasets met oblieke luchtfoto's beschikbaar. Foto's van beide datasets waren opgenomen onder een hoek van 45° tot 50°, maar met verschillende beeldresolutie en mate van overlap. Door de resultaten met deze datasets te vergelijken is het effect van deze beeldfactoren op de kwaliteit van de gebouwdetectie en gebouwomlijningen bepaald.

Geveldetectie
Om gevels te onderscheiden van andere objecten, zoals daken en grond, zijn beeldkenmerken geanalyseerd en geselecteerd. In de voorgestelde methode zijn meerdere overlappende foto's, die in dezelfde richting kijken, nodig om gevels dwars op de kijkrichting te detecteren. De gevels worden gedetecteerd door gebruik te maken van textuur op enkele beelden en hoogteveranderingen bepaald uit meerdere beelden. Dit proces omvat ook het bepalen van patronen met horizontale en verticale richtingen in beelduitsneden en het toetsen van hun dichtheid. Hoogteveranderingen worden verkregen uit een puntenwolk, die door *dense matching* is bepaald.

Deze puntenwolk is op de originele foto's geprojecteerd om aan ieder pixel een hoogte toe te kennen. Hypotheses voor de aanwezigheid van gevels worden gegenereerd door de twee kenmerken in elk beeld te combineren. Omdat de pixels ook voorzien zijn van hoogte-informatie, worden de hypotheses gevormd in de objectruimte, zodat zij geverifieerd kunnen worden door het toetsen van 3D hypotheses vanuit meerdere beelden.

Het resultaat van deze aanpak is een set van 2D rechthoeken in het platte vlak die de locatie van de gevonden gevels weergeven. In de beelden van Slagboom en Peeters (SenP) en die van Pictometry worden resp. 90% en 70% van de gebouwen gedetecteerd. Voor beide datasets is de gebouwdetectie in 95% juist. De geometrische nauwkeurigheid van de gedetecteerde gevels varieert van 1.5° tot 2.5° in de oriëntering en van 0.5 tot 0.7 m in de locatie in beelden met hoge resolutie en hoge overlap (SenP) en beelden met 50% lagere resolutie en zeer geringe overlap (Pictometry). Zowel beeldresolutie als overlap zijn belangrijk om de geveldetectie te verbeteren in termen van volledigheid, juistheid en geometrische nauwkeurigheid. Het resultaat van dit werk wordt gebruikt als invoer voor de volgende fase van het omlijnen van de gebouwen.

Omlijnen van gebouwen
Rondom een gevonden gevel wordt een puntenwolk gegenereerd als aandachtsgebied voor een gebouwhypothese. De gevels rondom een boven het grondniveau gelegen component van de puntenwolk worden geclusterd en beschouwd als gedetecteerde gevels van een gebouw. Uit de component van de puntenwolk worden vlakken geëxtraheerd. Deze worden geclassificeerd als dak-, muur- en grondvlakken en bepalen zo een eerste gebouwmodel. Vervolgens worden de vlakken verfijnd door verdere informatie met een *graph-cut* methode te integreren in een 2.5D raster. De invoer hiervoor bevat dak- en muursegmenten, projecties van gedetecteerde gevels, projecties van muurvlakken, geëxtraheerde 3D lijnen en gesegmenteerde foto's. De omlijningen worden gegeneraliseerd uitgaande van de verfijnde dakvlakken. Een set met hoofdrichtingen voor lijnsegmenten wordt gebruikt om een hogere nauwkeurigheid te verkrijgen.

Een rastergebaseerde evaluatie van de omlijnde gebouwen laat zien dat de volledigheid en juistheid van beide datasets beter is dan resp. 79% en 89%. Het resultaat met SenP beelden is beter dan met Pictometry beelden. De geometrische nauwkeurigheid van de omlijning in SenP beelden is 0.9 m, 0.1 m beter dan die met Pictometry beelden. De geometrische nauwkeurigheid hangt af van de onzekerheid door dakoverstekken en van de onzekerheid in de methode van omlijnen. Beeldfactoren als hoge beeldresolutie, grote overlap, en beschikbaarheid van nadiropnamen zijn van belang voor het genereren van omlijningen van een hoge kwaliteit.

Aan het einde van het proces zijn in beelddata met hoge resolutie en overlap in totaal 92% van de gebouwen in het studiegebied succesvol gedetecteerd en omlijnd, terwijl dit in beelddata met lage resolutie en overlap bij slechts 68% van de gebouwen het geval is. Objecten die geen gebouwen zijn worden niet ten onrechte als gebouw aangemerkt.

Dit werk is een van de eerste onderzoeken gericht op gebouwdetectie door gebruik van informatie over gevels uit luchtfoto's. Het beschrijft een snelle manier voor de lokalisering van gebouwen uit enkel en alleen oblieke luchtfoto's. Na een beperkte training voor het bepalen van parameterwaarden loopt het gehele proces automatisch.

Omdat oblieke foto's recent wereldwijd worden opgenomen kan de ontwikkelde methode op grote schaal worden toegepast. Mogelijke toepassingen zijn het controleren van benodigde kaartactualisaties, lokaliseren van informele nederzettingen, navigatie en ondersteuning bij 3D gebouwmodellering.

Curriculum Vitae

Jing Xiao (Emily) was born in Wuhan, China on January 6th, 1984. She started her Bachelor study on Geo-Information Science and Cartography in School of Resource and Environmental Science, Wuhan University, China in September 2002. During the year 2006 to 2008 she obtained Master of Science degrees on Geo-Information Engineering and Cartography from Wuhan University, China and another MSc degree on Natural Resource Management from ITC, the Netherlands. The subject of her MSc research is "Change detection of hydro-acoustic environment for Yangtze finless porpoise using remote sensing in Poyang Lake".

In December 2008 she started her PhD study in Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, the Netherlands, by the support from Chinese Scholarship Council. The research is under the theme of Acquisition and quality of geo-spatial information. Her task is to automatically detect buildings from oblique imagery and the result is presented in this thesis.

ITC Dissertation List

http://www.itc.nl/Pub/research/Graduate-programme/Graduate-programme-PhD_Graduates.html